



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
ACATLÁN**

**REPRESENTACIÓN DE ESPACIO DE
ESTADOS Y FILTRO DE KALMAN
APLICADOS EN LA CONSTRUCCIÓN DE
LA ETTI DEL MODELO DE VASICEK**

T E S I S

PARA OBTENER EL TÍTULO DE

A C T U A R I O

P R E S E N T A N

**FUENTES OROZCO RABEL
GARAY RAMÍREZ IXCHEL**

ASESOR:

M. EN C. VÍCTOR MANUEL ULLOA ARELLANO



ESTADO DE MÉXICO

2010



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Dedicamos este esfuerzo a nuestros respectivos padres.

Ixchel y Rabel

Índice

Índice.....	3
Introducción.....	5
Capítulo 1: Estructura temporal de tipos de interés de Vasicek (1977) de un factor.....	8
1.1 Tasas instantáneas, precio y rendimiento de bonos cupón cero bajo la medida de probabilidad neutral al riesgo.....	8
1.2 Modelado de la estructura temporal de tipos de interés.....	10
1.3 Tasa de interés de corto plazo del modelo de Vasicek (1977)	16
1.4 Curva de rendimientos de Vasicek (1977) aplicando la técnica de valuación neutral al riesgo.....	20
1.5 Ecuación diferencial parcial fundamental para el precio de los bonos cupón cero.....	24
1.5.1 Obtención por medio del factor de descuento estocástico.....	25
1.5.2 Obtención por el argumento de ausencia de arbitraje.....	26
1.6 Solución de la ecuación diferencial parcial fundamental para el precio de los bonos cupón cero.....	28
Capítulo 2: Modelos y técnicas por aplicar.....	32
2.1 La representación de espacio de estados.....	32
2.1.1 Vector de observaciones y vector de estado.....	32
2.1.2 Condiciones iniciales para espacio de estados invariante en el tiempo y vector de estado estacionario.....	37
2.2 Filtro de Kalman.....	39
2.2.1 Las ecuaciones del filtro de Kalman.....	39
2.2.2 Deducción de las ecuaciones del filtro de Kalman.....	42
2.2.3 Función de verosimilitud del vector de observaciones en forma de descomposición del error de pronóstico.....	45
2.3 Maximización de la función de verosimilitud para estimar los hiperparámetros de las matrices del sistema.....	47
2.3.1 El vector de puntaje.....	49
2.3.2 La matriz hessiana y la matriz de información.....	51
2.3.3 Primeras derivadas del vector de innovación y su matriz de covarianzas respecto cada hiperparámetro.....	54
2.3.4 Segundas derivadas del vector de innovación y su matriz de covarianzas.....	56
Capítulo 3: Un espacio de estados para el modelo de Vasicek (1977) de un factor y estimación de sus hiperparámetros.....	59
3.1 Matrices del sistema.....	59
3.2 Condiciones iniciales del espacio de estados.....	62
3.3 Primeras derivadas de las matrices del sistema respecto el i-ésimo hiperparámetro.....	64
3.4 Primeras derivadas del vector de estado y covarianzas, adelantados y actualizados respecto el i-ésimo hiperparámetro.....	66
3.5 Primeras derivadas del error de pronóstico y su matriz de covarianzas respecto el i-ésimo hiperparámetro.....	68
3.6 Segundas derivadas de las matrices del sistema.....	69

3.7 Segundas derivadas del vector de estado y covarianzas, adelantados y actualizados.....	71
3.8 Segundas derivadas del error de pronóstico y su matriz de covarianzas.....	80
Capítulo 4: Resultados.....	83
Apéndices.....	91
Apéndice A: Procesos estocásticos estacionarios.....	91
Apéndice B: Otros conceptos de procesos estocásticos.	92
Apéndice C: Correspondencia del filtro de Kalman con el método de mínimos cuadrados recursivos.....	95
Apéndice D: Inversión de matrices mediante la descomposición de Cholesky...	98
Apéndice E: Métodos de optimización para funciones diferenciables no lineales de varias variables en problemas no restringidos.....	101
Referencias.....	107

Introducción

La estructura temporal de tipos de interés, ETTI, o estructura de plazos de tasas de interés es la disposición de la tasa cero o rendimiento de los bonos cupón cero en un lapso de tiempo determinado, para diferentes plazos de inversión o financiación en un determinado mercado. Las dinámicas de una ETTI no solamente deben ser realistas sino que también deben impedir condiciones bajo las cuales los participantes del mercado puedan obtener beneficios libres de riesgo adicional. La ETTI es la referencia imprescindible para la inclusión del componente de mercado en un análisis económico-financiero y tiene importantes aplicaciones: Determinación de precios de activos que dependen de las tasas de interés, análisis de estrategias de financiamiento para gobiernos o empresas, cálculo del valor en riesgo de un portafolio de instrumentos de renta fija, cálculo de primas y reservas de seguros, estudios de la formación de precios y estudios de la política monetaria.

Modelar una ETTI requiere especificar un proceso estocástico para cada uno de los factores o variables de estado que conducen sus dinámicas. Las *variables de estado* o *vectores de estado* son variables escalares o vectoriales que pueden no ser observables directamente. La mayoría de los modelos de ETTI explican la variabilidad en los rendimientos de los bonos usando como factor o variable de estado única la tasa de interés instantánea o de corto plazo. Tal es el caso del modelo de Vasicek (1977). Su factor único depende de cuatro parámetros constantes en el tiempo: La media, la desviación estándar, la velocidad de ajuste y el precio de riesgo de mercado.

Un proceso representado por una ecuación diferencial estocástica está definido en tiempo continuo pero los datos observados son muestreados en intervalos de tiempo discreto. Típicamente el método de estimación de parámetros elegido es la máxima verosimilitud, que requiere conocer la función de densidad del vector de estado. En el modelo de Vasicek la tasa de interés de corto plazo es un proceso estocástico con función de densidad gaussiana, lo cual es muy favorable analíticamente.

Un espacio de estados es una representación algebraica de una serie de tiempo que consiste básicamente en un sistema de dos ecuaciones recursivas: El vector de observaciones y el vector de estados. El vector de observaciones está en función del vector de estado. Los valores que toma el vector de estado son desconocidos pero conocemos las salidas del vector de observaciones. El *filtro de Kalman* es un conjunto de ecuaciones recursivas que, en palabras burdas, sirven para calcular los valores del vector de estado a partir de las salidas, conocidas, del vector de observaciones.

Las ecuaciones de un espacio de estados están en función de las *matrices del sistema*. Las matrices del sistema están en función de un conjunto de parámetros llamados *hiperparámetros*. En cada punto de la serie de tiempo, el filtro de Kalman, paralelamente a que proporciona los valores del *vector de estado adelantado* y el *vector de estado actualizado* y sus respectivas matrices de covarianzas, arroja los valores del *error de pronóstico* o *vector de innovación* y su matriz de covarianzas.

Los vectores de innovación y sus matrices de covarianzas sirven para construir una función de verosimilitud del vector de observaciones en forma de descomposición del error de pronóstico. Tal función de verosimilitud puede resolverse aplicando algún

método de optimización preferido. Solucionar esa función de verosimilitud significa hallar el valor de los hiperparámetros.

Esta tesis presenta un espacio de estados específico. Resolviendo la ecuación diferencial estocástica del proceso de la tasa de interés de corto plazo del modelo de Vasicek se llega a una expresión que se usa como vector de estado. Por su parte, aplicando la técnica de valuación de bonos cupón cero neutral al riesgo o bien la ecuación diferencial parcial fundamental para el precio de los bonos cupón cero es posible obtener una expresión en la cual se aprecia la curva de tasas cero como una función lineal de la tasa de interés instantánea de Vasicek, es decir, se utiliza la ecuación de rendimientos de los bonos cupón cero como vector de observaciones. Los hiperparámetros del espacio de estados antedicho son los cuatro parámetros del modelo de ETTI de Vasicek. Se aplica el filtro de Kalman para obtener la función de verosimilitud que posteriormente se resuelve aplicando cualquier método de optimización para funciones diferenciables no lineales de varias variables en problemas no restringidos. A continuación explicamos la forma en que está estructurado este estudio.

El capítulo 1 expone lo referente a las ETTIs y el modelo de Vasicek. El tema 1.1 presenta la fórmula de Feynman y Kac y otros conceptos básicos útiles para después. El tema 1.2 habla de las ETTIs. En 1.3 se presenta el modelo de Vasicek antedicho. Partiendo de dicho modelo, en 1.4 se deducen las ecuaciones para el precio de cualquier bono, permitiendo así encontrar toda la ETTI. En 1.5 se obtiene por dos caminos distintos pero complementarios, la ecuación diferencial parcial fundamental para el precio de los bonos cupón cero y en 1.6 la resolvemos, llegando a los mismos resultados que se lograron en 1.4.

El capítulo 2 toca lo referente a espacio de estados y filtro de Kalman. En 2.1 definiremos e ilustraremos los espacios de estados. Señalaremos cómo calcular las condiciones iniciales para un espacio de estados invariante en el tiempo y vector de estado estacionario. En 2.2 deduciremos las ecuaciones recursivas del filtro de Kalman y construiremos la función de verosimilitud del vector de observaciones en forma de descomposición del error de pronóstico. En 2.3 calcularemos el vector de puntaje y la matriz hessiana, dos elementos básicos para aplicar varios métodos numéricos de optimización posibles. Para obtener el vector de puntaje y la matriz hessiana se vuelve necesario disponer de las primeras y segundas derivadas del vector de innovación y su matriz de covarianzas, respecto cada hiperparámetro. Las derivadas anteriores requieren conocer las primeras y segundas derivadas del vector de estado y su matriz de covarianzas, adelantados y actualizados, respecto cada hiperparámetro. Todas estas derivadas se calculan analíticamente en esta misma sección. Se incluyen las derivadas del vector de estado actualizado inicial y su matriz de covarianzas, o sea, las derivadas de las condiciones iniciales.

En el capítulo 3 aplicamos todo lo del capítulo 2 en lo del capítulo 1, de la siguiente manera. En 3.1 mostramos una representación de espacio de estados invariante en el tiempo para el modelo de ETTI de Vasicek. Definiremos cada una de las matrices del sistema en función de cuatro hiperparámetros, que vienen siendo los cuatro parámetros de la tasa de interés de corto plazo o factor único del modelo de Vasicek mencionados anteriormente. Los datos que usaremos como salidas conocidas del vector de

observaciones serán 1377 rendimientos de CETES¹ capturados directamente de los registros del mercado de valores mexicano del 16/05/2003 al 21/10/2008 para cuatro diferentes tiempos al vencimiento: 28, 91, 182 y 364 días. Se usarán solamente cuatro diferentes tiempos al vencimiento porque esos son los datos de que se dispone, pero no tiene relación con el hecho de tener que calcular cuatro hiperparámetros. En 3.2 calculamos las condiciones iniciales del espacio de estados que tenemos, tal como se señaló en 2.1. En lo que resta del capítulo 3 presentamos las primeras y segundas derivadas respecto cada hiperparámetro, para las matrices del sistema que tenemos, así como para los vectores de estado y sus matrices de covarianzas, adelantados y actualizados, el error de pronóstico y su matriz de covarianzas. Se incluyen las derivadas de las condiciones iniciales.

En el capítulo 4 presentamos como resultado la construcción de la estructura de temporalidad de los tipos de interés o curva de tasas de rendimiento de bonos cupón cero con tiempos al vencimiento entre 28 y 364 días, de México, del 16/05/2003 al 21/10/2008 y se concluye. Al final el lector encontrará valiosos apéndices que complementan algunos puntos pero puede omitirlos si no los considera necesarios.

¹ Los Certificados de la Tesorería de la Federación o CETES son bonos de corto plazo emitidos por el gobierno federal mexicano.

Capítulo 1: Estructura temporal de tipos de interés de Vasicek (1977) de un factor

1.1 Tasas instantáneas, precio y rendimiento de bonos cupón cero bajo la medida de probabilidad neutral al riesgo

Veamos las relaciones existentes entre el precio de un bono de descuento puro libre de riesgo de incumplimiento¹ o bono cupón cero, su tasa de rendimiento o tasa cero y la tasa de interés de corto plazo o tasa de interés instantánea.

Sea $P(t, T)$ el precio de un bono que en el tiempo t promete entregar una unidad monetaria en el tiempo futuro T , igual al valor esperado del flujo descontado. La esperanza estará condicionada al valor actual de la *tasa de interés de corto plazo o tasa de interés instantánea*, r_t , dado que dicha tasa es aleatoria y desconocemos sus valores futuros.

$$P(t, T) = E_t \left(\exp \left(- \int_t^T r_v dv \right) \right)$$

Cuanto mayor es la tasa de corto plazo menor es el precio del bono. Consecuentemente, si un bono ofrece una tasa de corto plazo, r^* , más grande que la tasa libre de riesgo, r , su tenedor puede cometer arbitraje² si vende un instrumento igual pero ofrece solamente la tasa libre de riesgo,³ obteniendo así el beneficio extra $P(t, T) - P^*(t, T)$, sin haberse expuesto a riesgo adicional.

Una forma de modelar la ausencia de oportunidades de arbitraje⁴ es que el precio de un bono que en el tiempo t promete entregar una unidad monetaria en el tiempo futuro T , sea igual al valor esperado del flujo descontado, menos el premio por riesgo que exige el mercado:⁵

¹ También suele decirse solamente *libre de riesgo*.

² Entendamos por *arbitraje* la operación de compra o venta de valores negociables, haciendo simultáneamente la operación contraria en otro mercado de valores, con la finalidad de aprovechar la diferencia entre los precios de ese mismo activo. Las discrepancias de precios son fruto de la volatilidad de los activos, demora en las comunicaciones o en la respuesta de cada mercado a la misma información. Se diferencia de la especulación en que en ésta, la compraventa está separada por el tiempo.

³ David Jamieson B., *Affine Term Structure Models: Theory and Implementation* (Canadá: Bank of Canada, 2001), p. 12.

⁴ Augusto Castillo R. y Alejandro Valenzuela D., "Un modelo de valoración de bonos con rescate anticipado," *Estudios de Administración*, vol. 12, num. 1, 2005, p. 32.

⁵ Riesgo de mercado: Riesgo de una cartera de valores de renta fija relacionado con la variación de los tipos de interés del mercado. También se le dice riesgo sistemático pues es el riesgo inherente e irrefutable del mercado; no puede ser eliminado mediante la diversificación. En contraste, el riesgo no sistemático es el riesgo específico de una empresa o sector; este riesgo se puede eliminar de una cartera si ésta se diversifica.

$$P(t, T) = E_t \left(\exp \left(- \int_t^T r_v dv \right) \right) - \pi(t, T, r)$$

Por otra parte, también existe el *método de valuación neutral al riesgo*. Pese al nombre que tiene, no supone neutralidad al riesgo en beneficio de los agentes económicos. Calcula el precio del bono como la esperanza del flujo descontado empleando la medida de distribución de probabilidad correspondiente al proceso estocástico de la tasa de interés de corto plazo ajustado por riesgo. A la medida de distribución de probabilidad ajustada por riesgo, Q , se le llama neutral al riesgo. Bajo la medida neutral al riesgo Q , el precio de un bono cupón cero emitido en t y que paga una unidad monetaria en T , dado que disponemos de información relevante hasta el tiempo t es:

$$P(t, T) = E_t^Q \left(\exp \left(- \int_t^T r_v dv \right) \right) \quad (1.1.1)$$

La *ecuación de valuación neutral al riesgo*, (1.1.1) también es conocida como la *fórmula de Feynman y Kac*.

Ahora observemos esta otra fórmula para calcular en el momento t el precio de los bonos cupón cero:

$$P(t, T) = \exp(-R(t, T, r)(T - t))$$

El término $R(t, T, r)$ es el rendimiento del bono cupón cero durante su período de vida restante y podemos verlo como la tasa de interés promedio del tiempo t al T .¹ $R(t, T, r)$ también se llama tasa cero; despejándola de la ecuación anterior obtenemos en el momento t la estructura de las tasas cero o curva de rendimientos del bono cupón cero:

$$R(t, T, r) = - \frac{1}{(T - t)} \ln P(t, T) \quad (1.1.2)$$

Por (1.1.1) y (1.1.2) vemos que el desarrollo de la tasa cero depende de la marcha de la tasa de interés de corto plazo:

$$R(t, T, r) = - \frac{1}{(T - t)} \ln E_t^Q \left(\exp \left(- \int_t^T r_v dv \right) \right)$$

De esto podemos afirmar que toda la estructura de tasas cero podrá determinarse por el proceso de tasas de interés de corto plazo ajustado por riesgo.²

¹ Dervis Bayazit, "Yield curve estimation and prediction with Vasicek model" (Tesis de licenciatura, Middle East Technical University, Estados Unidos, 2004), p. 33.

² Castillo R., "Un modelo de valoración de bonos con rescate anticipado," p. 33.

1.2 Modelado de la estructura temporal de tipos de interés

Modelar la estructura temporal de tipos de interés consiste en construir la relación funcional que informa, para un período de tiempo y mercado dados, cuáles son las tasas de rendimiento de bonos cupón cero para diversos plazos. Esto se traduce en que, dado un conjunto de observaciones que generalmente son precios de bonos cupón cero o rendimientos de los mismos, podamos construir una gráfica tridimensional en la cual:

- El eje Z indica la tasa cero o rendimiento de los bonos cupón cero.
- El eje X muestra diferentes plazos τ de inversión.
- El eje Y señala los días de negociación entre dos fechas dadas en un determinado mercado.

Las construcciones de los planos XZ y YZ son dos tareas distintas pero relacionadas. El plano XZ muestra una curva de rendimientos de los bonos cupón cero para diferentes vencimientos, en un momento t específico. Es generalmente llamada *curva temporal de tipos de interés*, *curva de tasas cero*, *curva cupón cero* o incluso *estructura temporal de tipos de interés*, pero es más bien como una fotografía de la ETTI completa, en un instante dado. El plano YZ especifica las dinámicas intertemporales de la ETTI: Determina la evolución de los rendimientos de los bonos cupón cero al paso del tiempo. El plano XY es una cuadrícula que muestra diversos plazos al vencimiento en cada fecha con actividad bursátil.

En contraste con el modelado de otras variables aleatorias económicas-financieras como por ejemplo, tipos de cambio, precios de activos, tasas de crecimiento económico, inflación, Etc. que son procesos estocásticos escalares que exhiben características como media, volatilidad, reversión a la media, Etc., modelar las tasas de interés es más complicado pues se añade la dimensión del tiempo.¹

Las dinámicas de una ETTI no solamente deben ser razonables desde un punto de vista empírico sino que también deben impedir las condiciones bajo las cuales los participantes del mercado produzcan beneficios libres de riesgo adicional. Esto es, además de ser realista, un modelo de ETTI debe ser libre de arbitraje.

Algunas características que puede tener la curva cupón cero

- **Creciente:** Mayor a medida que el plazo al que se refiere se va ampliando. Es normal, ya que a mayor plazo el riesgo también es mayor.
- **Decreciente:** Los tipos de interés a corto plazo son más elevados que los de largo plazo. Es una situación especial cuando el mercado apuesta en el corto y mediano plazo por un descenso de los tipos de interés. Es más común mientras mayores son los tipos de interés que se negocian en el mercado.
- **Plana:** Los tipos de interés tienen muy pocas diferencias; es una situación anómala e inestable.

¹ Jamieson B., *Affine Term Structure Models: Theory and Implementation*, p. 2.

- Oscilante o con jorobas: Se presentan continuas variaciones cuando en el mercado hay inestabilidad e incertidumbre.

Aplicaciones del modelado de la estructura temporal de tipos de interés

La ETTI representa en cada momento las alternativas básicas de inversión-financiación que existen para diferentes plazos en un determinado mercado y es por lo tanto la referencia a la cual obligatoriamente hay que recurrir cada vez que se quiere introducir el componente de mercado en un determinado análisis. Algunas de las principales aplicaciones son en:

1. Determinación de precios de derechos que dependen de las tasas de interés: Techos, pisos, opciones de swaps y bonos reclamables.
2. Análisis de estrategias de financiamiento; por ejemplo, si conviene a un gobierno o a una empresa emitir deuda de corto o de largo plazo.
3. Administración del riesgo: Cálculo del *valor en riesgo*¹ de un portafolio de instrumentos de renta fija.
4. Cálculo de primas y reservas de seguros.
5. Estudios del nexo pronosticador entre la ETTI y la inflación o la actividad económica real –formación de precios-.
6. Estudios de la relación entre la ETTI y la política monetaria.

Número y tipo de factores del modelo

Construir un modelo de estructura de plazos de tipos de interés requiere especificar un proceso estocástico para cada uno de los factores o variables de estado que conducen sus dinámicas. Las *variables de estado* o *vectores de estado* son variables escalares o vectoriales que pueden no ser observables directamente. La mayoría de los modelos de ETTI utilizan un solo factor o variable de estado para explicar la variabilidad en los rendimientos de los bonos. Dicha variable única es ordinariamente la tasa de interés instantánea. La Tabla 1.2.1 muestra numerosos ejemplos de ello.² Un caso de dos factores es el modelo de Longstaff y Schwartz (1992), que sugiere como primer factor la misma tasa de interés de corto plazo de CIR (1985) y como segundo factor la varianza de los cambios en dicha tasa. La Tabla 1.2.2, muestra ejemplos de modelos de dos o tres factores.³ Litterman y Scheinkman (1991) propusieron como factores el nivel, la pendiente y la curvatura de dicha tasa pero encontraron que casi 90 por ciento de la variación en las tasas del Tesoro de Estados Unidos corresponde al primer factor.⁴ La literatura sugiere que tres factores son adecuados para explicar entre el 90 y 99% de la variabilidad en los rendimientos de los bonos. El factor más importante es el nivel de la tasa de interés de corto plazo debido a que representa a menudo entre el 80 y 90% de la varianza total.

¹ El valor en riesgo se define como la pérdida máxima esperada en el valor de un portafolio con una probabilidad definida sobre un horizonte de tiempo predeterminado.

² Bayazit, “Yield curve estimation and prediction with Vasicek model,” p. 2.

³ Mercedes Galisteo R., “Dinámica de la estructura temporal de tipos de interés: Modelo de tres factores” (Tesis doctoral, Universidad de Barcelona, España, 2002), pp. 92-93.

⁴ Somnath Chatterjee, *Application of the Kalman filter for estimating continuous time term structure models: The case of UK and Germany* (Reino Unido: University of Glasgow, 2005), p. 4.

Tabla 1.2.1: Modelos de la estructura temporal de tipos de interés que usan como factor único la tasa de interés de corto plazo

Modelo:	Proceso estocástico del factor tasa de interés instantánea:
Merton (1973)	$dr_t = \alpha dt + \sigma dW_t$
Vasicek (1977)	$dr_t = (\beta + \alpha r_t) dt + \sigma dW_t$
Dothan (1978)	$dr_t = \sigma r_t dW_t$
Brennan / Schwartz (1980)	$dr_t = (\beta + \alpha r_t) dt + \sigma r_t dW_t$
Movimiento Browniano Geométrico	$dr_t = \alpha r_t dt + \sigma r_t dW_t$
Marsh / Rosenfeld (1983)	$dr_t = (\beta r_t^{-(1-\gamma)} + \alpha r_t) dt + \sigma r_t^{\frac{\gamma}{2}} dW_t$
Constatinides / Ingersoll (1984)	$dr_t = \sigma r_t^{\frac{3}{2}} dW_t$
Cox / Ingersoll / Ross (CIR) (1985)	$dr_t = (\beta + \alpha r_t) dt + \sigma \sqrt{r_t} dW_t$
Ho / Lee (1986)	$dr_t = \theta_t dt + \sigma dW_t$
Vasicek Exponencial	$dr_t = r_t (\eta_t - a \ln r_t) dt + \sigma r_t dW_t$
Hull / White / Vasicek Extendido (1990)	$dr_t = (\beta_t + \alpha_t r_t) dt + \sigma_t dW_t$
Hull / White CIR Extendido (1990)	$dr_t = (\beta_t + \alpha_t r_t) dt + \sigma_t \sqrt{r_t} dW_t$
Black / Derman / Toy (1990)	$d(\ln r_t) = \left(\theta_t + \frac{\sigma'_t}{\sigma_t} \ln r_t \right) dt + \sigma_t dW_t$
Black / Karazinsky (1991)	$d(\ln r_t) = \phi_t (\ln \mu_t - \ln r_t) dt + \sigma_t dW_t$

Tabla 1.2.2: Modelos de la estructura temporal de tipos de interés de dos o tres factores

Modelo:	Procesos estocásticos de los factores:
Richard (1978)	$dR = -a(R - R^*) dt + \sigma_R R^{\frac{1}{2}} dz_R$ $d\pi = -c(\pi - \pi^*) dt + \sigma_\pi \pi^{\frac{1}{2}} dz_\pi$
Brennan / Schwartz (1979)	$dr = r \left(\alpha \ln \left(\frac{l}{pr} \right) + \left(\frac{1}{2} \right) \sigma_1^2 \right) dt + \sigma_1 r dz_r$ $dl = l (l - r + \sigma_2^2 + \lambda_2 \sigma_2) dt + l \sigma_2 dz_2$
Boyle (1980)	$dR = \alpha_R (\gamma_R - R) dt + \sigma_R dz_R$ $dj = \alpha_j (\gamma_j - j) dt + \sigma_j dz_j$
Schaefer / Schwartz (1984)	$ds = m(\mu - s) dt + \gamma dz_1$ $dl = \beta_2(s, l, t) dt + \sigma l^{\frac{1}{2}} dz_2$
Longstaff / Schwartz (1992)	$dX = (a - b X) dt + c X^{\frac{1}{2}} dz_2$ $dY = (d - e Y) dt + f Y^{\frac{1}{2}} dz_3$

Moreno (1997)	$ds = K_1 (\mu_1 - s) dt + \sigma_1 dz_1$ $dl = k_2 (\mu_2 - l) dt + \sigma_2 dz_2$
Kraus / Smith (1993)	$dr = \mu dt + \sigma dz_1$ $d\mu = m dt + s dz_2$ $d\alpha = dm - d\sigma_2 = b dt + v dz_3$
Chen (1995)	$dr = k(\theta - r) dt + \sqrt{\sigma} r^{\frac{1}{2}} dz_1$ $d\theta = v(\bar{\theta} - \theta) dt + \xi \theta^{\frac{1}{2}} dz_2$ $d\sigma = \mu(\bar{\sigma} - \sigma) dt + \eta \sigma^{\frac{1}{2}} dz_3$

Existe abundante literatura sobre qué modelo explica mejor el comportamiento de los tipos de interés. Sin embargo los resultados varían dependiendo del período de observación y del país seleccionado. Es decir, actualmente no existe un consenso en la literatura sobre cuál es el más adecuado para explicar el comportamiento de los tipos de interés.¹

Precio de riesgo de mercado de cada factor de un modelo

La estimación de la estructura temporal de tipos de interés está basada en la premisa de que bonos con diferentes fechas de vencimiento son comerciados en el mismo momento. Bonos con fecha de vencimiento alejada son más riesgosos cuando emplazamos sobre períodos cortos de tiempo y los inversionistas con aversión al riesgo demandan cierta compensación por abordar tal riesgo.² Para anular las oportunidades de que los participantes del mercado puedan obtener ganancias libres de riesgo que existen en estos mercados hay que ajustar las esperanzas de los rendimientos de los bonos de largo plazo por el riesgo de las tasas de interés de corto plazo futuras promedio: Se imponen restricciones sobre el comportamiento de la tasa de interés intertemporal buscando modelar la ausencia de arbitraje.

Con base en el supuesto de ausencia de arbitraje, cada modelo de estructura de plazos de tipos de interés acompaña cada factor propuesto con un determinado *precio de riesgo de mercado*; la Tabla 1.2.3 nos muestra varios ejemplos.³

Tabla 1.2.3: Precio de riesgo de mercado de algunos modelos de estructura temporal de tipos de interés

Modelo.	Precio de riesgo de mercado.
Merton (1973)	Para el factor único: λ
Vasicek (1977)	Para el factor único: λ
Dothan (1978)	Para el factor único: λ
Constatinides / Ingersoll (1984)	Para el factor único: λ

¹ María de L. Gómez del V., “Nuevos planteamientos en modelos unifactoriales de la estructura temporal de los tipos de interés” (Tesis de doctorado, Universidad de Valladolid, España, 2004), p. 210.

² Chatterjee, *Application of the Kalman filter for estimating continuous time term structure models: The case of UK and Germany*, p. 2.

³ Galisteo R., “Dinámica de la estructura temporal de tipos de interés: Modelo de tres factores,” p. 94.

Cox / Ingersoll / Ross (CIR) (1985)	Para el factor único:	$\frac{\lambda\sqrt{r}}{\sigma}$
Richard (1978)	Para el primer factor:	$\lambda_R R^{\frac{1}{2}}$
	Para el segundo factor:	$\lambda_\pi \pi^{\frac{1}{2}}$
Brennan / Schwartz (1979)	Para el primer factor:	λ
	Para el segundo factor:	Eliminado
Boyle (1980)	Para el primer factor:	λ
	Para el segundo factor:	λ
Schaefer / Schwartz (1984)	Para el primer factor:	λ
	Para el segundo factor:	Eliminado
Longstaff / Schwartz (1992)	Para el primer factor:	0
	Para el segundo factor:	λ
Moreno (1997)	Para el primer factor:	$a + bs(t)$
	Para el segundo factor:	$c + dl(t)$
Kraus / Smith (1993)	Para el primer factor:	λ
	Para el segundo factor:	λ
	Para el tercer factor:	λ
Chen (1995)	Para el primer factor:	λ
	Para el segundo factor:	λ
	Para el tercer factor:	λ

Como vemos, λ puede depender de r o t pero en ningún caso del período de vencimiento de los bonos utilizados para su estimación. Esto se debe a que de otro modo se introducirían oportunidades de arbitraje en el modelo.¹

El precio de riesgo de mercado no es un precio en el sentido estricto de la palabra. Determina el exceso de rendimiento que exige un inversor por aceptar una unidad adicional de riesgo asociado a variaciones no anticipadas de la(s) variable(s) de estado del modelo, en términos relativos.² El proceso que describe el comportamiento del precio de los bonos en un mercado depende de las relaciones entre la oferta y la demanda de ese mercado, determinadas por las preferencias de los inversores. Agentes adversos al riesgo dan lugar a precios del riesgo de mercado negativos, inversores propensos al riesgo dan lugar a precios del riesgo de mercado positivos e inversores neutrales al riesgo dan lugar a precios del riesgo de mercado nulos. El precio de riesgo de mercado no es observable y debe ser determinado a partir de la información disponible en el mercado.

Supuestos para los modelos de un factor

Los supuestos habituales para la forma matemática común de los modelos de estructura temporal de tipos de interés de un factor son:

¹ Gómez del V., "Nuevos planteamientos en modelos unifactoriales de la estructura temporal de los tipos de interés," p. 79.

² *Ibíd.*, p. 111.

- i. Son válidos los supuestos económicos estándar para un modelo continuo: Es posible el comercio continuo en los mercados de bonos, no hay impuestos ni costos de transacción, no hay restricciones para las ventas en corto ni problemas de divisibilidad y los inversionistas prefieren mejorar su utilidad marginal en todos los niveles.
- ii. La dinámica de un proceso de tasa de interés instantánea r_t puede describirse por la ecuación diferencial estocástica de primer orden:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dW_t$$

donde $\mu(r_t)dt$ es un componente determinístico y $\sigma(r_t)dW$ aleatorio. La propiedad de poder dividir en estas dos partes los cambios en la tasa en cuestión, dr_t en nuestro caso, es muy común en los modelos de tasas de interés. El factor $\mu(r_t)$ describe la tendencia del proceso, $\sigma(r_t)$ su desviación estándar y W_t es un movimiento browniano¹ estándar² que sirve para simular ruido.

- iii. El precio de riesgo de mercado, $\lambda(\cdot)$ sólo depende de la tasa de interés instantánea r_t .
- iv. Como consecuencia de los dos supuestos anteriores, $\forall T > t$, los precios de los bonos $P(t, T)$ son funciones de t , T y la variable de estado única, la tasa de interés instantánea r_t . Esto implica que los cambios en la curva de rendimientos del bono cupón cero a diferentes fechas de vencimiento están perfectamente correlacionados. En los modelos multifactoriales esto tiene varias restricciones.

Modelos dinámicos y modelos binomiales

La Tabla 1.2.1 maneja una notación un poquito más completa que la de la Tabla 1.2.2; particularmente podemos ver en ella que algunas variables y parámetros del factor único tienen un subíndice t , lo cual indica dependencia del tiempo. Específicamente, observe que ninguno de los parámetros de los primeros ocho ejemplos de proceso estocástico para la tasa de interés instantánea en la Tabla 1.2.1 dependen del tiempo; de tal forma, corresponden a los llamados modelos *dinámicos* de estructura temporal de tipos de interés. En contraste, los últimos seis ejemplos de la Tabla 1.2.1 son factores de modelos *binomiales*. En otras palabras, en los modelos dinámicos los parámetros de los procesos estocásticos que describen a cada factor son constantes; en los modelos binomiales uno o varios de los parámetros de los factores se actualizan con el tiempo. Actualmente también se les dice modelos de equilibrio o modelos clásicos a los modelos dinámicos y modelos libres de arbitraje a los modelos binomiales.

Modelos afines

Existe una clase general de modelos de estructura temporal de tipos de interés multifactoriales llamados afines³ debido a que se basan en la suposición de que los

¹ Puede consultar el apéndice B.

² Algunos sugieren el movimiento browniano fraccional.

³ Se le debe a Darrell Duffie y Rui Kan, "A yield factor model of interest rates," *Mathematical Finance*, vol. 6, num. 4, 1996, pp. 379 - 406.

rendimientos de los bonos son una función lineal de las variables de estado subyacentes que proveen incertidumbre al modelo.¹ Ejemplos de modelos afines son Vasicek (1977), CIR (1985) y Longstaff y Schwartz (1992), entre muchos otros.

La fórmula del precio de un bono de descuento puro, $P(t, T)$, en un modelo de estructura temporal de tipos de interés afín puede tomar la forma:

$$P(t, T) = \exp(F(\tau)X + G(\tau))$$

X es el vector de estado o arreglo vectorial conformado por todos los factores. $F(\tau)$ y $G(\tau)$ son funciones del tiempo $\tau = T - t$ que falta para el vencimiento del contrato, los parámetros de los factores y el precio de riesgo de mercado del modelo. El conjunto de precios de bonos cupón cero como una función del tiempo que falta para el vencimiento del contrato, $\tau = T - t$, definirá la curva de rendimientos de los bonos cupón cero:

$$R(t, T) = -\frac{1}{\tau} \ln P(t, T) = \frac{F(\tau)X + G(\tau)}{\tau}$$

La propiedad de los rendimientos de la clase afín puede verse claramente en la ecuación anterior: Los rendimientos de los bonos cupón cero son funciones afines de los factores subyacentes.

Para modelos donde tanto la tendencia como la volatilidad son afines en r , es posible obtener expresiones concretas para $F(\tau)$ y $G(\tau)$. Tanto el modelo Vasicek (1977) como el CIR (1985) cumplen dicho criterio resultando en soluciones con una forma específica para los precios de los bonos de descuento puro.²

Cálculo de los parámetros de los factores de un modelo

Un proceso representado por una ecuación diferencial estocástica está definido en tiempo continuo pero los datos observados son muestreados en intervalos de tiempo discreto. Típicamente el método de estimación de parámetros elegido es la máxima verosimilitud, pese a tener la desventaja de que la función de densidad del vector de estado es generalmente desconocida y debe aproximarse.

1.3 Tasa de interés de corto plazo del modelo de Vasicek (1977)

En el **modelo de Vasicek** se asume que el proceso de la tasa de interés de corto plazo r_t satisface la siguiente ecuación diferencial estocástica:

¹ Una función $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ es afín si tiene la forma $f(\mathbf{x}) = \mathbf{a}'\mathbf{x} + b$ con $\mathbf{a}' \in \mathfrak{R}^n$, $b \in \mathfrak{R}$ y $\mathbf{x} \in \mathfrak{R}^n$.

² Chatterjee, *Application of the Kalman filter for estimating continuous time term structure models: The case of UK and Germany*, p. 7.

$$dr_t = (\beta + \alpha r_t) dt + \sigma dW_t$$

o bien:

$$dr_t = -\alpha \left(-\frac{\beta}{\alpha} - r_t \right) dt + \sigma dW_t$$

donde $\alpha < 0$, $\sigma > 0$, $\beta \geq 0$ y W_t es un movimiento browniano.

Haciendo en la expresión previa $\kappa = -\alpha$ y $\mu = -\frac{\beta}{\alpha}$, obtenemos la ecuación diferencial estocástica:

$$dr_t = \kappa(\mu - r_t) dt + \sigma dW_t \quad (1.3.1)$$

donde κ , μ y σ son constantes reales no negativas, r_t es el nivel vigente de la tasa de interés y W_t un movimiento browniano.

La ecuación anterior es un proceso gaussiano con reversión a la media μ , a la velocidad de ajuste κ . El parámetro μ es la tasa de interés normal de largo plazo. El modelo exhibe “reversión a la media” porque si la tasa de interés está sobre la media de largo plazo, $r_t > \mu$, entonces la tendencia se vuelve negativa para que la tasa sea acercada a la μ promedio. Análogamente, si la tasa es menor que la media de largo plazo, $r_t < \mu$, entonces la tendencia permanece positiva para que la tasa sea acercada al nivel de μ . El coeficiente $\kappa > 0$ determina la velocidad con que la tasa de interés se acerca a su nivel normal de largo plazo.¹

Según la literatura de las ecuaciones diferenciales estocásticas, la ecuación (1.3.1) es un *proceso de Ornstein-Uhlenbeck* y es el equivalente en tiempo continuo de un proceso autorregresivo de primer orden o modelo $AR(1)$.

Para encontrar la r_t que satisface (1.3.1) hacemos las siguientes manipulaciones algebraicas sucesivas:

$$\frac{dr_t}{dt} = \kappa(\mu - r_t) + \frac{\sigma dW_t}{dt}$$

$$\frac{dr_t}{dt} = \kappa\mu - \kappa r_t + \frac{\sigma dW_t}{dt}$$

¹ El supuesto de reversión a la media concuerda con el fenómeno económico de que la tasa de interés parece ser atraída hacia algún valor promedio de largo plazo: Cuando la tasa de interés crece, la economía decrece y hay menor demanda para los préstamos y una tendencia natural de las tasa a caer; el caso opuesto puede explicarse análogamente. Bayazit, “Yield curve estimation and prediction with Vasicek model,” p. 8.

$$\frac{dr_t}{dt} + \kappa r_t = \kappa \mu + \frac{\sigma dW_t}{dt} \quad (1.3.2)$$

Recordemos que se le llama *ecuación lineal* a una ecuación diferencial de la forma:

$$a_1(x) \frac{dy}{dx} + a_0(x)y = g(x) \quad (1.3.3)$$

Si en (1.3.3) hacemos $y = r_t$, $x = t$, $a_1(x) = 1$, $a_0(x) = \kappa$ y $g(x) = \kappa \mu + \frac{\sigma dW_t}{dt}$, podemos ver que la ecuación (1.3.2), es una *ecuación lineal*. Además es *no homogénea* dado que $g(x) \neq 0$, o mejor dicho, $\kappa \mu + \frac{\sigma dW_t}{dt} \neq 0$.

Por ser equivalente con (1.3.2), la ecuación (1.3.1) es una ecuación lineal no homogénea y por ello aplicaremos el *método de variación de las constantes*.

Haciendo en (1.3.2) que el término análogo con el $g(x)$ de (1.3.3) sea igual con cero obtenemos la *ecuación homogénea*:

$$\frac{dr_t}{dt} + \kappa r_t = 0$$

Obtengamos la solución de la ecuación anterior:

$$\frac{dr_t}{dt} = -\kappa r_t$$

$$dr_t = -\kappa r_t dt$$

$$\frac{dr_t}{r_t} = -\kappa dt$$

$$\int \frac{1}{r_t} dr_t = -\kappa \int dt$$

$$\ln(r_t) = -\kappa t + C_1$$

$$e^{\ln(r_t)} = e^{(-\kappa t + C_1)}$$

$$r_t = e^{-\kappa t + C_1}$$

Por lo tanto, la solución de la ecuación homogénea, o *solución homogénea*, es:

$$r_t = Ce^{-\kappa t}$$

donde $C = e^{C_1}$ es una constante arbitraria.

Busquemos la *solución particular* reemplazando en la solución homogénea la constante C por una función desconocida $\psi(t)$, como sigue:

$$r_{t_1} = \psi(t)e^{-\kappa t} \quad (1.3.4)$$

Luego, debe cumplirse que la derivada de la ecuación anterior con respecto a t , sea igual con el término de (1.3.2) que es análogo con el $g(x)$ de (1.3.3), esto es:

$$\frac{d\psi(t)e^{-\kappa t}}{dt} = \kappa\mu + \frac{\sigma dW_t}{dt}$$

De lo anterior sigue:

$$\begin{aligned} d\psi(t)e^{-\kappa t} &= \kappa\mu dt + \sigma dW_t \\ d\psi(t) &= \kappa\mu e^{\kappa t} dt + \sigma e^{\kappa t} dW_t \\ \int d\psi(t) &= \int \kappa\mu e^{\kappa t} dt + \int \sigma e^{\kappa t} dW_t \\ \int d\psi(t) &= \mu \int \kappa e^{\kappa t} dt + \sigma \int e^{\kappa t} dW_t \\ \psi(t) &= \mu e^{\kappa t} + \sigma \int_0^t e^{\kappa v} dW_v \end{aligned}$$

Sustituyendo el resultado anterior en (1.3.4) obtenemos la *solución particular*:

$$r_{t_1} = \mu + e^{-\kappa t} \sigma \int_0^t e^{\kappa v} dW_v$$

Solución general es la suma de la solución homogénea más la solución particular:

$$r_t = Ce^{-\kappa t} + \mu + \sigma e^{-\kappa t} \int_0^t e^{\kappa v} dW_v$$

Al sustituir en la expresión anterior la condición inicial $r_0 = r_s$, obtenemos $C = r_s - \mu$ y por lo tanto la solución de (1.3.1), es:

$$\begin{aligned} r_t &= (r_s - \mu)e^{-\kappa(t-s)} + \mu + \sigma \int_s^t e^{-\kappa(t-v)} dW_v \\ &= r_s e^{-\kappa(t-s)} - \mu e^{-\kappa(t-s)} + \mu + \sigma \int_s^t e^{-\kappa(t-v)} dW_v \\ &= r_s e^{-\kappa(t-s)} + \mu(1 - e^{-\kappa(t-s)}) + \sigma \int_s^t e^{-\kappa(t-v)} dW_v \end{aligned} \quad (1.3.5)$$

Finalmente, la solución de la ecuación diferencial estocástica (1.3.1) es (1.3.5).

Debido a que en el modelo de Vasicek la tasa de interés de corto plazo $\{r_t\}$ es un proceso gaussiano, tiene la ventaja de ser muy maleable analíticamente, pero la desventaja de que para cada t hay una probabilidad positiva de que r sea negativa,¹ pese a que la reversión a la media contrarresta dicho problema.

1.4 Curva de rendimientos de Vasicek (1977) aplicando la técnica de valuación neutral al riesgo

La ecuación (1.3.1) no evita las oportunidades de arbitraje en la valuación de bonos. Aplicaremos el teorema de Girsanov² para cambiar su medida de distribución de probabilidad por una llamada neutral al riesgo, bajo la cual los precios de los bonos tendrán la propiedad de martingala³ y satisfarán la condición de ausencia de arbitraje.⁴

Del teorema de Girsanov sabemos que

$$W_t^Q = W_t + \int_0^t \lambda ds = W_t + \lambda t$$

es un movimiento browniano bajo la medida de distribución de probabilidad neutral al riesgo, Q . De la ecuación anterior obtenemos:

$$dW_t = dW_t^Q - \lambda dt$$

Sustituyendo el resultado anterior en (1.3.1) obtenemos la dinámica ajustada por riesgo de la tasa de interés instantánea:

$$\begin{aligned} dr_t &= \kappa(\mu - r_t) dt + \sigma (dW_t^Q - \lambda dt) \\ &= (\kappa(\mu - r_t) - \lambda\sigma) dt + \sigma dW_t^Q \\ &= (\kappa\mu - \kappa r_t - \lambda\sigma) dt + \sigma dW_t^Q \\ &= \kappa \left(\mu - \frac{\lambda\sigma}{\kappa} - r_t \right) dt + \sigma dW_t^Q \end{aligned} \tag{1.4.1}$$

Definimos la media de largo plazo ajustada por riesgo:

$$\mu^Q = \mu - \frac{\lambda\sigma}{\kappa} \tag{1.4.2}$$

¹ Ello es ilógico desde el punto de vista económico: Debido a que la tasa nominal de interés no puede caer bajo cero siempre y cuando las personas puedan poseer dinero en efectivo, ésta puede estancarse en cero por largos períodos, no obstante que los precios caigan persistentemente y substancialmente. Bayazit, "Yield curve estimation and prediction with Vasicek model," p. 10.

² Puede consultar el apéndice B.

³ *Ibíd.*

⁴ Bayazit, *op. cit.*, p. 27.

Sustituyendo la expresión anterior en (1.4.1) obtenemos la forma más común del proceso estocástico neutral al riesgo para la tasa de interés de corto plazo:

$$d r_t = \kappa \left(\mu^Q - r_t \right) dt + \sigma dW_t^Q \quad (1.4.3)$$

Observe que la ecuación anterior es la misma ecuación que (1.3.1), pero con medida de distribución de probabilidad neutral al riesgo. De la misma manera en que obtuvimos la solución de (1.3.1), que es (1.3.5), se logra la solución de (1.4.3):¹

$$r_t = e^{-\kappa(t-s)} r_s + \mu^Q \left(1 - e^{-\kappa(t-s)} \right) + \sigma \int_s^t e^{-\kappa(t-v)} dW_v^Q \quad (1.4.4)$$

con $s \leq t$.

Puesto que la distribución de dW_v^Q es normal con media cero y varianza uno, la integral del extremo derecho de la ecuación anterior también es gaussiana. El único elemento estocástico del cual depende r_t es dicha integral y lo demás es determinístico. Por lo tanto la distribución de r_t condicionada en r_s , bajo la medida de distribución de probabilidad neutral al riesgo también es gaussiana, pero con media y varianza dadas por:²

$$E_s^Q(r_t) = e^{-\kappa(t-s)} r_s + \mu^Q \left(1 - e^{-\kappa(t-s)} \right) \quad (1.4.5)$$

$$\begin{aligned} \text{var}_s^Q(r_t) &= \sigma^2 V \left(\int_s^t e^{-\kappa(t-v)} dW_v^Q \right) \\ &= \sigma^2 E_s^Q \left(\left(\int_s^t e^{-\kappa(t-v)} dW_v^Q \right)^2 \right) \end{aligned}$$

Por la propiedad de isometría³ lo anterior se vuelve:

$$\text{var}_s^Q(r_t) = \sigma^2 \int_s^t e^{-2\kappa(t-v)} dv$$

Es decir:

$$\text{var}_s^Q(r_t) = \frac{\sigma^2}{2\kappa} \left(1 - e^{-2\kappa(t-s)} \right) \quad (1.4.6)$$

Por su parte:

¹ Bayazit, "Yield curve estimation and prediction with Vasicek model," p. 44.

² Ibíd. y Jesper Lund, "Dynamic models of the term structure of interest rates" (Tesis de maestría, Aarhus University, Dinamarca, 1993), pp. 33-34.

³ Puede consultar el apéndice B.

$$\begin{aligned} \text{cov}_s^Q(r_t, r_u) &= \sigma^2 \int_s^{m=\min(t,u)} e^{-\kappa(t-v)} e^{-\kappa(u-v)} dv \\ &= e^{-\kappa|t-u|} \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa(m-s)}) \end{aligned} \quad (1.4.7)$$

La distribución condicional de r_t definida por (1.4.5), (1.4.6) y (1.4.7) es equivalente a un proceso $AR(1)$ ¹ en tiempo discreto y como veremos en la sección 3.1.2 se requiere que $\kappa > 0$ para que el proceso sea estacionario.²

La integral $\int_t^T r_v dv$ también es normalmente distribuida. La función generadora de momentos de una variable aleatoria gaussiana X es:

$$M_X(u) = E(e^{uX}) = \exp\left(u E(X) + \frac{1}{2} u^2 \text{var}(X)\right)$$

Para $u = -1$ tenemos:

$$E(e^{-X}) = \exp\left(-E(X) + \frac{1}{2} \text{var}(X)\right)$$

Aplicando la fórmula anterior en la ecuación (1.1.1) obtenemos:

$$P(t, T) = E_t^Q\left(\exp\left(-\int_t^T r_v dv\right)\right) = \exp\left(-E_t^Q\left(\int_t^T r_v dv\right) + \frac{1}{2} \text{var}_t^Q\left(\int_t^T r_v dv\right)\right) \quad (1.4.8)$$

Calculemos E_t^Q :

$$\begin{aligned} E_t^Q\left(\int_t^T r_v dv\right) &= \int_t^T E_t^Q(r_v) dv \\ &= \int_t^T e^{-\kappa(v-t)} r_t + \mu^Q (1 - e^{-\kappa(v-t)}) dv \\ &= \mu^Q (T - t) + (r_t - \mu^Q) \left(\frac{1 - e^{-\kappa(T-t)}}{\kappa}\right) \end{aligned}$$

Calculemos var_t^Q :

¹ Un proceso $AR(1)$ para la serie $\{Y_t\}$: $Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$ donde $Y_t = r_t$, $Y_{t-1} = r_{t-1}$,

$\phi_1 = e^{-\kappa(t-s)}$, $c = \mu^Q (1 - e^{-\kappa(t-s)})$ y $\varepsilon_t = \sigma \int_s^t e^{-\kappa(t-v)} dW_v^Q$.

² Puede consultar el apéndice A.

$$\begin{aligned}
\text{var}_t^Q \left(\int_t^T r_v dv \right) &= \int_t^T \int_t^T \left(\sigma^2 \int_t^{m=\min(u,w)} e^{-\kappa(u-v)} e^{-\kappa(w-v)} dv \right) du dw \\
&= \sigma^2 \int_t^T \left(\int_v^T \int_v^T (e^{-\kappa(u-v)} e^{-\kappa(w-v)}) dw du \right) dv \\
&= \sigma^2 \int_t^T \left(\int_v^T e^{-\kappa(u-v)} du \right) \left(\int_v^T e^{-\kappa(w-v)} dw \right) dv \\
&= \frac{\sigma^2}{\kappa^2} \int_t^T (e^{-\kappa(T-v)} - 1)^2 dv \\
&= \frac{\sigma^2}{\kappa^2} \int_t^T (e^{-2\kappa(T-v)} - 2e^{-\kappa(T-v)} + 1) dv \\
&= \frac{\sigma^2}{2\kappa^3} (4e^{-\kappa(T-t)} - e^{-2\kappa(T-t)} + 2\kappa(T-t) - 3)
\end{aligned}$$

Sustituyendo los dos resultados anteriores en (1.4.8) se obtiene la expresión para el precio del bono:

$$\begin{aligned}
P(t, T) &= \exp \left(-\mu^Q (T-t) - (r_t - \mu^Q) \left(\frac{1 - e^{-\kappa(T-t)}}{\kappa} \right) \right. \\
&\quad \left. + \frac{\sigma^2}{4\kappa^3} (4e^{-\kappa(T-t)} - e^{-2\kappa(T-t)} + 2\kappa(T-t) - 3) \right)
\end{aligned}$$

Reacomodando términos obtenemos:

$$\begin{aligned}
P(t, T) &= \exp \left(-r_t \left(\frac{1 - e^{-\kappa(T-t)}}{\kappa} \right) - \frac{\sigma^2}{4\kappa^3} (1 - e^{-\kappa(T-t)})^2 \right. \\
&\quad \left. - \left(\mu^Q - \frac{\sigma^2}{2\kappa^2} \right) \left(T-t - \left(\frac{1 - e^{-\kappa(T-t)}}{\kappa} \right) \right) \right)
\end{aligned}$$

En otros términos, la expresión para el precio de cualquier bono de descuento puro que en el período t ofrece entregar una unidad monetaria en el tiempo T es:

$$P(t, T) = \exp(F(t, T)r_t + G(t, T)) \quad (1.4.9)$$

donde

$$F(t, T) = - \left(\frac{1 - e^{-\kappa(T-t)}}{\kappa} \right) \quad (1.4.10)$$

$$G(t,T) = -\frac{\sigma^2(F(t,T))^2}{4\kappa} - \left(\mu^Q - \frac{\sigma^2}{2\kappa^2} \right) (T - t + F(t,T)) \quad (1.4.11)$$

La media de largo plazo ajustada por riesgo μ^Q está dada por (1.4.2).

Las ecuaciones (1.1.2), (1.4.9), (1.4.10) y (1.4.11), puestas en función del tiempo $\tau = T - t$ de los plazos para el vencimiento del contrato, en vez de t que mide las fechas del calendario, nos permiten encontrar toda la curva de tasas cero según el modelo de estructura temporal de tipos de interés de un factor de Vasicek (1977). También nos permiten valorar cualquier bono libre de riesgo de incumplimiento, si consideramos que éste puede verse como un portafolio de bonos cupón cero.¹

En resumen: La técnica de valuación de bonos cupón cero neutral al riesgo sugiere como solución la *ecuación de valuación neutral al riesgo* o *fórmula de Feynman y Kac* (1.1.1), calculando su esperanza bajo la medida de distribución de probabilidad correspondiente al proceso estocástico ajustado por riesgo (1.4.3). Empleando el modelo de estructura temporal de tipos de interés de un factor de Vasicek (1977), la curva de tasas cero dada por la ecuación (1.1.2) está dada por las ecuaciones (1.4.9), (1.4.10) y (1.4.11). Como ya dijimos, las cuatro ecuaciones anteriormente mencionadas deberán ponerse en función de los plazos al vencimiento.

En el siguiente subcapítulo deduciremos la *ecuación diferencial parcial fundamental para el precio de los bonos cupón cero* basándonos en los supuestos para la forma matemática común de los modelos de estructura temporal de tipos de interés de un factor expuestos anteriormente, dado que es posible obtener una solución analítica libre de arbitraje de dicha ecuación para muchos modelos de un factor, entre ellos Vasicek (1977). Esto último lo haremos en el tema 1.6, donde obtendremos las mismas fórmulas para la curva de rendimientos con las que concluyó este subcapítulo.

1.5 Ecuación diferencial parcial fundamental para el precio de los bonos cupón cero

La *curva de rendimientos de equilibrio* es representada a través de los *precios de equilibrio de los bonos*, $P(t,T)$, que son la solución de la *ecuación diferencial parcial fundamental para el precio de los bonos cupón cero*. A continuación deduciremos la ecuación fundamental de dos modos distintos pero complementarios.² El primer procedimiento hace suposiciones directas respecto el *factor estocástico de descuento* y el segundo usa el *argumento clásico de ausencia de arbitraje* de Black y Scholes (1973) para construir un portafolio libre de riesgo.

¹ Castillo R., "Un modelo de valoración de bonos con rescate anticipado," p. 35.

² Jesper Lund, *Review of Continuous Time Term Structure Models. Part I: equilibrium models* (Dinamarca: Nykredit Markets, 1997), pp. 1 - 11.

1.5.1 Obtención por medio del factor de descuento estocástico

Además de los supuestos (i) al (iv), asumiremos que el factor de descuento estocástico, Λ_t , sigue la ecuación diferencial estocástica:

$$d\Lambda_t = -r_t \Lambda_t dt - \Lambda_t \lambda(r_t) dW_t \quad (1.5.1)$$

donde $\lambda(r_t)$ es el precio de riesgo de mercado y W_t un movimiento browniano.

Dado que un bono cupón cero da una unidad monetaria en la fecha de vencimiento T , su precio de equilibrio en el tiempo actual t es:

$$P(t, T) = E_t \left(\frac{\Lambda_T}{\Lambda_t} \right)$$

El subíndice t en el operador de la esperanza significa que ésta está condicionada en la información disponible hasta el tiempo presente. De la expresión anterior obtenemos:

$$E_t (d(P(t, T) \Lambda_T)) = 0$$

Se ve que el producto $P(t, T) \Lambda_T$ es una martingala porque la pendiente, o sea la tendencia, de su ecuación diferencial estocástica es cero. Calculemos la tendencia de la ecuación diferencial estocástica de $P(t, T) \Lambda_T$.

El proceso estocástico de Λ_T es (1.5.1). Para obtener el de $P(t, T)$ se retoma el supuesto (iv), se aplica el lema de Ito¹ y resulta la siguiente ecuación diferencial estocástica:

$$dP(t, T) = \mu_p(t, T) P(t, T) dt + \sigma_p(t, T) P(t, T) dW_t \quad (1.5.2)$$

donde

$$\mu_p(t, T) P(t, T) = \frac{\partial P}{\partial r} \mu(r) + \frac{\partial P}{\partial t} + \frac{1}{2} \frac{\partial^2 P}{\partial r^2} \sigma^2(r) \quad (1.5.3)$$

$$\sigma_p(t, T) P(t, T) = \frac{\partial P}{\partial r} \sigma(r) \quad (1.5.4)$$

¹ Se asume que $P(t, r, T)$ es una función $C^{1,2}$ de sus argumentos, es decir, es continuamente diferenciable una vez respecto su primer argumento y doblemente respecto el segundo. Esto es un requerimiento técnico para aplicar el teorema de Ito. Puede consultar el apéndice B.

$\mu_p(t, T)$ es el rendimiento instantáneo esperado de equilibrio del bono con vencimiento T y $\sigma_p(t, T)$ su desviación estándar. Ambas dependen de la tasa de interés instantánea pero lo hemos omitido por sencillez. Se aplica nuevamente el lema de Ito para obtener una expresión para la tendencia de $P(t, T)\Lambda_T$; resulta:

$$\begin{aligned} & E_t \left(\frac{d(P(t, T)\Lambda_T)}{dt} \right) \\ &= E_t(dP(t, T)\Lambda_T) + E_t(d\Lambda_T P(t, T)) + E_t(d\Lambda_T dP(t, T)) \\ &= \mu_p(t, T)P(t, T)\Lambda_T - r_t \Lambda_T P(t, T) - \Lambda_T \lambda(r_t) \sigma_p(t, T) P(t, T) \end{aligned}$$

El tercer término de la última línea es el término de segundo orden en el lema de Ito, adicional en comparación con una expansión en series de Taylor de primer orden.

Se iguala con cero la última línea para tener una restricción económica de los precios de los bonos:

$$\mu_p(t, T)P(t, T)\Lambda_T - r_t \Lambda_T P(t, T) - \Lambda_T \lambda(r_t) \sigma_p(t, T) P(t, T) = 0$$

Se divide la ecuación anterior por $\Lambda_T > 0$ y se sustituyen (1.5.3) y (1.5.4); finalmente obtenemos la *ecuación diferencial parcial fundamental para el precio de los bonos cupón cero*:

$$\frac{1}{2} \frac{\partial^2 P}{\partial r^2} \sigma^2(r) + \frac{\partial P}{\partial r} (\mu(r) - \lambda(r) \sigma(r)) + \frac{\partial P}{\partial t} - rP = 0 \quad (1.5.5)$$

cuya condición límite es:

$$P(T, T) = 1 \quad (1.5.6)$$

1.5.2 Obtención por el argumento de ausencia de arbitraje

Sin hacer suposiciones sobre el factor de descuento estocástico, se asume que $P(t, T)$ sigue la dinámica (1.5.2). Desconocemos $P(t, T)$ porque no conocemos el rendimiento esperado, $\mu_p(t, T)$. A continuación, usando el principio de ausencia de arbitraje se consigue un modelo de equilibrio para $\mu_p(t, T) \forall T$, que depende del precio de riesgo de mercado $\lambda(r_t)$ pero no de T .

De manera muy parecida a como Black y Scholes hicieron para obtener su método de valuación de opciones accionarias, Vasicek (1977) supone un portafolio de w_1 bonos con fecha de vencimiento T_1 y w_2 bonos con fecha de vencimiento $T_2 \neq T_1$, cuyo valor en el tiempo t es:

$$\Pi_t = w_1 P(t, T_1) + w_2 P(t, T_2)$$

y satisface la ecuación diferencial estocástica:

$$d\Pi_t = (w_1 \mu_p(t, T_1) P(t, T_1) + w_2 \mu_p(t, T_2) P(t, T_2)) dt + (w_1 \sigma_p(t, T_1) P(t, T_1) + w_2 \sigma_p(t, T_2) P(t, T_2)) dW_t \quad (1.5.7)$$

Supongamos, dado que (i) lo hace posible, que ajustamos continuamente las cantidades w_1 y w_2 de tal forma que:

$$w_1 \sigma_p(t, T_1) P(t, T_1) + w_2 \sigma_p(t, T_2) P(t, T_2) = 0 \quad (1.5.8)$$

Consecuentemente, el portafolio de (1.5.7) se reduce a un portafolio localmente determinístico y libre de riesgo, que sigue la ecuación diferencial estocástica:

$$d\Pi_t = (w_1 \mu_p(t, T_1) P(t, T_1) + w_2 \mu_p(t, T_2) P(t, T_2)) dt$$

Para evitar oportunidades de arbitraje, debe cumplirse que el exceso de rendimiento sobre r_t sea cero, es decir:

$$w_1 (\mu_p(t, T_1) - r_t) P(t, T_1) + w_2 (\mu_p(t, T_2) - r_t) P(t, T_2) = 0 \quad (1.5.9)$$

Sea (1.5.8) en forma matricial:

$$\begin{pmatrix} \sigma_p(t, T_1) P(t, T_1) & \sigma_p(t, T_2) P(t, T_2) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \mathbf{M}_1 \mathbf{w} = 0$$

Sea el sistema de ecuaciones homogéneo de (1.5.8) y (1.5.9) en forma matricial:

$$\begin{pmatrix} \sigma_p(t, T_1) P(t, T_1) & \sigma_p(t, T_2) P(t, T_2) \\ (\mu_p(t, T_1) - r_t) P(t, T_1) & (\mu_p(t, T_2) - r_t) P(t, T_2) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \mathbf{M}_2 \mathbf{w} = 0$$

Si \mathbf{M}_2 tiene rango dos, la solución de la ecuación anterior es la solución trivial, la cual no nos interesa. Sólo queda suponer que \mathbf{M}_2 es de rango uno. Siendo así, es posible escribir la segunda fila de \mathbf{M}_2 como una combinación lineal de las demás. Dado que la única fila que hay además de la segunda, es la primera, la combinación lineal susodicha será simplemente una multiplicación escalar del factor elegido por cada componente de la primera fila. Sea $\lambda(r_t)$ el factor designado. Igualado cada uno de los productos obtenidos con su elemento homólogo de la segunda fila obtenemos el resultado:

$$\lambda(r_t) \cdot (\sigma_p(t, T_j) P(t, T_j)) = (\mu_p(t, T_j) - r_t) P(t, T_j), \quad j = 1, 2$$

es decir,

$$\mu_p(t, T_j) = r_t + \lambda(r_t) \cdot \sigma_p(t, T_j), \quad j = 1, 2 \quad (1.5.10)$$

donde $\lambda(r_t)$ es el precio de riesgo de mercado.¹

Podemos darnos cuenta de que son irrelevantes valores específicos que demos a T_1 y T_2 , lo cual significa que (1.5.10) es válida para cualquier T . Entonces el problema de determinar $\mu_p(t, T) \forall T$ se reduce a especificar el parámetro escalar $\lambda(r)$, el cual es cuando mucho una función de la tasa de corto plazo r . Esto último requiere suposiciones adicionales sobre las preferencias del mercado y como vimos, cada modelo de ETTI especifica diferente $\lambda(r)$.

Si sustituimos en (1.5.10) la $\sigma_p(t, T)$ de (1.5.4) y la $\mu_p(t, T)$ de (1.5.3) y reacomodamos términos, obtenemos un resultado idéntico al señalado en la ecuación (1.5.5), con la misma condición límite. Esta equivalencia señala que el proceso del factor de descuento estocástico debe tener la forma (1.5.1) para desechar oportunidades de arbitraje.

1.6 Solución de la ecuación diferencial parcial fundamental para el precio de los bonos cupón cero

Reescribimos la ecuación diferencial parcial fundamental para el precio de los bonos cupón cero (1.5.5) aplicando un precio de riesgo de mercado constante $\lambda(r) = \lambda$ y los términos correspondientes con (1.3.1):

$$\frac{1}{2} \frac{\partial^2 P}{\partial r^2} \sigma^2 + \frac{\partial P}{\partial r} (\kappa(\mu - r) - \lambda \sigma) + \frac{\partial P}{\partial t} - rP = 0 \quad (1.6.1)$$

con condición límite (1.5.6).

Conjeturamos que su solución tiene la forma exponencial afín:

$$P(t, T) = \exp(F(\tau)r_t + G(\tau)) \text{ con } \tau = T - t \quad (1.6.2)$$

Diferenciamos respecto r , r^2 y t :

$$\frac{\partial P}{\partial r} = F(\tau)P(t, T)$$

¹ Escrito en la forma $\lambda(r_t) = \frac{\mu_p(t, T_j) - r_t}{\sigma_p(t, T_j)}$ el precio de riesgo de mercado representa claramente el exceso estandarizado de rendimiento sobre la tasa libre de riesgo.

$$\frac{\partial^2 P}{\partial r^2} = (F(\tau))^2 P(t, T)$$

$$\begin{aligned} \frac{\partial P}{\partial t} &= -\frac{\partial P}{\partial \tau} \\ &= -(F'(\tau)r + G'(\tau)) \cdot P(t, T) \\ &= -\left(\frac{dF(\tau)}{d\tau}r + \frac{dG(\tau)}{d\tau}\right) \cdot P(t, T) \end{aligned}$$

Sustituimos las derivadas anteriores en (1.6.1) y obtenemos:

$$\begin{aligned} \frac{1}{2}(F(\tau))^2 P(t, T)\sigma^2 + F(\tau)P(t, T)(\kappa(\mu - r) - \lambda\sigma) \\ - \left(\frac{dF(\tau)}{d\tau}r + \frac{dG(\tau)}{d\tau}\right) \cdot P(t, T) - rP(t, T) = 0 \end{aligned}$$

Dividimos por $P(t, T)$ y reunimos los términos que tienen el factor r :

$$\begin{aligned} \left(\frac{1}{2}(F(\tau))^2\sigma^2 + F(\tau)(\kappa\mu - \lambda\sigma) - \frac{dG(\tau)}{d\tau}\right) \\ - \left(\kappa F(\tau) + 1 + \frac{dF(\tau)}{d\tau}\right)r = 0 \end{aligned}$$

La ecuación diferencial parcial anterior debe satisfacerse $\forall r$, lo cual es posible si cada uno de sus dos términos más grandes es cero. Igualamos con cero dichos términos y obtenemos dos ecuaciones diferenciales ordinarias:

$$\frac{dG(\tau)}{d\tau} = \frac{1}{2}\sigma^2(F(\tau))^2 + (\kappa\mu - \lambda\sigma)F(\tau) \quad (1.6.3)$$

$$\frac{dF(\tau)}{d\tau} = -\kappa F(\tau) - 1 \quad (1.6.4)$$

Si podemos resolver (1.6.3) y (1.6.4), se podrá aceptar (1.6.2) como solución de (1.6.1).

La condición límite de (1.6.1) es (1.5.6), o sea $\lim_{t \rightarrow T} P(t, T) = 1$. Además, la solución de (1.6.1) debe tener la forma (1.6.2). Entonces:

$$\lim_{t \rightarrow T} \exp(F(\tau)r_t + G(\tau)) = 1$$

Dado que $t \rightarrow T \Rightarrow \tau \rightarrow 0$ las condiciones límite iniciales del sistema de ecuaciones formado por (1.6.3) y (1.6.4) son:

$$F(0) = 0 \text{ y } G(0) = 0$$

Primero resolveremos (1.6.4) por tener una sola incógnita. La reescribimos de la siguiente forma:

$$\frac{dF(\tau)}{d\tau} + \kappa F(\tau) = -1$$

Multiplicamos por $e^{\kappa\tau}$ y nos queda:

$$\frac{dF(\tau)}{d\tau} e^{\kappa\tau} + \kappa F(\tau) e^{\kappa\tau} = -e^{\kappa\tau}$$

Por la regla para derivar productos, el lado derecho de la igualdad anterior también puede escribirse:

$$\frac{d}{d\tau} (e^{\kappa\tau} F(\tau)) = -e^{\kappa\tau}$$

Por la relación entre diferenciación e integración, de lo anterior se tiene:

$$\int_0^{\tau} \frac{d}{dv} (e^{\kappa v} F(v)) dv = -\int_0^{\tau} e^{\kappa v} dv$$

Desarrollamos el lado izquierdo de la ecuación anterior y queda:

$$\int_0^{\tau} \frac{d}{dv} (e^{\kappa v} F(v)) dv = e^{\kappa\tau} F(\tau) - e^{\kappa \cdot 0} F(0) = e^{\kappa\tau} F(\tau)$$

Igualando los dos resultados previos obtenemos:

$$e^{\kappa\tau} F(\tau) = -\int_0^{\tau} e^{\kappa v} dv$$

Despejamos $F(\tau)$ de la expresión anterior:

$$F(\tau) = -e^{-\kappa\tau} \int_0^{\tau} e^{\kappa v} dv = -e^{-\kappa\tau} \left(\frac{1}{\kappa} e^{\kappa v} \right) \Big|_{v=0}^{v=\tau} = -e^{-\kappa\tau} \left(\frac{e^{\kappa\tau} - 1}{\kappa} \right)$$

y queda:

$$F(\tau) = \frac{e^{-\kappa\tau} - 1}{\kappa} \tag{1.6.5}$$

Por la relación entre diferenciación e integración, de (1.6.3) se tiene:

$$\int_0^{\tau} \frac{dG(v)}{dv} dv = \int_0^{\tau} \left(\frac{1}{2} \sigma^2 (F(v))^2 + (\kappa \mu - \lambda \sigma) F(v) \right) dv \quad (1.6.6)$$

Sustituyendo (1.6.5) en el lado derecho de (1.6.6) obtenemos:

$$\begin{aligned} \int_0^{\tau} \frac{dG(v)}{dv} dv &= \frac{1}{2} \sigma^2 \int_0^{\tau} \left(\frac{e^{-\kappa v} - 1}{\kappa} \right)^2 dv + (\kappa \mu - \lambda \sigma) \int_0^{\tau} \left(\frac{e^{-\kappa v} - 1}{\kappa} \right) dv \\ &= -\frac{\sigma^2}{4\kappa} \left(\frac{e^{-\kappa \tau} - 1}{\kappa} \right)^2 - \left(\mu - \frac{\lambda \sigma}{\kappa} - \frac{1}{2} \left(\frac{\sigma}{\kappa} \right)^2 \right) \left(\tau + \left(\frac{e^{-\kappa \tau} - 1}{\kappa} \right) \right) \end{aligned}$$

Desarrollamos el lado izquierdo de (1.6.6):

$$\int_0^{\tau} \frac{dG(v)}{dv} dv = G(\tau) - G(0) = G(\tau)$$

Igualando los dos resultados anteriores queda:

$$G(\tau) = -\frac{\sigma^2}{4\kappa} \left(\frac{e^{-\kappa \tau} - 1}{\kappa} \right)^2 - \left(\mu - \frac{\lambda \sigma}{\kappa} - \frac{1}{2} \left(\frac{\sigma}{\kappa} \right)^2 \right) \left(\tau + \left(\frac{e^{-\kappa \tau} - 1}{\kappa} \right) \right)$$

Sustituyendo (1.6.5) en la igualdad anterior, obtenemos:

$$G(\tau) = -\frac{\sigma^2}{4\kappa} (F(\tau))^2 - \left(\mu - \frac{\lambda \sigma}{\kappa} - \frac{1}{2} \left(\frac{\sigma}{\kappa} \right)^2 \right) (\tau + F(\tau)) \quad (1.6.7)$$

Finalmente, la solución de (1.6.1) es (1.6.2) junto con (1.6.5) y (1.6.7). Este resultado es el mismo de las ecuaciones (1.4.9), (1.4.10) y (1.4.11) puestas en función de τ , como se concluyó en el tema 1.4.

Capítulo 2: Modelos y técnicas por aplicar

2.1 La representación de espacio de estados

2.1.1 Vector de observaciones y vector de estado

La representación de espacio de estados es una poderosa herramienta que ayuda a manipular una gran variedad de modelos de series de tiempo.¹ Pueden construirse varias formas distintas para describir la dinámica de una misma serie de tiempo. Básicamente, un *modelo de espacio de estados en tiempo discreto* está dado por dos ecuaciones:²

1. Ecuación de estado o de transición:³

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \mathbf{v}_t \quad (2.1.1)$$

2. Ecuación de medidas u observaciones:

$$\mathbf{y}_t = \mathbf{C}_t \mathbf{x}_t + \mathbf{D}_t \mathbf{u}_t + \mathbf{w}_t \quad (2.1.2)$$

Para las dos ecuaciones anteriores tenemos:

- $t = 1, 2, \dots, N$.
- \mathbf{A}_t tiene dimensiones $m \times m$. Es llamada matriz de transición.
- \mathbf{B}_t tiene dimensiones $m \times n$.
- \mathbf{C}_t tiene dimensiones $n \times m$.
- \mathbf{D}_t tiene dimensiones $n \times n$.
- \mathbf{u}_t es el *vector de entradas o variables de control*. Tiene dimensiones $n \times 1$.
- \mathbf{x}_t es el *vector de estado o variable de estado* en el tiempo t . Tiene dimensiones $m \times 1$. En general sus elementos no son observables.
- \mathbf{y}_t es el *vector de observaciones, medidas o salidas* en el tiempo t . Tiene dimensiones $n \times 1$. Se llama así porque sus elementos son observables.
- \mathbf{v}_t y \mathbf{w}_t son vectores de ruido, de dimensiones respectivas $m \times 1$ y $n \times 1$.

Los vectores \mathbf{v}_t y \mathbf{w}_t tienen media cero y matrices de covarianza \mathbf{V}_t y \mathbf{W}_t , están incorrelacionados el uno con el otro¹ e incorrelacionados con el vector de estado inicial \mathbf{x}_0 para $t = 1, 2, \dots, N$, es decir:

¹ Andrew C. Harvey, *Forecasting, structural time series models and the Kalman filter* (Estados Unidos: Cambridge University Press, 1989), p. 100.

² Mariano J. Valderrama B., et. al., *Filtrado de Kalman, Aplicaciones en Economía e Ingeniería* (España: EUB, 1996), pp. 21 – 23 y 25.

³ Esta ecuación es un proceso de Markov de primer orden. Harvey, op. cit., p. 101 y Valderrama B., op. cit., p. 21. Puede consultar el apéndice B.

- $E(\mathbf{v}_t) = 0$
- $E(\mathbf{w}_t) = 0$
- $E(\mathbf{v}_t \mathbf{v}_t') = \mathbf{V}_t$
- $E(\mathbf{w}_t \mathbf{w}_t') = \mathbf{W}_t$
- $E(\mathbf{v}_t \mathbf{w}_t') = 0$
- $E(\mathbf{v}_t \mathbf{x}_0') = 0$
- $E(\mathbf{w}_t \mathbf{x}_0') = 0$

\mathbf{V}_t tiene dimensiones $m \times m$ y \mathbf{W}_t es de $n \times n$.

La especificación más simple para un ruido o componente irregular es el *ruido blanco*, que es una secuencia de variables aleatorias serialmente incorrelacionadas, con media y varianza constantes.² Por lo tanto, si \mathbf{V}_t y \mathbf{W}_t son constantes los ruidos \mathbf{v}_t y \mathbf{w}_t serán ruidos blancos.

El espacio de estados se termina de especificar suponiendo conocidas la media $E(\mathbf{x}_0) = \hat{\mathbf{x}}_0$ y la matriz de covarianzas $\text{var}(\mathbf{x}_0) = \mathbf{P}_0$ del vector de estado inicial \mathbf{x}_0 .

$\mathbf{A}_t, \mathbf{B}_t, \mathbf{C}_t, \mathbf{D}_t, \mathbf{V}_t$ y \mathbf{W}_t son llamadas *matrices del sistema*. Son generalmente determinísticas pero pueden ser estocásticas. Aquí las supondremos determinísticas.

Los subíndices t en las matrices del sistema indican que tenemos un modelo *variante en el tiempo*. Si por el contrario no cambian con el tiempo, omitimos sus subíndices t y decimos que tenemos un modelo *invariante en el tiempo* u *homogéneo en el tiempo*.

Las matrices del sistema pueden depender de un conjunto de parámetros $\Psi = \{\psi_1, \psi_2, \dots, \psi_p\}$ llamados *hiperparámetros*.

El sistema de ecuaciones de (2.1.1) y (2.1.2) es *lineal* porque para cualquier valor de t , \mathbf{y}_t puede expresarse como una combinación lineal de los \mathbf{v}_t y \mathbf{w}_t actuales y previos, así como del vector de estado inicial \mathbf{x}_0 .³

Si $N=1$ entonces $\mathbf{u}_t, \mathbf{y}_t$ y \mathbf{w}_t son escalares y tenemos un *modelo univariante o sistema de una entrada y una salida*; en inglés se llama SISO: Single Input Single Output. En un *modelo multivariante* hay múltiples entradas y múltiples salidas; en inglés se llama MIMO: Multiple Input Multiple Output. Análogamente existen sistemas SIMO y MISO.

¹ Este supuesto puede relajarse. Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 101.

² *Ibid.*, p. 18.

³ Si \mathbf{y}_t no es una función lineal de los \mathbf{v}_t y \mathbf{w}_t actuales y previos, no se aplica el filtro de Kalman que se muestra más adelante en la sección 2.2.1, sino el *filtro de Kalman extendido*. No lo enseñaremos en esta tesis, pero si el lector tiene curiosidad lo puede conocer en Harvey, op. cit., pp. 160-162.

En esta tesis no manejaremos variables de control. Supondremos que \mathbf{u}_t es un vector de unos y haremos un ligero cambio de notación al sistema (2.1.1) y (2.1.2). Usaremos \mathbf{b}_t en vez de $\mathbf{B}_t\mathbf{u}_t$ y \mathbf{d}_t en vez de $\mathbf{D}_t\mathbf{u}_t$. Nuestro espacio de estados para $t=1,\dots,N$ queda:

1. Ecuación de estado o de transición:

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{b}_t + \mathbf{v}_t \quad (2.1.3)$$

2. Ecuación de medidas u observaciones:

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{d}_t + \mathbf{w}_t \quad (2.1.4)$$

donde:

- \mathbf{b}_t tiene dimensiones $m \times 1$.
- \mathbf{d}_t tiene dimensiones $n \times 1$.
- El resto de las variables son como anteriormente se definió.

Observemos algunos ejemplos de cómo una serie de tiempo es representada en forma de espacio de estados.

Ejemplo 2.1.1:

Tenemos un modelo $AR(2)$:¹

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Sean ε_t v.a.i.i.d.² $\sim N(0, \sigma^2)$.

Definimos el vector de estado:

$$\mathbf{x}_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} c \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

Las matrices del sistema de la ecuación de estado son:

$$\mathbf{A} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} c \\ 0 \end{pmatrix}, \quad \mathbf{v}_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \text{ y } \mathbf{V} = \sigma^2.$$

Definimos el vector de medidas:

¹ AR: Siglas de Auto Regressive o Autorregresivo. Un proceso $AR(p)$ para la serie $\{Y_t\}$ se define por: $(\phi_0 + \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p)Y_t = \varepsilon_t$ donde B es el operador de retardo, es decir $BY_t = Y_{t-1}$.

² Variables aleatorias independientes e idénticamente distribuidas.

$$\mathbf{y}_t = y_t = (1 \ 0) \mathbf{x}_t$$

Las matrices del sistema de la ecuación de medidas son:

$$\mathbf{C} = (1 \ 0), \mathbf{d} = 0, \mathbf{w}_t = 0 \text{ y } \mathbf{W} = 0.$$

Ejemplo 2.1.2:

Otra representación de espacio de estados del modelo $AR(2)$ del ejemplo 2.1.1 puede ser con el vector de estado:

$$\mathbf{x}_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \phi_2 y_{t-2} \end{pmatrix} + \begin{pmatrix} c \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

Las matrices del sistema de la ecuación de transición son:

$$\mathbf{A} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} c \\ 0 \end{pmatrix}, \mathbf{v}_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \text{ y } \mathbf{V} = \sigma^2.$$

Ejemplo 2.1.3:

Tenemos un modelo $MA(1)$:¹

$$y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

Definimos el vector de estado:

$$\mathbf{x}_t = \begin{pmatrix} y_t - \mu \\ \theta \varepsilon_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} - \mu \\ \theta \varepsilon_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \theta \varepsilon_t \end{pmatrix}$$

Las matrices del sistema de la ecuación de transición son:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{b} = 0 \text{ y } \mathbf{v}_t = \begin{pmatrix} \varepsilon_t \\ \theta \varepsilon_t \end{pmatrix}.$$

Definimos el vector de observaciones:

$$\mathbf{y}_t = y_t = (1 \ 0) \mathbf{x}_t + \mu$$

Las matrices del sistema de la ecuación de observaciones son:

¹ MA: Siglas de Moving Average o Promedios Móviles. Un proceso $MA(q)$ para la serie $\{Y_t\}$ se define por: $Y_t = \varepsilon_t (\theta_0 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$.

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \mathbf{d} = \mu \text{ y } \mathbf{W} = 0.$$

El primer elemento de \mathbf{x}_t es $\theta \varepsilon_{t-1} + \varepsilon_t$, que es de hecho $y_t - \mu$.

Ejemplo 2.1.4:

Tenemos un modelo $ARMA(1,1)$:¹

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t + \theta \varepsilon_{t-1}$$

Definimos el vector de estado:

$$\mathbf{x}_t = \begin{pmatrix} y_t - \mu \\ \theta \varepsilon_t \end{pmatrix} = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} - \mu \\ \theta \varepsilon_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \theta \varepsilon_t \end{pmatrix}$$

Las matrices del sistema de la ecuación de estado son:

$$\mathbf{A} = \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{b} = 0 \text{ y } \mathbf{v} = \begin{pmatrix} \varepsilon_t \\ \theta \varepsilon_t \end{pmatrix}.$$

Definimos el vector de observaciones:

$$\mathbf{y}_t = y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \mathbf{x}_t + \mu$$

Donde las matrices del sistema de la ecuación de observaciones son:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \mathbf{d} = \mu \text{ y } \mathbf{W} = 0.$$

El primer elemento de \mathbf{x}_t es $\phi(y_{t-1} - \mu) + \theta \varepsilon_{t-1}$, que es de hecho $y_t - \mu$.

Ejemplo 2.1.5:

Tenemos el modelo $ARMA(p,q)$ general:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Definimos $m = \max(p, q + 1)$ y reescribimos la ecuación anterior como

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-m} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_{m-1} \varepsilon_{t-m+1}$$

¹ ARMA: Siglas de Auto Regressive Moving Average o Promedios Móviles Autorregresivos. Un proceso $ARMA(p, q)$ para la serie $\{Y_t\}$ se define por:

$$(\phi_0 + \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p) Y_t = \varepsilon_t (\theta_0 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q).$$

Este cambio de variable en los subíndices hará que haya la misma cantidad de términos autorregresivos ϕ_j que de promedios móviles θ_j , pese a que algunos valdrán cero a no ser que $p = q + 1$.

Definimos el vector de estado:

$$\mathbf{x}_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} + \dots + \phi_p y_{t-m+1} + \theta_1 \varepsilon_t + \dots + \theta_{m-1} \varepsilon_{t-m+2} \\ \vdots \\ \phi_m y_{t-1} + \theta_m \varepsilon_t \end{pmatrix}$$

La ecuación de transición es:

$$\mathbf{x}_t = \begin{pmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{m-1} & 0 & 0 & \dots & 1 \\ \phi_m & 0 & 0 & \dots & 0 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} \varepsilon_t \\ \theta_1 \varepsilon_t \\ \vdots \\ \theta_{m-2} \varepsilon_t \\ \theta_{m-1} \varepsilon_t \end{pmatrix}$$

La ecuación de observaciones es:

$$\mathbf{y}_t = y_t = (1 \quad \mathbf{0}'_{m-1}) \mathbf{x}_t$$

Cabe mencionar que, aunque no vimos ejemplos más complicados, es posible hacer una representación de espacio de estados con varios *vectores o variables de estado*. Una representación de varias variables de estado, puede tener su equivalente en otro espacio de estados de un solo vector o variable de estado.

2.1.2 Condiciones iniciales para espacio de estados invariante en el tiempo y vector de estado estacionario

Supongamos el espacio de estados de (2.1.3) y (2.1.4) pero invariante en el tiempo. El *vector de estado* será *estacionario* si además de ser invariante en el tiempo, cumple que las raíces características¹ de la matriz de transición \mathbf{A} tienen módulo² menor que uno, es decir:³

$$|\lambda_i(\mathbf{A})| < 1 \quad i = 1, \dots, m. \quad (2.1.5)$$

¹ Recuerde que las raíces características o valores propios de una matriz cuadrada \mathbf{M} son las raíces de su polinomio característico: $p(\lambda) = \det(\mathbf{M} - \lambda \mathbf{I})$. No confundamos esta lambda con la que utilizamos para el precio de riesgo de mercado.

² El valor numérico del determinante de una matriz también es conocido como módulo de la matriz.

³ Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 121.

Si además de ser estacionario, el vector de estado anterior tiene matriz de transición \mathbf{A} no singular¹, la media y la matriz de covarianzas del vector de estado inicial \mathbf{x}_0 están dados por:²

$$\hat{\mathbf{x}}_0 = E(\mathbf{x}_0) = (\mathbf{I}_m - \mathbf{A})^{-1} \mathbf{b} \quad (2.1.6)$$

$$\text{vec}(\mathbf{P}_0) = (\mathbf{I}_{m^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec}(\mathbf{V}) \quad (2.1.7)$$

El operador $\text{vec}(\cdot)$ indica que las columnas de la matriz argumento son colocadas una sobre otra, por ejemplo para una matriz $\mathbf{A} = \{a_{ij}\}$ de $m \times n$:

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \\ a_{12} \\ \vdots \\ a_{m2} \\ \vdots \\ a_{mn} \end{pmatrix}$$

El símbolo \otimes indica el producto de Kronecker. El producto de Kronecker de $\mathbf{A} \otimes \mathbf{B}$ donde $\mathbf{A} = \{a_{ij}\}$ es de $m \times n$ y $\mathbf{B} = \{b_{ij}\}$ de $l \times p$ es una matriz de tamaño $ml \times np$:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}$$

Ejemplo 2.1.6:

Retomemos el modelo y espacio de estados del ejemplo 2.1.1. El vector de estado \mathbf{x}_t es invariante en el tiempo.

El polinomio característico de la matriz de transición \mathbf{A} es:

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = \det \begin{pmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - \phi_1 \lambda - \phi_2.$$

Los valores propios de \mathbf{A} son:

¹ Una matriz singular es una matriz que es cuadrada pero tiene determinante cero y por lo tanto no es invertible. Sinónimos de matriz no singular son: Regular, invertible o inversible.

² Eric Zivot, *State Space Models and the Kalman Filter* (Estados Unidos: University of Washington, 2006), p. 2 o Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 121.

$$\lambda = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4 \cdot \phi_2}}{2}$$

Supongamos que ϕ_1 y ϕ_2 son tales que:

$$\left| \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4 \cdot \phi_2}}{2} \right| < 1$$

Consecuentemente \mathbf{x}_t es estacionario.

El vector de estado inicial es $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, \mathbf{P}_0)$

Aplicamos (2.1.6):

$$\hat{\mathbf{x}}_0 = E(\mathbf{x}_0) = (\mathbf{I}_2 - \mathbf{A})^{-1} \mathbf{b} = \begin{pmatrix} 1 - \phi_1 & -\phi_2 \\ -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} c \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{c}{1 - \phi_1 - \phi_2} \\ \frac{c}{1 - \phi_1 - \phi_2} \end{pmatrix}$$

Aplicamos (2.1.7):

$$\begin{aligned} \text{vec}(\mathbf{P}_0) &= (\mathbf{I}_4 - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec}(\mathbf{V}) \\ &= \begin{pmatrix} 1 - \phi_2^2 & -\phi_1 \phi_2 & -\phi_1 \phi_2 & -\phi_2^2 \\ -\phi_1 & 1 & -\phi_2 & 0 \\ -\phi_1 & -\phi_2 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

2.2 Filtro de Kalman

A un espacio de estados definido cuyo vector de estado inicial \mathbf{x}_0 y vectores de ruido \mathbf{v}_t y \mathbf{w}_t son gaussianos podemos aplicarle el filtro de Kalman. El filtro de Kalman es un procedimiento recursivo para calcular el estimador óptimo del vector de estado en el tiempo t y la matriz de covarianzas del error de dicha estimación, dado que tenemos información hasta el tiempo t . La sección 2.2.1 presenta las ecuaciones del filtro de Kalman y 2.2.2 dice cómo fueron obtenidas.

2.2.1 Las ecuaciones del filtro de Kalman

El filtro de Kalman es un conjunto de ecuaciones recursivas para determinar un estimador óptimo de los vectores de estado subsecuentes haciendo uso de las

observaciones de entrada-salida, las matrices del sistema disponibles y las matrices de covarianza conocidas. A la estimación de \mathbf{x}_t , dado que tenemos información hasta el tiempo $t-1$ la denotaremos $\hat{\mathbf{x}}_{t|t-1}$, pero si ya tenemos datos hasta t la representaremos $\hat{\mathbf{x}}_{t|t}$. Otros autores consideran esto una notación exagerada y prefieren sólo escribir $\hat{\mathbf{x}}_t$.

Junto con cada vector de estado estimado se calculará la matriz de covarianzas de dimensiones $m \times m$ de los errores en su estimación:

$$\mathbf{P}_{t|t-1} = E\left(\left(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}\right)\left(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}\right)'\right)$$

$$\mathbf{P}_{t|t} = E\left(\left(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}\right)\left(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}\right)'\right)$$

Algunos prefieren la notación \mathbf{P}_t en vez de $\mathbf{P}_{t|t}$.

Las ecuaciones recursivas para implementar el filtro de Kalman en $t = 1, \dots, N$ son:¹

1. Estimador adelantado una etapa:

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{A}_t \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{b}_t \quad (2.2.1)$$

Por lo que vimos en la sección 2.1 es claro que $\hat{\mathbf{x}}_{t|t-1}$ es de dimensiones $m \times 1$.

2. Covarianza del error de estimación adelantada una etapa:

$$\mathbf{P}_{t|t-1} = \mathbf{A}_t \mathbf{P}_{t-1|t-1} \mathbf{A}_t' + \mathbf{V}_t \quad (2.2.2)$$

Sabemos que $\mathbf{P}_{t|t-1}$ es de dimensiones $m \times m$.

3. Innovación:

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{C}_t \hat{\mathbf{x}}_{t|t-1} - \mathbf{d}_t \quad (2.2.3)$$

Se tiene que \mathbf{v}_t es de dimensiones $n \times 1$.

4. Matriz de ganancia:

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{C}_t' \mathbf{F}_t^{-1} \quad (2.2.4)$$

Donde \mathbf{K}_t es de dimensiones $m \times n$ y se tiene que:

¹ Puede consultar el apéndice C.

$$\mathbf{F}_t = \mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}'_t + \mathbf{W}_t \quad (2.2.5)$$

Se tiene que \mathbf{F}_t es de dimensiones $n \times n$.

5. Estimador filtrado:

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \mathbf{v}_t \quad (2.2.6)$$

Tenemos que $\hat{\mathbf{x}}_{t|t}$ es de dimensiones $m \times 1$.

6. Covarianza del error de estimación:

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) \mathbf{P}_{t|t-1} \quad (2.2.7)$$

Obviamente $\mathbf{P}_{t|t}$ es de dimensiones $m \times m$.

A las ecuaciones (2.2.1) y (2.2.2) se les llama *ecuaciones de pronóstico* y a (2.2.6) y (2.2.7) junto con (2.2.3) y (2.2.4) de por medio se les llama *ecuaciones de actualización*, dado que se calculan una vez que ya se dispone de la nueva observación, correspondiente al tiempo t vigente.

La ecuación recursiva resultante de sustituir (2.2.2) en (2.2.7) es conocida como *ecuación de Riccati*.¹

Para un espacio de estados en el cual las matrices del sistema son constantes conocidas, se cumple con exactitud que $\mathbf{v}_t \sim N(0, \mathbf{F}_t)$. Por otra parte, en general eso no se sostiene cuando las matrices del sistema dependen de hiperparámetros desconocidos que deben estimarse.

La ecuación (2.2.4) requiere calcular la inversa de una matriz, lo cual puede ser exhaustivo. Veamos el siguiente lema de inversión matricial.

Lema 2.2.1. Sean \mathbf{A} , \mathbf{B} y \mathbf{C} matrices de dimensiones $n \times n$, $m \times m$ y $m \times n$ respectivamente. Si existe la inversa de \mathbf{A} y \mathbf{C} entonces

$$(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{B}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1}$$

Aplicando el lema anterior en (2.2.5) obtenemos una igualdad que puede ser de utilidad para agilizar la implementación del filtro:

$$\mathbf{F}_t^{-1} = (\mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}'_t + \mathbf{W}_t)^{-1} = \mathbf{W}_t^{-1} - \mathbf{W}_t^{-1} \mathbf{C}_t (\mathbf{C}'_t \mathbf{W}_t^{-1} \mathbf{C}_t + \mathbf{P}_{t|t-1}^{-1})^{-1} \mathbf{C}'_t \mathbf{W}_t^{-1}$$

¹ Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 106.

Una modalidad del filtro de Kalman exige que $\mathbf{P}_{t|t-1} = \mathbf{P}_{t|t-1}^{\frac{1}{2}} \mathbf{P}_{t|t-1}'^{\frac{1}{2}}$ donde $\mathbf{P}_{t|t-1}^{\frac{1}{2}}$ es matriz triangular superior de Cholesky¹ o raíz cuadrada de $\mathbf{P}_{t|t-1}$. Esto garantiza que $\mathbf{P}_{t|t-1}$ es definida positiva y existe su inversa. Lo mismo se exige para las matrices $\mathbf{P}_{t|t}$, \mathbf{V}_t y \mathbf{W}_t .

2.2.2 Deducción de las ecuaciones del filtro de Kalman

Dijimos que definir un espacio de estados requiere sugerir la media $\hat{\mathbf{x}}_{0|0}$ y la matriz de covarianzas $\mathbf{P}_{0|0}$ del vector de estado inicial \mathbf{x}_0 y suponer que para $t=1, 2, \dots, N$ los ruidos \mathbf{v}_t y \mathbf{w}_t son independientes uno del otro, así como de \mathbf{x}_0 .

Si adicionalmente suponemos que \mathbf{x}_0 y los vectores de ruido \mathbf{v}_t y \mathbf{w}_t tienen cada uno distribución normal multivariante, el vector de estado en $t=1$ dado por:

$$\mathbf{x}_1 = \mathbf{A}_1 \mathbf{x}_0 + \mathbf{b}_1 + \mathbf{v}_1$$

también tiene distribución normal multivariante con media:

$$\hat{\mathbf{x}}_{1|0} = \mathbf{A}_1 \hat{\mathbf{x}}_{0|0} + \mathbf{b}_1$$

y matriz de covarianzas:

$$\mathbf{P}_{1|0} = \mathbf{A}_1 \mathbf{P}_{0|0} \mathbf{A}_1' + \mathbf{V}_1$$

Para obtener la distribución de \mathbf{x}_1 condicional en \mathbf{y}_1 escribimos:

$$\mathbf{x}_1 = \hat{\mathbf{x}}_{1|0} + (\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0}) \quad (2.2.8)$$

$$\mathbf{y}_1 = \mathbf{C}_1 \hat{\mathbf{x}}_{1|0} + \mathbf{d}_1 + \mathbf{C}_1 (\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0}) + \mathbf{w}_1 \quad (2.2.9)$$

La última es la ecuación de medida (2.1.4) más un cero.

Así puede verse que el vector $\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{pmatrix}$ tiene distribución normal multivariante con media

$$\begin{pmatrix} \hat{\mathbf{x}}_{1|0} \\ \mathbf{C}_1 \hat{\mathbf{x}}_{1|0} + \mathbf{d}_1 \end{pmatrix}.$$

¹ Puede consultar el apéndice D.

Por su parte, la matriz de covarianzas de $\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{pmatrix}$ es $\begin{pmatrix} \text{cov}(\mathbf{x}_1, \mathbf{x}_1) & \text{cov}(\mathbf{x}_1, \mathbf{y}_1) \\ \text{cov}(\mathbf{y}_1, \mathbf{x}_1) & \text{cov}(\mathbf{y}_1, \mathbf{y}_1) \end{pmatrix}$, donde:

$$\text{cov}(\mathbf{x}_1, \mathbf{x}_1) = E\left((\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \right) = \mathbf{P}_{1|0}$$

$$\begin{aligned} \text{cov}(\mathbf{x}_1, \mathbf{y}_1) &= E\left((\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})(\mathbf{y}_1 - \hat{\mathbf{y}}_{1|0})' \right) \\ &= E\left((\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})(\mathbf{C}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0}) + \mathbf{w}_1)' \right) \\ &= E\left((\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \mathbf{C}_1' \right) + E\left((\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})\mathbf{w}_1' \right) \\ &= \mathbf{P}_{1|0} \mathbf{C}_1' \end{aligned}$$

$$\begin{aligned} \text{cov}(\mathbf{y}_1, \mathbf{x}_1) &= E\left((\mathbf{y}_1 - \hat{\mathbf{y}}_{1|0})(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \right) \\ &= E\left((\mathbf{C}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0}) + \mathbf{w}_1)(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \right) \\ &= E\left(\mathbf{C}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \right) + E\left(\mathbf{w}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \right) \\ &= \mathbf{C}_1 \mathbf{P}_{1|0} \end{aligned}$$

$$\begin{aligned} \text{cov}(\mathbf{y}_1, \mathbf{y}_1) &= E\left((\mathbf{y}_1 - \hat{\mathbf{y}}_{1|0})(\mathbf{y}_1 - \hat{\mathbf{y}}_{1|0})' \right) \\ &= E\left((\mathbf{C}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0}) + \mathbf{w}_1)(\mathbf{C}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0}) + \mathbf{w}_1)' \right) \\ &= E\left(\mathbf{C}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \mathbf{C}_1' \right) + E\left(\mathbf{C}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})\mathbf{w}_1' \right) \\ &\quad + E\left(\mathbf{w}_1(\mathbf{x}_1 - \hat{\mathbf{x}}_{1|0})' \mathbf{C}_1' \right) + E\left(\mathbf{w}_1 \mathbf{w}_1' \right) \\ &= \mathbf{C}_1 \mathbf{P}_{1|0} \mathbf{C}_1' + \mathbf{W}_1 \end{aligned}$$

Es decir:

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{\mathbf{x}}_{1|0} \\ \mathbf{C}_1 \hat{\mathbf{x}}_{1|0} + \mathbf{d}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{1|0} & \mathbf{P}_{1|0} \mathbf{C}_1' \\ \mathbf{C}_1 \mathbf{P}_{1|0} & \mathbf{C}_1 \mathbf{P}_{1|0} \mathbf{C}_1' + \mathbf{W}_1 \end{pmatrix} \right)$$

Usaremos el siguiente lema.¹

¹ La demostración de dicho lema puede consultarse en Harvey, *Forecasting, structural time series models and the Kalman filter*, pp. 165-166.

Lema 2.2.2: Sean \mathbf{x} y \mathbf{y} un par de vectores con función de distribución normal multivariante conjunta tal que $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$ tiene media $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}$ y matriz de covarianzas

$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}$. Entonces la distribución de \mathbf{x} condicional en \mathbf{y} es también normal multivariante con media $\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y)$ y matriz de covarianzas $\boldsymbol{\Sigma}_{x|y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}$. Las matrices $\boldsymbol{\Sigma}$ y $\boldsymbol{\Sigma}_{yy}$ se asumen no singulares pero $\boldsymbol{\Sigma}_{yy}^{-1}$ puede ser reemplazada por una pseudo inversa.

Aplicando el lema 2.2.2 resulta que la función de distribución de \mathbf{x}_1 condicional en un valor particular de \mathbf{y}_1 es normal multivariante con media:

$$\mathbf{x}_{1|1} = \hat{\mathbf{x}}_{1|0} + \mathbf{P}_{1|0} \mathbf{C}'_1 \mathbf{F}_1^{-1} (\mathbf{y}_1 - \mathbf{C}_1 \hat{\mathbf{x}}_{1|0} - \mathbf{d}_1) \quad (2.2.10)$$

y matriz de covarianzas:

$$\mathbf{P}_{1|1} = \mathbf{P}_{1|0} - \mathbf{P}_{1|0} \mathbf{C}'_1 \mathbf{F}_1^{-1} \mathbf{C}_1 \mathbf{P}_{1|0} \quad (2.2.11)$$

donde:

$$\mathbf{F}_1 = \mathbf{C}_1 \mathbf{P}_{1|0} \mathbf{C}'_1 + \mathbf{W}_1 \quad (2.2.12)$$

Repetiendo del paso (2.2.8) al (2.2.12) para $t = 2, \dots, N$ se obtiene un conjunto de ecuaciones con la misma forma que las recursiones del filtro de Kalman (2.2.1) a (2.2.7). Esta deducción nos permite interpretar $\hat{\mathbf{x}}_{t|t}$ y $\mathbf{P}_{t|t}$ como la media y matriz de covarianzas de la distribución condicional de \mathbf{x}_t .¹ Puede argumentarse² que $\hat{\mathbf{x}}_{t|t}$ es el *estimador de mínimo error cuadrático medio* de \mathbf{x}_t . Es por eso que $\hat{\mathbf{x}}_{t|t}$ es un estimador óptimo³ de \mathbf{x}_t y $\mathbf{P}_{t|t}$ la matriz de covarianzas del error de la estimación, usando la información disponible en el tiempo t .

Cuando los ruidos del modelo de espacio de estados no son normalmente distribuidos, en general no sigue siendo cierto que el filtro de Kalman proporcione la media condicional del vector de estado.⁴ El espacio de estados específico que presentaremos en el capítulo 3 sí tiene ruidos gaussianos y por lo tanto no expondremos aquí la forma

¹ Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 110.

² B. D. O. Anderson y J. B. Moore, *Optimal filtering* (Estados Unidos: Prentice-Hall, 1979), pp. 29 – 32.

³ Recordemos que el *error cuadrático medio* de un estimador $\hat{\theta}$ del parámetro θ está definido como: $ECM(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}))^2 + E(\theta - E(\hat{\theta}))^2 = \text{var}(\hat{\theta}) + (\text{sesgo})^2$.

Recordemos además que un estimador $\hat{\theta}$ que tiene un error cuadrático medio menor o igual que el error cuadrático medio de cualquier otro estimador, para todos los valores del parámetro θ , recibe el nombre de *estimador óptimo* de θ .

⁴ Harvey, op. cit., p. 111.

de calcular la media condicional en los casos en que los ruidos no son gaussianos.¹ No obstante, cabe mencionar que el estimador del vector de estado que proporciona el filtro de Kalman, bajo las modificaciones respectivas, sigue siendo un estimador óptimo porque genera el estimador que minimiza el error cuadrático medio dentro de la clase de todos los estimadores que son combinaciones lineales de las observaciones.

2.2.3 Función de verosimilitud del vector de observaciones en forma de descomposición del error de pronóstico

Ya dijimos que las matrices del sistema de un espacio de estados adquieren su valor en función de un conjunto de parámetros $\psi_1, \psi_2, \dots, \psi_p$ llamados hiperparámetros. A continuación deduciremos la *función de máxima verosimilitud* de \mathbf{y}_t , pues será

necesaria para encontrar el estimador máximo verosímil de $\boldsymbol{\psi} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_p \end{pmatrix}$.

Cuando se tienen N conjuntos de observaciones $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ que son independientes e idénticamente distribuidos, la función de densidad conjunta o función de verosimilitud está dada por:

$$L_{\boldsymbol{\psi}}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) = \prod_{t=1}^N p(\mathbf{y}_t)$$

La principal característica de un modelo de series de tiempo es que las observaciones no son independientes.² Por lo tanto, la fórmula anterior no es aplicable. En vez de la función de densidad de probabilidad marginal $p(\mathbf{y}_t)$, utilizamos la función de densidad de probabilidad condicional $p(\mathbf{y}_t | \mathbf{Y}_{t-1})$ y obtenemos la función de verosimilitud:

$$L_{\boldsymbol{\psi}}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) = \prod_{t=1}^N p(\mathbf{y}_t | \mathbf{Y}_{t-1})$$

Suponiendo que las observaciones han sido tomadas del modelo de espacio de estados (2.1.3) y (2.1.4) con ruidos perturbadores y vector de estado inicial gaussianos, la función de distribución de \mathbf{y}_t condicionada en \mathbf{Y}_{t-1} también tiene función de densidad normal multivariante. Reescribimos (2.1.4) de siguiente manera:

¹ Si el lector tiene curiosidad por conocer las modificaciones requeridas en el filtro de Kalman para cuando los espacios de estados tienen ruidos no gaussianos, puede consultar *Ibíd.*, pp. 162-165.

² Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 125.

$$\mathbf{y}_t = \mathbf{C}_t \hat{\mathbf{x}}_{t|t-1} + \mathbf{d}_t + \mathbf{C}_t (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1}) + \mathbf{w}_t$$

Se vuelve evidente que la media de \mathbf{y}_t condicionada en \mathbf{Y}_{t-1} es:

$$E(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \hat{\mathbf{y}}_{t|t-1} = \mathbf{C}_t \hat{\mathbf{x}}_{t|t-1} + \mathbf{d}_t$$

y la matriz de covarianzas es:

$$E\left(\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)\left(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}\right)'\right) = \mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}_t' + \mathbf{W}_t = \mathbf{F}_t$$

Dado que el vector \mathbf{y}_t tiene dimensión n su función de densidad de probabilidad condicionada en \mathbf{Y}_{t-1} es:

$$p(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n |\mathbf{F}_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})' \mathbf{F}_t^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})\right)$$

Dado que $\mathbf{v}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}$ tenemos:

$$p(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n |\mathbf{F}_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t\right) \quad (2.2.13)$$

La función de verosimilitud de \mathbf{y}_t es:

$$L_\psi(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) = \prod_{t=1}^N \left(\left(\frac{1}{\sqrt{2\pi}}\right)^n |\mathbf{F}_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t\right) \right) \quad (2.2.14)$$

La función de log-verosimilitud de \mathbf{y}_t es:

$$\begin{aligned} \ln L_\psi(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) &= \ln \left(\prod_{t=1}^N \left(\left(\frac{1}{\sqrt{2\pi}}\right)^n |\mathbf{F}_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t\right) \right) \right) \\ &= \sum_{t=1}^N \ln \left(\left(\frac{1}{\sqrt{2\pi}}\right)^n |\mathbf{F}_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t\right) \right) \\ &= \sum_{t=1}^N \left(-\frac{n}{2} \cdot \ln(2\pi) - \frac{1}{2} \ln |\mathbf{F}_t| - \frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right) \\ &= -\frac{nN}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^N \ln |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^N \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \end{aligned} \quad (2.2.15)$$

Dado que $\hat{\mathbf{y}}_{t|t-1}$ es un estimador de \mathbf{y}_t que minimiza el error cuadrático medio, el vector de innovación \mathbf{v}_t puede ser interpretado como un vector de errores de

pronóstico; consecuentemente, (2.2.14) es conocida como la *función de verosimilitud de las observaciones en forma de descomposición del error de pronóstico*.¹

La igualdad siguiente² puede ayudarnos en evaluar la última expresión:

$$|\mathbf{F}_t| = |\mathbf{W}_t| \cdot |\mathbf{P}_{t|t-1}| \cdot |\mathbf{C}'_t \mathbf{W}_t^{-1} \mathbf{C}_t + \mathbf{P}_{t|t-1}^{-1}|$$

Hemos visto cómo el algoritmo recursivo de Kalman proporciona los valores \mathbf{v}_t y \mathbf{F}_t , necesarios para evaluar la función de máxima verosimilitud.

2.3 Maximización de la función de verosimilitud para estimar los hiperparámetros de las matrices del sistema

Condiciones de optimalidad

La condición de optimalidad necesaria para que $L_\psi(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ sea máxima (mínima) en $\psi = \psi^*$ es:

$$\nabla L_{\psi^*}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) = \mathbf{0} \quad (2.3.1)$$

Esto significa que para $i = 1, \dots, p$

$$\left. \frac{\partial L_\psi(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)}{\partial \psi_i} \right|_{\psi=\psi^*} = 0$$

Las dos ecuaciones anteriores equivalen a que $\|\nabla L_{\psi^*}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)\| \leq \varepsilon$ donde ε es un pequeño escalar positivo. Los puntos ψ^* que satisfacen esta condición se llaman *puntos estacionarios*.

Una vez cumplido lo anterior, la condición suficiente de optimalidad para que ψ^* sea un *máximo local estricto* de $L_\psi(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ es que $\nabla^2 L_{\psi^*}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ sea definida negativa. Análogamente, si queremos que ψ^* sea un *mínimo local estricto* de $L_\psi(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ deberá cumplirse que $\nabla^2 L_{\psi^*}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ sea definida positiva. De ahora en adelante cuando digamos *punto óptimo* nos referiremos al máximo o mínimo local estricto que se busca según sea el caso.

¹ Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 126.

² *Ibíd.*, p. 108.

Hacemos una pausa para recordar que una matriz definida positiva es el análogo de un escalar real positivo y una matriz definida negativa es el análogo de un escalar real negativo.¹

A la función de verosimilitud $L_{\psi}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ a partir de ahora la denotaremos abreviadamente L .

Si una matriz es definida positiva, también lo es su inversa. Análogamente se tiene para una matriz definida negativa. Por ende, si $\nabla^2 L$ es definida positiva se cumplirá que $-\nabla L' (\nabla^2 L)^{-1} \nabla L < 0$ y si $\nabla^2 L$ es definida negativa se satisfará $-\nabla L' (\nabla^2 L)^{-1} \nabla L > 0$.

Cuando ψ no es escalar no es fácil resolver la *ecuación de verosimilitud* (2.3.1), haciéndose necesario aplicar métodos recursivos de optimización.² El vector de puntaje $\mathbf{S}(\psi)$ y la matriz de Hess o hessiana $\mathbf{H}(\psi)$, son dos elementos utilizados en la implementación de dichos métodos. Cabe señalar que los más eficientes métodos modernos evitan el uso de la matriz hessiana por requerir en exceso memoria de cómputo.

Propiedades asintóticas del estimador máximo verosímil

La *matriz de información* $\mathbf{I}(\psi) = E(-\mathbf{H}(\psi))$ desempeña un papel esencial en la teoría de estimación de máxima verosimilitud. Sea ψ^* el estimador máximo verosímil del vector ψ de $p \times 1$, obtenido tras maximizar (2.2.15) con algún método y suponga que la matriz de información, $\mathbf{I}(\psi)$, converge a una matriz definida positiva $\mathbf{IA}(\psi)$ tal que

$$\mathbf{IA}(\psi) = \lim_{N \rightarrow \infty} \frac{1}{N} \cdot \mathbf{I}(\psi)$$

Sujeto a ciertas condiciones de regularidad,³ $\sqrt{N}(\psi^* - \psi)$ tiende a tener una distribución normal multivariante con media vector de ceros y matriz de covarianzas $(\mathbf{IA}(\psi))^{-1}$. Este resultado significa que ψ^* tiene una distribución normal asintótica con media ψ y matriz de covarianzas:

$$\text{A var}(\psi^*) = \frac{1}{N} \cdot (\mathbf{IA}(\psi))^{-1}$$

A la inversa de la matriz de información también se le llama *información límite a la varianza* o límite inferior de Cramer-Rao para la varianza del estimador máximo verosímil. La *desigualdad de Cramer-Rao* está dada por:

¹ Puede consultar la primera parte del apéndice D.

² Puede consultar el apéndice E.

³ Puede consultarlas en Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 128.

$$\text{var}(\boldsymbol{\psi}^*) \geq \mathbf{I}^{-1}(\boldsymbol{\psi})$$

Usando términos más estrictos, si $\boldsymbol{\psi}^*$ es el verdadero valor del vector de parámetros $\boldsymbol{\psi}$ y $\hat{\boldsymbol{\psi}}^*$ es el estimador máximo verosímil de $\boldsymbol{\psi}$ dada una muestra de tamaño N , las propiedades asintóticas de dicho estimador son:¹

- a) Consistencia: $\lim_{N \rightarrow \infty} \hat{\boldsymbol{\psi}}^* = \boldsymbol{\psi}^*$
- b) Insesgadez asintótica: $\lim_{N \rightarrow \infty} E(\hat{\boldsymbol{\psi}}^*) = \boldsymbol{\psi}^*$
- c) Normalidad asintótica: $\hat{\boldsymbol{\psi}}^* \underset{\text{asintóticamente}}{\sim} N(\boldsymbol{\psi}^*, \mathbf{I}^{-1}(\boldsymbol{\psi}^*))$
- d) Eficiencia asintótica: $\lim_{N \rightarrow \infty} \text{cov}(\hat{\boldsymbol{\psi}}^*) = \mathbf{I}^{-1}(\boldsymbol{\psi}^*)$

En las secciones siguientes deduciremos el vector de puntaje, la matriz hessiana y la matriz de información de (2.2.15).

2.3.1 El vector de puntaje

Sea $\mathbf{S}(\hat{\boldsymbol{\psi}}^k)$ el vector de primeras derivadas de $\ln L$ con respecto a los elementos de $\boldsymbol{\psi}$, tomados en el estimador iterado más reciente $\hat{\boldsymbol{\psi}}^k$ de $\boldsymbol{\psi}$. Es decir, es el vector gradiente $\nabla \ln L_{\boldsymbol{\psi}}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ pero transpuesto, evaluado en $\hat{\boldsymbol{\psi}}^k$. Llamaremos a $\mathbf{S}(\hat{\boldsymbol{\psi}}^k)$ *vector de puntaje en $\hat{\boldsymbol{\psi}}^k$* .

Sea l_t el logaritmo de la función de densidad de probabilidad condicional $p(\mathbf{y}_t | \mathbf{Y}_{t-1})$. De (2.2.13) tenemos:

$$l_t = \ln p(\mathbf{y}_t | \mathbf{Y}_{t-1}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{F}_t| - \frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \quad (2.3.2)$$

La función de log-verosimilitud de \mathbf{y}_t puede escribirse:

$$\ln L = \sum_{t=1}^N l_t$$

El i -ésimo elemento del vector $\mathbf{S}(\boldsymbol{\psi})$ es:

$$\mathbf{S}_i(\boldsymbol{\psi}) = \frac{\partial \ln L}{\partial \psi_i} = \frac{\partial}{\partial \psi_i} \left(\sum_{t=1}^N l_t \right) = \sum_{t=1}^N \frac{\partial l_t}{\partial \psi_i} \quad (2.3.3)$$

¹ Sonia Sotoca L., “Aplicaciones Econométricas del filtro de Kalman y algunas variaciones numéricas: El Filtro de Chandrasekhar” (Tesis de doctorado, Universidad Complutense de Madrid, España, 1992), pp. 73-74.

De (2.3.2) tenemos:

$$\begin{aligned}\frac{\partial l_t}{\partial \psi_i} &= \frac{\partial}{\partial \psi_i} \left(-\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{F}_t| - \frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right) = \\ &= -\frac{1}{2} \cdot |\mathbf{F}_t|^{-1} \cdot \frac{\partial |\mathbf{F}_t|}{\partial \psi_i} - \frac{1}{2} \left(\left(\frac{\partial \mathbf{v}_t'}{\partial \psi_i} \right) \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{v}_t' \left(\frac{\partial \mathbf{F}_t^{-1}}{\partial \psi_i} \right) \mathbf{v}_t + \mathbf{v}_t' \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{v}_t}{\partial \psi_i} \right) \right)\end{aligned}\quad (2.3.4)$$

Para cualquier matriz simétrica \mathbf{A} , las derivadas del determinante y la inversa respecto a la variable \mathbf{x} son:

$$\frac{\partial |\mathbf{A}|}{\partial \mathbf{x}} = |\mathbf{A}| \cdot \text{traza} \left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{x}} \right) \quad (2.3.5)$$

$$\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{x}} = -\mathbf{A}^{-1} \left(\frac{\partial \mathbf{A}}{\partial \mathbf{x}} \right) \mathbf{A}^{-1} \quad (2.3.6)$$

Aplicando (2.3.5) y (2.3.6) en (2.3.4) obtenemos:

$$\begin{aligned}\frac{\partial l_t}{\partial \psi_i} &= -\frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \right) \\ &\quad - \frac{1}{2} \left(\frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{v}_t' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{v}_t' \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \right)\end{aligned}$$

Los términos dentro del segundo paréntesis son escalares y por ende si les aplicamos la función traza siguen valiendo lo mismo. Reescribiendo tenemos:

$$\begin{aligned}\frac{\partial l_t}{\partial \psi_i} &= -\frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \right) + \frac{1}{2} \cdot \text{traza} \left(\mathbf{v}_t' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \right) \\ &\quad - \frac{1}{2} \cdot \text{traza} \left(\frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \right) - \frac{1}{2} \cdot \text{traza} \left(\mathbf{v}_t' \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \right)\end{aligned}$$

Dada la propiedad de la función traza que dice que $\text{traza}(\mathbf{AB}) = \text{traza}(\mathbf{BA})$ siempre que \mathbf{A} y \mathbf{B} sean multiplicables aunque $\mathbf{AB} \neq \mathbf{BA}$ modificamos el segundo término de la expresión anterior. Por su parte, el cuarto término también puede verse diferente puesto que $\text{traza}(\mathbf{A}) = \text{traza}(\mathbf{A}')$.

$$\begin{aligned}\frac{\partial l_t}{\partial \psi_i} &= -\frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \right) + \frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t' \right) \\ &\quad - \frac{1}{2} \cdot \text{traza} \left(\frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \right) - \frac{1}{2} \cdot \text{traza} \left(\left(\mathbf{v}_t' \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \right)' \right)\end{aligned}$$

Dado que $\text{traza}(\mathbf{A}) + \text{traza}(\mathbf{B}) = \text{traza}(\mathbf{A} + \mathbf{B})$ podemos unir los dos primeros términos, así como también los dos últimos.

$$\begin{aligned} \frac{\partial l_t}{\partial \psi_i} = & -\frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} - \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}' \right) \\ & - \frac{1}{2} \cdot \text{traza} \left(\frac{\partial \mathbf{v}'_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t + \left(\mathbf{v}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \right)' \right) \end{aligned}$$

Tenemos $\left(\mathbf{v}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \right)' = \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \right)' \mathbf{v}_t = \frac{\partial \mathbf{v}'_t}{\partial \psi_i} (\mathbf{F}_t^{-1})' \mathbf{v}_t$. Como la matriz de covarianzas \mathbf{F}_t es simétrica, también lo es su inversa \mathbf{F}_t^{-1} y por ende $(\mathbf{F}_t^{-1})' = \mathbf{F}_t^{-1}$.

Luego $\left(\mathbf{v}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \right)' = \frac{\partial \mathbf{v}'_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t$ y obtenemos:

$$\frac{\partial l_t}{\partial \psi_i} = -\frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} - \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}' \right) - \text{traza} \left(\frac{\partial \mathbf{v}'_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \right)$$

Simplificando:

$$\frac{\partial l_t}{\partial \psi_i} = -\frac{1}{2} \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} (\mathbf{I} - \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}') \right) - \frac{\partial \mathbf{v}'_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \quad (2.3.7)$$

Sustituyendo el resultado anterior en (2.3.3) obtenemos el i -ésimo elemento del vector de puntaje:

$$\begin{aligned} \mathbf{S}_i(\boldsymbol{\psi}) = & -\frac{1}{2} \sum_{t=1}^N \left(\text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} (\mathbf{I} - \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}') \right) + 2 \frac{\partial \mathbf{v}'_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \right) \\ & i=1, \dots, p. \end{aligned} \quad (2.3.8)$$

2.3.2 La matriz hessiana y la matriz de información

La matriz hessiana

Sea $\mathbf{H}(\hat{\boldsymbol{\psi}}^k)$ la matriz de derivadas segundas de $\ln L$ con respecto a los elementos de $\boldsymbol{\psi}$, tomados en el estimador iterado más reciente $\hat{\boldsymbol{\psi}}^k$ de $\boldsymbol{\psi}$. Es decir, tenemos $\nabla^2 \ln L_{\boldsymbol{\psi}}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ evaluada en $\hat{\boldsymbol{\psi}}^k$. Llamaremos a $\mathbf{H}(\hat{\boldsymbol{\psi}}^k)$ *matriz hessiana en* $\hat{\boldsymbol{\psi}}^k$.

El (i, j) -ésimo elemento de la matriz $\mathbf{H}(\boldsymbol{\psi})$ es:

$$\mathbf{H}_{i,j}(\boldsymbol{\psi}) = \frac{\partial^2 \ln L}{\partial \psi_i \partial \psi_j} = \frac{\partial^2}{\partial \psi_i \partial \psi_j} \left(\sum_{t=1}^N l_t \right) = \sum_{t=1}^N \frac{\partial^2 l_t}{\partial \psi_i \partial \psi_j}$$

$i, j = 1, \dots, p.$

Diferenciaremos (2.3.7) respecto el j -ésimo elemento de $\boldsymbol{\psi}$ recordando que por (2.3.6) se tiene $\frac{\partial \mathbf{F}_t^{-1}}{\partial \psi_j} = -\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1}$. Dado que estamos derivando respecto escalares podemos introducir el operador de diferencias dentro de la función traza.

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \psi_i \partial \psi_j} &= -\frac{1}{2} \cdot \text{traza} \left(\frac{\partial}{\partial \psi_j} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} (\mathbf{I} - \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t') \right) \right) - \frac{\partial}{\partial \psi_j} \left(\frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{v}_t \right) \\ &= -\frac{1}{2} \cdot \text{traza} \left\{ \left(-\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} + \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \psi_i \partial \psi_j} \right) (\mathbf{I} - \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t') \right. \\ &\quad \left. + \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t' - \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_j} \mathbf{v}_t' - \mathbf{F}_t^{-1} \mathbf{v}_t \frac{\partial \mathbf{v}_t'}{\partial \psi_j} \right) \right\} \\ &\quad - \left(\frac{\partial^2 \mathbf{v}_t'}{\partial \psi_i \partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t - \frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t + \frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_j} \right) \end{aligned}$$

Reacomodando términos:

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \psi_i \partial \psi_j} &= -\frac{1}{2} \cdot \text{traza} \left(\left(\mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \psi_i \partial \psi_j} - \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \right) (\mathbf{I} - \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t') \right) \\ &\quad - \frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t' \right) \\ &\quad + \frac{1}{2} \cdot \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{v}_t}{\partial \psi_j} \mathbf{v}_t' + \mathbf{v}_t \frac{\partial \mathbf{v}_t'}{\partial \psi_j} \right) \right) \\ &\quad - \frac{\partial^2 \mathbf{v}_t'}{\partial \psi_i \partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t + \frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t - \frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_j} \end{aligned}$$

Finalmente

$$\begin{aligned} \mathbf{H}_{i,j}(\boldsymbol{\psi}) &= -\frac{1}{2} \sum_{t=1}^N \text{traza} \left(\left(\mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \psi_i \partial \psi_j} - \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \right) (\mathbf{I} - \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t') \right) \\ &\quad - \frac{1}{2} \sum_{t=1}^N \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t \mathbf{v}_t' \right) \end{aligned} \quad (2.3.9)$$

$$\begin{aligned}
& + \frac{1}{2} \sum_{t=1}^N \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{v}_t}{\partial \psi_j} \mathbf{v}_t' + \mathbf{v}_t \frac{\partial \mathbf{v}_t'}{\partial \psi_j} \right) \right) \\
& - \sum_{t=1}^N \frac{\partial^2 \mathbf{v}_t'}{\partial \psi_i \partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t + \sum_{t=1}^N \frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{v}_t - \sum_{t=1}^N \frac{\partial \mathbf{v}_t'}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_j}
\end{aligned}$$

Tanto el i -ésimo elemento del vector de puntaje, dado por (2.3.8), como el (i, j) -ésimo elemento de la matriz hessiana, presentado por (2.3.9), están en función de los vectores de primeras y segundas derivadas $\frac{\partial \mathbf{v}_t}{\partial \psi_i}$ y $\frac{\partial^2 \mathbf{v}_t}{\partial \psi_i \partial \psi_j}$ de $n \times 1$ y de las matrices de primeras y segundas derivadas $\frac{\partial \mathbf{F}_t}{\partial \psi_i}$ y $\frac{\partial^2 \mathbf{F}_t}{\partial \psi_i \partial \psi_j}$ de $n \times n$ para $i, j = 1, \dots, p$ y $t = 1, \dots, N$.

La matriz de información

Sea $\mathbf{I}(\boldsymbol{\psi}^k)$ la *información de Fisher sobre $\boldsymbol{\psi}^k$ en la muestra* o más comúnmente *matriz de información*, dada por:

$$\mathbf{I}(\boldsymbol{\psi}^k) = E(-\mathbf{H}(\boldsymbol{\psi}^k))$$

El (i, j) -ésimo elemento de la matriz de información es:

$$\mathbf{I}_{i,j}(\boldsymbol{\psi}) = E(-\mathbf{H}_{i,j}(\boldsymbol{\psi})) = E\left(-\sum_{t=1}^N \frac{\partial^2 l_t}{\partial \psi_i \partial \psi_j}\right) = -\sum_{t=1}^N E\left(\frac{\partial^2 l_t}{\partial \psi_i \partial \psi_j}\right) \quad (2.3.10)$$

$$i, j = 1, \dots, p.$$

Sustituyendo (2.3.9) en (2.3.10) obtenemos el (i, j) -ésimo elemento de la matriz de información. Para poder simplificar de antemano dicha expresión primero tengamos en cuenta que:¹

$$E(l_t) = E(E(l_t | \mathbf{Y}_{t-1}))$$

Si tomamos la esperanza de $\frac{\partial^2 l_t}{\partial \psi_i \partial \psi_j}$ condicionada en \mathbf{Y}_{t-1} las únicas variables aleatorias son los elementos del vector de innovación y sus derivadas primera y segunda. Sin embargo las derivadas son fijas respecto el operador esperanza en el instante $t-1$ ya que:

$$\mathbf{v}_t = \mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1}) \Rightarrow \frac{\partial \mathbf{v}_t}{\partial \psi_i} = -\frac{\partial}{\partial \psi_i} E(\mathbf{y}_t | \mathbf{Y}_{t-1})$$

¹ Harvey, *Forecasting, structural time series models and the Kalman filter*, p. 141.

La esperanza condicionada de \mathbf{v}_t es cero y por tanto:

$$E\left(\left(\frac{\partial \mathbf{v}'_t}{\partial \psi_i}\right) \mathbf{v}_t \middle| \mathbf{Y}_{t-1}\right) = \left(\frac{\partial \mathbf{v}'_t}{\partial \psi_i}\right) E(\mathbf{v}_t | \mathbf{Y}_{t-1}) = 0$$

De forma similar se puede obtener un resultado análogo para los términos que contienen a \mathbf{v}_t y a su derivada segunda. Por tanto los términos tercero, cuarto y quinto de (2.3.9) se anulan. El primer término desaparece ya que la esperanza condicionada de $\mathbf{v}_t \mathbf{v}'_t$ es \mathbf{F}_t y por la misma razón el segundo se simplifica. Entonces (2.3.10) queda:

$$\mathbf{I}_{i,j}(\boldsymbol{\psi}) = \frac{1}{2} \sum_{t=1}^N \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \right) + E \left(\sum_{t=1}^N \frac{\partial \mathbf{v}'_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_j} \right)$$

Si se elimina el operador esperanza del segundo término se obtiene una expresión que es asintóticamente equivalente con la anterior.

$$\mathbf{I}_{i,j}(\boldsymbol{\psi}) = \frac{1}{2} \sum_{t=1}^N \text{traza} \left(\mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \right) + \sum_{t=1}^N \frac{\partial \mathbf{v}'_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_j}$$

$i, j = 1, \dots, p.$

Vemos que también la matriz de información está en función de los vectores de derivadas $\frac{\partial \mathbf{v}_t}{\partial \psi_i}$ de $n \times 1$ y las matrices de derivadas $\frac{\partial \mathbf{F}_t}{\partial \psi_i}$ de $n \times n$ para $i=1, \dots, p$ y $t=1, \dots, N$.

2.3.3 Primeras derivadas del vector de innovación y su matriz de covarianzas respecto cada hiperparámetro

Como definimos en (2.2.3), el vector de innovaciones es:

$$\mathbf{v}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t = \mathbf{y}_t - \mathbf{C}_t \hat{\mathbf{x}}_{t|t-1} - \mathbf{d}_t$$

Por lo tanto:

$$\frac{\partial \mathbf{v}_t}{\partial \psi_i} = -\mathbf{C}_t \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_i} - \left(\frac{\partial \mathbf{C}_t}{\partial \psi_i} \right) \hat{\mathbf{x}}_{t|t-1} - \frac{\partial \mathbf{d}_t}{\partial \psi_i} \quad (2.3.11)$$

Como definimos en (2.2.5), la matriz de covarianzas del vector de innovaciones es:

$$\mathbf{F}_t = \mathbf{C}_t \mathbf{P}_{t|t-1} \mathbf{C}'_t + \mathbf{W}_t$$

Por lo tanto:

$$\frac{\partial \mathbf{F}_t}{\partial \psi_i} = \left(\frac{\partial \mathbf{C}_t}{\partial \psi_i} \right) \mathbf{P}_{t|t-1} \mathbf{C}'_t + \mathbf{C}_t \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} \right) \mathbf{C}'_t + \mathbf{C}_t \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'_t}{\partial \psi_i} + \frac{\partial \mathbf{W}_t}{\partial \psi_i} \quad (2.3.12)$$

La ecuación (2.3.11) requiere $\frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_i}$ y (2.3.12) necesita $\frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i}$.

Como definimos en (2.2.1) y (2.2.2), las ecuaciones de pronóstico son:

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{A}_t \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{b}_t$$

$$\mathbf{P}_{t|t-1} = \mathbf{A}_t \mathbf{P}_{t-1|t-1} \mathbf{A}'_t + \mathbf{V}_t$$

Por lo tanto:

$$\frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_i} = \left(\frac{\partial \mathbf{A}_t}{\partial \psi_i} \right) \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{A}_t \frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \psi_i} + \frac{\partial \mathbf{b}_t}{\partial \psi_i} \quad (2.3.13)$$

$$\frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} = \left(\frac{\partial \mathbf{A}_t}{\partial \psi_i} \right) \mathbf{P}_{t-1|t-1} \mathbf{A}'_t + \mathbf{A}_t \left(\frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \psi_i} \right) \mathbf{A}'_t + \mathbf{A}_t \mathbf{P}_{t-1|t-1} \frac{\partial \mathbf{A}'_t}{\partial \psi_i} + \frac{\partial \mathbf{V}_t}{\partial \psi_i} \quad (2.3.14)$$

La ecuación (2.3.13) requiere $\frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \psi_i}$ y (2.3.14) necesita $\frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \psi_i}$.

Como definimos en (2.2.6) y (2.2.7) con (2.2.4), las ecuaciones de actualización son:

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \mathbf{v}_t$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \mathbf{C}_t \mathbf{P}_{t|t-1}$$

Por lo tanto, aplicando la fórmula de (2.3.6) para la derivada de una matriz inversa tenemos:

$$\begin{aligned} \frac{\partial \hat{\mathbf{x}}_{t|t}}{\partial \psi_i} &= \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_i} + \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} \right) \mathbf{C}'_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{P}_{t|t-1} \left(\frac{\partial \mathbf{C}'_t}{\partial \psi_i} \right) \mathbf{F}_t^{-1} \mathbf{v}_t \\ &\quad - \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{F}_t}{\partial \psi_i} \right) \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{v}_t}{\partial \psi_i} \end{aligned} \quad (2.3.15)$$

$$\frac{\partial \mathbf{P}_{t|t}}{\partial \psi_i} = \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} \right) \mathbf{C}'_t \mathbf{F}_t^{-1} \mathbf{C}_t \mathbf{P}_{t|t-1} \quad (2.3.16)$$

$$\begin{aligned}
& -\mathbf{P}_{t|t-1} \left(\frac{\partial \mathbf{C}'_t}{\partial \psi_i} \right) \mathbf{F}_t^{-1} \mathbf{C}_t \mathbf{P}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{F}_t}{\partial \psi_i} \right) \mathbf{F}_t^{-1} \mathbf{C}_t \mathbf{P}_{t|t-1} \\
& -\mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{C}_t}{\partial \psi_i} \right) \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \mathbf{C}_t \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i}
\end{aligned}$$

Calculando (2.3.13), (2.3.14), (2.3.15) y (2.3.16) para $t=1, \dots, N$ obtenemos (2.3.11) y (2.3.12) en $t=1, \dots, N$. Los valores de partida dependen de los valores iniciales para el filtro de Kalman, pero si $\hat{\mathbf{x}}_{0|0}$ y $\mathbf{P}_{0|0}$ son independientes de $\boldsymbol{\psi}$ entonces:

$$\frac{\partial \hat{\mathbf{x}}_{0|0}}{\partial \psi_i} = 0 \text{ y } \frac{\partial \mathbf{P}_{0|0}}{\partial \psi_i} = 0$$

2.3.4 Segundas derivadas del vector de innovación y su matriz de covarianzas

De la sección anterior se siguen:

$$\begin{aligned}
\frac{\partial^2 \mathbf{v}_t}{\partial \psi_i \partial \psi_j} &= -\frac{\partial \mathbf{C}_t}{\partial \psi_j} \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_i} - \mathbf{C}_t \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_i \partial \psi_j} \\
&- \frac{\partial^2 \mathbf{C}_t}{\partial \psi_i \partial \psi_j} \hat{\mathbf{x}}_{t|t-1} - \frac{\partial \mathbf{C}_t}{\partial \psi_i} \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_j} - \frac{\partial^2 \mathbf{d}_t}{\partial \psi_i \partial \psi_j}
\end{aligned} \tag{2.3.17}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{F}_t}{\partial \psi_i \partial \psi_j} &= \frac{\partial^2 \mathbf{C}_t}{\partial \psi_i \partial \psi_j} \mathbf{P}_{t|t-1} \mathbf{C}' + \frac{\partial \mathbf{C}_t}{\partial \psi_i} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_j} \mathbf{C}' + \frac{\partial \mathbf{C}_t}{\partial \psi_i} \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'_t}{\partial \psi_j} \\
&+ \frac{\partial \mathbf{C}_t}{\partial \psi_j} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} \mathbf{C}' + \mathbf{C}_t \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \psi_i \partial \psi_j} \mathbf{C}' + \mathbf{C}_t \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} \frac{\partial \mathbf{C}'_t}{\partial \psi_j} \\
&+ \frac{\partial \mathbf{C}_t}{\partial \psi_j} \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'_t}{\partial \psi_i} + \mathbf{C}_t \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_j} \frac{\partial \mathbf{C}'_t}{\partial \psi_i} + \mathbf{C}_t \mathbf{P}_{t|t-1} \frac{\partial^2 \mathbf{C}'_t}{\partial \psi_i \partial \psi_j} + \frac{\partial^2 \mathbf{W}_t}{\partial \psi_i \partial \psi_j}
\end{aligned} \tag{2.3.18}$$

$$\begin{aligned}
\frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \psi_i \partial \psi_j} &= \frac{\partial^2 \mathbf{A}_t}{\partial \psi_i \partial \psi_j} \hat{\mathbf{x}}_{t-1|t-1} + \frac{\partial \mathbf{A}_t}{\partial \psi_i} \frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \psi_j} \\
&+ \frac{\partial \mathbf{A}_t}{\partial \psi_j} \frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \psi_i} + \mathbf{A}_t \frac{\partial^2 \hat{\mathbf{x}}_{t-1|t-1}}{\partial \psi_i \partial \psi_j} + \frac{\partial^2 \mathbf{b}_t}{\partial \psi_i \partial \psi_j}
\end{aligned} \tag{2.3.19}$$

$$\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \psi_i \partial \psi_j} = \frac{\partial^2 \mathbf{A}_t}{\partial \psi_i \partial \psi_j} \mathbf{P}_{t-1|t-1} \mathbf{A}'_t + \frac{\partial \mathbf{A}_t}{\partial \psi_i} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \psi_j} \mathbf{A}'_t + \frac{\partial \mathbf{A}_t}{\partial \psi_i} \mathbf{P}_{t-1|t-1} \frac{\partial \mathbf{A}'_t}{\partial \psi_j} \tag{2.3.20}$$

$$\begin{aligned}
& -\mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{C}_t \mathbf{P}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \frac{\partial \mathbf{C}_t}{\partial \psi_j} \mathbf{P}_{t|t-1} \\
& + \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_i} \mathbf{F}_t^{-1} \mathbf{C}_t \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_j} - \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_j} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{C}_t}{\partial \psi_i} \mathbf{P}_{t|t-1} \\
& - \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'_t}{\partial \psi_j} \mathbf{F}_t^{-1} \frac{\partial \mathbf{C}_t}{\partial \psi_i} \mathbf{P}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \frac{\partial \mathbf{C}_t}{\partial \psi_i} \mathbf{P}_{t|t-1} \\
& - \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{C}_t}{\partial \psi_i \partial \psi_j} \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{C}_t}{\partial \psi_i} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_j} \\
& - \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_j} \mathbf{C}'_t \mathbf{F}_t^{-1} \mathbf{C}_t \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} - \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{C}_t \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} \\
& + \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \psi_j} \mathbf{F}_t^{-1} \mathbf{C}_t \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} - \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \frac{\partial \mathbf{C}_t}{\partial \psi_j} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \psi_i} \\
& - \mathbf{P}_{t|t-1} \mathbf{C}'_t \mathbf{F}_t^{-1} \mathbf{C}_t \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \psi_i \partial \psi_j}
\end{aligned}$$

Capítulo 3: Un espacio de estados para el modelo de Vasicek (1977) de un factor y estimación de sus hiperparámetros

3.1 Matrices del sistema

A continuación presentamos un espacio de estados para el modelo de estructura temporal de tipos de interés de Vasicek (1977) de un factor. Como veremos, las matrices de la ecuación de transición, así como las de la ecuación de medidas son invariantes en el tiempo y por lo tanto se omite el subíndice t en su notación.

El vector de medidas u observaciones es:

$$\mathbf{y}_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{pmatrix} = \mathbf{C} \mathbf{x}_t + \mathbf{d} + \mathbf{w}_t \quad (3.1.1)$$

donde cada observación $y_{i,t}$ con $t=1, \dots, N$ e $i=1, \dots, n$ es el rendimiento que el bono cupón cero con plazo al vencimiento τ_i ofrece el día t . En total tendremos $N=1377$ observaciones tomadas directamente de los registros del mercado de valores mexicano del 16/05/2003 al 21/10/2008 para $n=4$ diferentes tiempos al vencimiento: $\tau_1 = \frac{28}{364}$, $\tau_2 = \frac{91}{364}$, $\tau_3 = \frac{182}{364}$ y $\tau_4 = \frac{364}{364}$. Los valores de τ_i son así porque contamos el tiempo en años y los bonos de un año tienen 364 días de vigencia.

El vector de estado o variable desconocida es el escalar tasa de interés de corto plazo r_t :

$$\mathbf{x}_t = r_t = \mathbf{A} \mathbf{x}_{t-1} + \mathbf{b} + \mathbf{v}_t \quad (3.1.2)$$

Sustituimos la fórmula (1.6.2) del precio de un bono cupón cero en la fórmula (1.1.2) del rendimiento de un bono cupón cero en términos del tiempo τ_i que falta para el vencimiento del instrumento y lo adaptamos para ser el i -ésimo elemento de (3.1.1), de forma que obtenemos:

$$y_{i,t} = -\frac{F(\tau_i)r_t + G(\tau_i)}{\tau_i} + w_{i,t} \quad (3.1.3)$$

Por (3.1.1), (3.1.3) y sabiendo que la dimensión de nuestro vector de observaciones es de 4×1 tenemos:

$$\mathbf{y}_t = \begin{pmatrix} -\frac{F(\tau_1)r_t + G(\tau_1)}{\tau_1} + w_{i,t} \\ -\frac{F(\tau_2)r_t + G(\tau_2)}{\tau_2} + w_{i,t} \\ -\frac{F(\tau_3)r_t + G(\tau_3)}{\tau_3} + w_{i,t} \\ -\frac{F(\tau_4)r_t + G(\tau_4)}{\tau_4} + w_{i,t} \end{pmatrix} = \begin{pmatrix} -\frac{F(\tau_1)}{\tau_1} \\ -\frac{F(\tau_2)}{\tau_2} \\ -\frac{F(\tau_3)}{\tau_3} \\ -\frac{F(\tau_4)}{\tau_4} \end{pmatrix} r_t + \begin{pmatrix} -\frac{G(\tau_1)}{\tau_1} \\ -\frac{G(\tau_2)}{\tau_2} \\ -\frac{G(\tau_3)}{\tau_3} \\ -\frac{G(\tau_4)}{\tau_4} \end{pmatrix} + \mathbf{w}_t$$

Dado que \mathbf{x}_t es r_t , es inmediato pensar en las siguientes matrices del vector de observaciones:

$$\mathbf{C} = \begin{pmatrix} -\frac{F(\tau_1)}{\tau_1} \\ -\frac{F(\tau_2)}{\tau_2} \\ -\frac{F(\tau_3)}{\tau_3} \\ -\frac{F(\tau_4)}{\tau_4} \end{pmatrix}$$

$$\mathbf{d} = \begin{pmatrix} -\frac{G(\tau_1)}{\tau_1} \\ -\frac{G(\tau_2)}{\tau_2} \\ -\frac{G(\tau_3)}{\tau_3} \\ -\frac{G(\tau_4)}{\tau_4} \end{pmatrix}$$

Note que las dos expresiones anteriores son invariantes en el tiempo porque dependen de cada τ_i fija en vez de t . Los valores $F(\tau_i)$ y $G(\tau_i)$ son como se definió en (1.6.5) y (1.6.7). Luego:

$$\mathbf{C} = \begin{pmatrix} \frac{1 - \exp(-\kappa \tau_1)}{\kappa \tau_1} \\ \frac{1 - \exp(-\kappa \tau_2)}{\kappa \tau_2} \\ \frac{1 - \exp(-\kappa \tau_3)}{\kappa \tau_3} \\ \frac{1 - \exp(-\kappa \tau_4)}{\kappa \tau_4} \end{pmatrix} \quad (3.1.4)$$

$$\mathbf{d} = \begin{pmatrix} \frac{\sigma^2}{4\kappa\tau_1} \left(\frac{1 - \exp(-\kappa\tau_1)}{\kappa} \right)^2 + \left(\mu - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2} \right) \left(1 + \frac{\exp(-\kappa\tau_1) - 1}{\kappa\tau_1} \right) \\ \frac{\sigma^2}{4\kappa\tau_2} \left(\frac{1 - \exp(-\kappa\tau_2)}{\kappa} \right)^2 + \left(\mu - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2} \right) \left(1 + \frac{\exp(-\kappa\tau_2) - 1}{\kappa\tau_2} \right) \\ \frac{\sigma^2}{4\kappa\tau_3} \left(\frac{1 - \exp(-\kappa\tau_3)}{\kappa} \right)^2 + \left(\mu - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2} \right) \left(1 + \frac{\exp(-\kappa\tau_3) - 1}{\kappa\tau_3} \right) \\ \frac{\sigma^2}{4\kappa\tau_4} \left(\frac{1 - \exp(-\kappa\tau_4)}{\kappa} \right)^2 + \left(\mu - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2} \right) \left(1 + \frac{\exp(-\kappa\tau_4) - 1}{\kappa\tau_4} \right) \end{pmatrix} \quad (3.1.5)$$

Supondremos que el ruido \mathbf{w}_t , debido a errores en la captura de información, es un valor gaussiano con media cero y covarianzas pequeñas y constantes, por ejemplo:

$$\mathbf{W} = \begin{pmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 0.01 \end{pmatrix} \quad (3.1.6)$$

Basándonos en las ecuaciones (1.4.4) a (1.4.6) tenemos las siguientes matrices de la ecuación de transición invariantes en el tiempo:

$$\mathbf{A} = e^{-\kappa(t-s)} \quad (3.1.7)$$

$$\mathbf{b} = \left(\mu - \frac{\lambda\sigma}{\kappa} \right) (1 - e^{-\kappa(t-s)}) \quad (3.1.8)$$

$$\mathbf{V} = \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa(t-s)}) \quad (3.1.9)$$

Por lo que vimos en el subcapítulo 1.4 la perturbación \mathbf{v}_t de nuestro vector de estados es gaussiana y tiene media cero.

Note que tanto \mathbf{v}_t como \mathbf{w}_t son *ruidos blancos*.

Tenemos $t - s = \frac{1}{250}$ pues existen aproximadamente 250 días de negociación en un año.

Dadas las dimensiones de las matrices del sistema, tendremos que $\mathbf{x}_{t|t-1}$, $\mathbf{P}_{t|t-1}$, $\mathbf{x}_{t|t}$ y $\mathbf{P}_{t|t}$ son de dimensión escalar, \mathbf{v}_t es de 4×1 , \mathbf{K}_t es de 1×4 y \mathbf{F}_t es de 4×4 .

La matriz \mathbf{W} propuesta es claramente definida positiva. Por su parte, \mathbf{V} , $\mathbf{P}_{t|t-1}$ y $\mathbf{P}_{t|t}$ son escalares y su inversa claramente existe, evitándonos tener que revisar que sean matrices triangulares superiores de Cholesky.

3.2 Condiciones iniciales del espacio de estados

Veamos si nuestro vector de estado cumple (2.1.5) siendo por ende estacionario. Dado que \mathbf{A} es escalar sólo hay una raíz característica $\lambda(\mathbf{A})$ cuyo polinomio característico por resolver también es escalar:

$$p(\lambda) = |\mathbf{A} - \lambda|$$

Sustituyendo (3.1.7) en lo anterior queda:

$$p(\lambda) = |e^{-\kappa(t-s)} - \lambda|$$

Para resolverlo se hace:

$$|e^{-\kappa(t-s)} - \lambda| = 0$$

Recordemos que $\forall a \in \mathfrak{R}$ se tiene

$$|a| = \begin{cases} a & \text{si } a > 0 \\ -a & \text{si } a \leq 0 \end{cases}$$

Análogamente:

$$|e^{-\kappa(t-s)} - \lambda| = \begin{cases} e^{-\kappa(t-s)} - \lambda & \text{si } e^{-\kappa(t-s)} > \lambda \\ \lambda - e^{-\kappa(t-s)} & \text{si } e^{-\kappa(t-s)} \leq \lambda \end{cases}$$

Puede ser que $e^{-\kappa(t-s)} > \lambda \geq 1$ o bien $e^{-\kappa(t-s)} \leq \lambda < 1$. Para asegurar la estacionariedad de r_t debe cumplirse este último caso, que viene siendo:

$$\begin{aligned} e^{-\kappa(t-s)} < 1 &\Rightarrow 1 < e^{\kappa(t-s)} \Rightarrow \ln 1 < \kappa(t-s) \\ &\Rightarrow 0 < \kappa(t-s) \Rightarrow 0 < \kappa \end{aligned}$$

Por lo tanto para que r_t sea estacionario se requiere κ positivo. En el subcapítulo 1.3 expusimos que el parámetro κ del modelo de Vasicek se define positivo. De tal forma (3.1.2) es estacionario.

Dado que además \mathbf{A} tiene inversa pues es escalar podemos aplicar (2.1.6) y (2.1.7) para obtener la media y la matriz de covarianzas del vector de estado inicial \mathbf{x}_0 :

$$\mathbf{x}_{0|0} = \mu - \frac{\lambda \sigma}{\kappa} \tag{3.2.1}$$

$$\text{vec}(\mathbf{P}_{0|0}) = \frac{\sigma^2}{2\kappa} \quad (3.2.2)$$

Estimación de los hiperparámetros

Recapitulando, tenemos el espacio de estados descrito por las ecuaciones (3.1.2) y (3.1.1), con condiciones iniciales (3.2.1) y (3.2.2). Dicho espacio de estados es invariante en el tiempo porque tanto las matrices (3.1.7), (3.1.8) y (3.1.9) de su ecuación de transición, como las matrices (3.1.4), (3.1.5) y (3.1.6) del vector de observaciones permanecen invariantes en el tiempo. Como pudimos ver, tanto las condiciones iniciales como las matrices del sistema adquieren su valor en función del conjunto de cuatro

hiperparámetros ψ_i , que escritos dentro un vector son $\boldsymbol{\psi} = \begin{pmatrix} \mu \\ \sigma \\ \kappa \\ \lambda \end{pmatrix}$.

Sabemos que las ecuaciones del filtro de Kalman proporcionarán en cada t los valores necesarios para evaluar al final¹ una función de verosimilitud del vector de observaciones en forma de descomposición del error de pronóstico. Podremos resolver nuestra función de verosimilitud aplicando algún método numérico recursivo de optimización no lineal sin restricciones, es decir, podremos encontrar el valor de los hiperparámetros. El vector de puntaje y la matriz hessiana son dos elementos básicos para aplicar varios métodos posibles.²

Dado que tenemos cuatro hiperparámetros, el vector de puntaje será de dimensiones 4×1 y la matriz de Hess de 4×4 . Hemos dicho que para obtener el vector de puntaje y la matriz hessiana se vuelve necesario disponer de las primeras y segundas derivadas del vector de innovación y su matriz de covarianzas, respecto cada hiperparámetro. Las derivadas anteriores requieren conocer las primeras y segundas derivadas del vector de estado y su matriz de covarianzas, adelantados y actualizados, respecto cada hiperparámetro. También se necesitan de las primeras y segundas derivadas de nuestras matrices del sistema respecto cada hiperparámetro. A continuación calcularemos las derivadas de las matrices del sistema que tenemos respecto cada hiperparámetro. En lo que resta del capítulo 3 utilizaremos tanto esos resultados como los provenientes de las secciones 2.3.3 y 2.3.4 del capítulo anterior para presentar las primeras y segundas derivadas respecto cada hiperparámetro de nuestros vectores de estado y sus matrices de covarianza, adelantados y actualizados, el error de pronóstico y su matriz de covarianzas. Se incluyen las derivadas de las condiciones iniciales.

¹ Después de procesarse el N –ésimo dato.

² Puede consultar el apéndice E. Contiene cuatro métodos de optimización para funciones diferenciables no lineales de varias variables en problemas no restringidos: El Método de Newton, el DFP, el Fletcher-Reeves y el BFGS.

3.3 Primeras derivadas de las matrices del sistema respecto el i-ésimo hiperparámetro

Derivaremos: $\frac{\partial \mathbf{A}}{\partial \psi_i}$, $\frac{\partial \mathbf{b}}{\partial \psi_i}$, $\frac{\partial \mathbf{V}}{\partial \psi_i}$, $\frac{\partial \mathbf{C}}{\partial \psi_i}$, $\frac{\partial \mathbf{d}}{\partial \psi_i}$ y $\frac{\partial \mathbf{W}}{\partial \psi_i}$ para $\psi_1 = \mu$, $\psi_2 = \sigma$, $\psi_3 = \kappa$ y $\psi_4 = \lambda$.

$$\frac{\partial \mathbf{A}}{\partial \mu} = \frac{\partial \mathbf{A}}{\partial \sigma} = \frac{\partial \mathbf{A}}{\partial \lambda} = 0$$

$$\frac{\partial \mathbf{A}}{\partial \kappa} = -(t-s)e^{-\kappa(t-s)}$$

$$\frac{\partial \mathbf{b}}{\partial \mu} = 1 - e^{-\kappa(t-s)}$$

$$\frac{\partial \mathbf{b}}{\partial \sigma} = \frac{\lambda}{\kappa} (e^{-\kappa(t-s)} - 1)$$

$$\frac{\partial \mathbf{b}}{\partial \kappa} = \frac{\lambda \sigma}{\kappa^2} (1 - e^{-\kappa(t-s)}) + (t-s) \left(\mu - \frac{\lambda \sigma}{\kappa} \right) e^{-\kappa(t-s)}$$

$$\frac{\partial \mathbf{b}}{\partial \lambda} = \frac{\sigma}{\kappa} (e^{-\kappa(t-s)} - 1)$$

$$\frac{\partial \mathbf{V}}{\partial \mu} = \frac{\partial \mathbf{V}}{\partial \lambda} = 0$$

$$\frac{\partial \mathbf{V}}{\partial \sigma} = \frac{\sigma}{\kappa} (1 - e^{-2\kappa(t-s)})$$

$$\frac{\partial \mathbf{V}}{\partial \kappa} = \frac{\sigma^2}{2\kappa^2} ((2\kappa(t-s) + 1)e^{-2\kappa(t-s)} - 1)$$

El i -ésimo elemento de cada uno de los vectores $\frac{\partial \mathbf{C}}{\partial \mu}$, $\frac{\partial \mathbf{C}}{\partial \sigma}$ o $\frac{\partial \mathbf{C}}{\partial \lambda}$ de 4×1 es:

$$\frac{\partial C_i}{\partial \mu} = \frac{\partial C_i}{\partial \sigma} = \frac{\partial C_i}{\partial \lambda} = 0$$

El i -ésimo elemento del vector $\frac{\partial \mathbf{C}}{\partial \kappa}$ de 4×1 es:

$$\frac{\partial C_i}{\partial \kappa} = \frac{(\kappa \tau_i + 1) \cdot \exp(-\kappa \tau_i) - 1}{\tau_i \kappa^2}$$

El i -ésimo elemento del vector $\frac{\partial \mathbf{d}}{\partial \mu}$ de 4×1 es:

$$\frac{\partial d_i}{\partial \mu} = 1 + \frac{\exp(-\kappa \tau_i) - 1}{\kappa \tau_i}$$

El i -ésimo elemento del vector $\frac{\partial \mathbf{d}}{\partial \sigma}$ de 4×1 es:

$$\frac{\partial d_i}{\partial \sigma} = \frac{\sigma}{2\kappa \tau_i} \left(\frac{1 - \exp(-\kappa \tau_i)}{\kappa} \right)^2 + \left(\frac{\lambda}{\kappa} + \frac{\sigma}{\kappa^2} \right) \left(\frac{1 - \exp(-\kappa \tau_i)}{\kappa \tau_i} - 1 \right)$$

El i -ésimo elemento del vector $\frac{\partial \mathbf{d}}{\partial \kappa}$ de 4×1 es:

$$\begin{aligned} \frac{\partial d_i}{\partial \kappa} &= \left(\mu - \frac{\lambda \sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2} \right) \left(\frac{1 - (\kappa \tau_i + 1) \exp(-\kappa \tau_i)}{\tau_i \kappa^2} \right) \\ &+ \frac{\sigma^2}{4\kappa^4 \tau_i} (1 - \exp(-\kappa \tau_i)) (2\kappa \tau_i \exp(-\kappa \tau_i) + 3 \exp(-\kappa \tau_i) - 3) \\ &+ \left(\frac{\lambda \sigma}{\kappa^2} + \frac{\sigma^2}{\kappa^3} \right) \left(1 + \frac{\exp(-\kappa \tau_i) - 1}{\kappa \tau_i} \right) \end{aligned}$$

El i -ésimo elemento del vector $\frac{\partial \mathbf{d}}{\partial \lambda}$ de 4×1 es:

$$\frac{\partial d_i}{\partial \lambda} = \frac{\sigma}{\kappa} \left(\frac{1 - \exp(-\kappa \tau_i)}{\kappa \tau_i} - 1 \right)$$

El i -ésimo elemento de cada una de las matrices $\frac{\partial \mathbf{W}}{\partial \mu}$, $\frac{\partial \mathbf{W}}{\partial \sigma}$, $\frac{\partial \mathbf{W}}{\partial \kappa}$ o $\frac{\partial \mathbf{W}}{\partial \lambda}$ de 4×4 es:

$$\frac{\partial W_i}{\partial \mu} = \frac{\partial W_i}{\partial \sigma} = \frac{\partial W_i}{\partial \kappa} = \frac{\partial W_i}{\partial \lambda} = 0$$

3.4 Primeras derivadas del vector de estado y covarianzas, adelantados y actualizados respecto al i-ésimo hiperparámetro

Algunas de las derivadas de la sección 3.3 valen cero. Si sustituimos estos casos en (2.3.13), (2.3.14), (2.3.15) y (2.3.16) para $\psi_1 = \mu$, $\psi_2 = \sigma$, $\psi_3 = \kappa$ y $\psi_4 = \lambda$ obtenemos:

$$\frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \mu} = \mathbf{A} \frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \mu} + \frac{\partial \mathbf{b}}{\partial \mu}$$

$$\frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \sigma} = \mathbf{A} \frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \sigma} + \frac{\partial \mathbf{b}}{\partial \sigma}$$

$$\frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \kappa} = \frac{\partial \mathbf{A}}{\partial \kappa} \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{A} \frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \kappa} + \frac{\partial \mathbf{b}}{\partial \kappa}$$

$$\frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \lambda} = \mathbf{A} \frac{\partial \hat{\mathbf{x}}_{t-1|t-1}}{\partial \lambda} + \frac{\partial \mathbf{b}}{\partial \lambda}$$

$$\frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} = \mathbf{A} \left(\frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \mu} \right) \mathbf{A}'$$

$$\frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} = \mathbf{A} \left(\frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \sigma} \right) \mathbf{A}' + \frac{\partial \mathbf{V}}{\partial \sigma}$$

$$\begin{aligned} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} &= \frac{\partial \mathbf{A}}{\partial \kappa} \mathbf{P}_{t-1|t-1} \mathbf{A}' + \mathbf{A} \left(\frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \kappa} \right) \mathbf{A}' \\ &\quad + \mathbf{A} \mathbf{P}_{t-1|t-1} \frac{\partial \mathbf{A}'}{\partial \kappa} + \frac{\partial \mathbf{V}}{\partial \kappa} \end{aligned}$$

$$\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} = \mathbf{A} \left(\frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \lambda} \right) \mathbf{A}'$$

$$\begin{aligned} \frac{\partial \hat{\mathbf{x}}_{t|t}}{\partial \mu} &= \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \mu} + \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \right) \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{v}_t \\ &\quad - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \mu} \right) \mathbf{F}_t^{-1} \mathbf{v}_t - \frac{\partial \mathbf{v}_t}{\partial \mu} \right) \end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{x}_{t|t}}{\partial \sigma} &= \frac{\partial \mathbf{x}_{t|t-1}}{\partial \sigma} + \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \right) \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{v}_t \\ &\quad - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \sigma} \right) \mathbf{F}_t^{-1} \mathbf{v}_t - \frac{\partial \mathbf{v}_t}{\partial \sigma} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{x}_{t|t}}{\partial \kappa} &= \frac{\partial \mathbf{x}_{t|t-1}}{\partial \kappa} + \left(\left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \right) \mathbf{C}' + \mathbf{P}_{t|t-1} \left(\frac{\partial \mathbf{C}'}{\partial \kappa} \right) \right) \mathbf{F}_t^{-1} \mathbf{v}_t \\ &\quad - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \kappa} \right) \mathbf{F}_t^{-1} \mathbf{v}_t - \frac{\partial \mathbf{v}_t}{\partial \kappa} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{x}_{t|t}}{\partial \lambda} &= \frac{\partial \mathbf{x}_{t|t-1}}{\partial \lambda} + \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \right) \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{v}_t \\ &\quad - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \lambda} \right) \mathbf{F}_t^{-1} \mathbf{v}_t - \frac{\partial \mathbf{v}_t}{\partial \lambda} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{P}_{t|t}}{\partial \mu} &= \frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \right) \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\ &\quad + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \mu} \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} - \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{P}_{t|t}}{\partial \sigma} &= \frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \right) \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\ &\quad + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \sigma} \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} - \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{P}_{t|t}}{\partial \kappa} &= \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} - \left(\left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \right) \mathbf{C}' + \mathbf{P}_{t|t-1} \left(\frac{\partial \mathbf{C}'}{\partial \kappa} \right) \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\ &\quad + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \kappa} \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} - \left(\frac{\partial \mathbf{C}}{\partial \kappa} \right) \mathbf{P}_{t|t-1} - \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{P}_{t|t}}{\partial \lambda} &= \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \right) \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\ &\quad + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\left(\frac{\partial \mathbf{F}_t}{\partial \lambda} \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} - \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \right)\end{aligned}$$

Aplicamos (3.2.1) y (3.2.2) y obtenemos las condiciones iniciales de las recursiones de esta sección:

$$\frac{\partial \mathbf{x}_{0|0}}{\partial \mu} = 1$$

$$\frac{\partial \mathbf{x}_{0|0}}{\partial \sigma} = -\frac{\lambda}{\kappa}$$

$$\frac{\partial \mathbf{x}_{0|0}}{\partial \kappa} = \frac{\lambda \sigma}{\kappa^2}$$

$$\frac{\partial \mathbf{x}_{0|0}}{\partial \lambda} = -\frac{\sigma}{\kappa}$$

$$\frac{\partial \mathbf{P}_{0|0}}{\partial \mu} = \frac{\partial \mathbf{P}_{0|0}}{\partial \lambda} = 0$$

$$\frac{\partial \mathbf{P}_{0|0}}{\partial \sigma} = \frac{\sigma}{\kappa}$$

$$\frac{\partial \mathbf{P}_{0|0}}{\partial \kappa} = -\frac{\sigma^2}{2\kappa^2}$$

3.5 Primeras derivadas del error de pronóstico y su matriz de covarianzas respecto el i-ésimo hiperparámetro

Algunas de las derivadas de la sección 3.3 valen cero. Si sustituimos estos casos en (2.3.11) y (2.3.12) para $\psi_1 = \mu$, $\psi_2 = \sigma$, $\psi_3 = \kappa$ y $\psi_4 = \lambda$ obtenemos:

$$\frac{\partial \mathbf{v}_t}{\partial \mu} = -\mathbf{C} \frac{\partial \mathbf{x}_{t|t-1}}{\partial \mu} - \frac{\partial \mathbf{d}}{\partial \mu}$$

$$\frac{\partial \mathbf{v}_t}{\partial \sigma} = -\mathbf{C} \frac{\partial \mathbf{x}_{t|t-1}}{\partial \sigma} - \frac{\partial \mathbf{d}}{\partial \sigma}$$

$$\frac{\partial \mathbf{v}_t}{\partial \kappa} = -\mathbf{C} \frac{\partial \mathbf{x}_{t|t-1}}{\partial \kappa} - \left(\frac{\partial \mathbf{C}}{\partial \kappa} \right) \mathbf{x}_{t|t-1} - \frac{\partial \mathbf{d}}{\partial \kappa}$$

$$\frac{\partial \mathbf{v}_t}{\partial \lambda} = -\mathbf{C} \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \lambda} - \frac{\partial \mathbf{d}}{\partial \lambda}$$

$$\frac{\partial \mathbf{F}_t}{\partial \mu} = \mathbf{C} \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \right) \mathbf{C}'$$

$$\frac{\partial \mathbf{F}_t}{\partial \sigma} = \mathbf{C} \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \right) \mathbf{C}'$$

$$\frac{\partial \mathbf{F}_t}{\partial \kappa} = \left(\frac{\partial \mathbf{C}}{\partial \kappa} \right) \mathbf{P}_{t|t-1} \mathbf{C}' + \mathbf{C} \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \right) \mathbf{C}' + \mathbf{C} \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'}{\partial \kappa}$$

$$\frac{\partial \mathbf{F}_t}{\partial \lambda} = \mathbf{C} \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \right) \mathbf{C}'$$

3.6 Segundas derivadas de las matrices del sistema

$$\frac{\partial^2 \mathbf{A}}{\partial \psi_i \partial \psi_j} = 0 \text{ si } \psi_i \text{ o } \psi_j \text{ son } \mu, \sigma \text{ o } \lambda.$$

$$\frac{\partial^2 \mathbf{A}}{\partial \kappa^2} = (t-s)^2 e^{-\kappa(t-s)}$$

$$\frac{\partial^2 \mathbf{b}}{\partial \mu^2} = \frac{\partial^2 \mathbf{b}}{\partial \mu \partial \sigma} = \frac{\partial^2 \mathbf{b}}{\partial \mu \partial \lambda} = \frac{\partial^2 \mathbf{b}}{\partial \sigma \partial \mu} = \frac{\partial^2 \mathbf{b}}{\partial \sigma^2} = \frac{\partial^2 \mathbf{b}}{\partial \lambda \partial \mu} = \frac{\partial^2 \mathbf{b}}{\partial \lambda^2} = 0$$

$$\frac{\partial^2 \mathbf{b}}{\partial \mu \partial \kappa} = \frac{\partial^2 \mathbf{b}}{\partial \kappa \partial \mu} = (t-s) e^{-\kappa(t-s)}$$

$$\frac{\partial^2 \mathbf{b}}{\partial \sigma \partial \kappa} = \frac{\partial^2 \mathbf{b}}{\partial \kappa \partial \sigma} = \frac{\lambda}{\kappa^2} (1 - e^{-\kappa(t-s)}) - \frac{\lambda}{\kappa} (t-s) e^{-\kappa(t-s)}$$

$$\frac{\partial^2 \mathbf{b}}{\partial \sigma \partial \lambda} = \frac{\partial^2 \mathbf{b}}{\partial \lambda \partial \sigma} = \frac{1}{\kappa} (e^{-\kappa(t-s)} - 1)$$

$$\frac{\partial^2 \mathbf{b}}{\partial \kappa^2} = -\frac{2\lambda\sigma}{\kappa^3} (1 - e^{-\kappa(t-s)}) + \frac{\lambda\sigma}{\kappa^2} (t-s) e^{-\kappa(t-s)}$$

$$-(t-s)^2 \mu e^{-\kappa(t-s)} + (t-s) \frac{\lambda \sigma}{\kappa^2} e^{-\kappa(t-s)} + (t-s)^2 \frac{\lambda \sigma}{\kappa} e^{-\kappa(t-s)}$$

$$\frac{\partial^2 \mathbf{b}}{\partial \kappa \partial \lambda} = \frac{\partial^2 \mathbf{b}}{\partial \lambda \partial \kappa} = \frac{\sigma}{\kappa^2} (1 - e^{-\kappa(t-s)}) - \frac{\sigma}{\kappa} (t-s) e^{-\kappa(t-s)}$$

$$\frac{\partial^2 \mathbf{V}}{\partial \psi_i \partial \psi_j} = 0 \text{ si } \psi_i \text{ o } \psi_j \text{ son } \mu \text{ o } \lambda.$$

$$\frac{\partial^2 \mathbf{V}}{\partial \sigma^2} = \frac{1}{\kappa} (1 - e^{-2\kappa(t-s)})$$

$$\frac{\partial^2 \mathbf{V}}{\partial \sigma \partial \kappa} = \frac{\partial^2 \mathbf{V}}{\partial \kappa \partial \sigma} = \frac{\sigma}{\kappa^2} (e^{-2\kappa(t-s)} - 1) + \frac{2\sigma}{\kappa} (t-s) e^{-2\kappa(t-s)}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{V}}{\partial \kappa^2} &= -\frac{\sigma^2}{\kappa^3} ((2\kappa(t-s)+1)e^{-2\kappa(t-s)} - 1) + \frac{\sigma^2}{\kappa^2} (t-s) e^{-2\kappa(t-s)} \\ &\quad - \frac{\sigma^2}{\kappa^2} (t-s)(2\kappa(t-s)+1)e^{-2\kappa(t-s)} \end{aligned}$$

$$\frac{\partial^2 \mathbf{C}}{\partial \psi_i \partial \psi_j} = \mathbf{0} \text{ si } \psi_i \text{ o } \psi_j \text{ son } \mu, \sigma \text{ o } \lambda.$$

$$\frac{\partial^2 \mathbf{C}}{\partial \kappa^2} = -\frac{\tau_i \exp(-\kappa \tau_i)}{\kappa} - \frac{2((\kappa \tau_i + 1) \cdot \exp(-\kappa \tau_i) - 1)}{\tau_i \kappa^3}$$

$$\frac{\partial^2 \mathbf{d}}{\partial \mu^2} = \frac{\partial^2 \mathbf{d}}{\partial \mu \partial \sigma} = \frac{\partial^2 \mathbf{d}}{\partial \mu \partial \lambda} = \frac{\partial^2 \mathbf{d}}{\partial \sigma \partial \mu} = \frac{\partial^2 \mathbf{d}}{\partial \lambda \partial \mu} = \frac{\partial^2 \mathbf{d}}{\partial \lambda^2} = 0$$

$$\frac{\partial^2 \mathbf{d}}{\partial \mu \partial \kappa} = \frac{\partial^2 \mathbf{d}}{\partial \kappa \partial \mu} = \frac{1 - (\kappa \tau_i + 1) \exp(-\kappa \tau_i)}{\kappa^2 \tau_i}$$

$$\frac{\partial^2 \mathbf{d}}{\partial \sigma^2} = \frac{1}{2\kappa \tau_i} \left(\frac{1 - \exp(-\kappa \tau_i)}{\kappa} \right)^2 + \frac{1}{\kappa^2} \left(\frac{1 - \exp(-\kappa \tau_i)}{\kappa \tau_i} - 1 \right)$$

$$\frac{\partial^2 \mathbf{d}}{\partial \sigma \partial \kappa} = \frac{\partial^2 \mathbf{d}}{\partial \kappa \partial \sigma} =$$

$$= \frac{\sigma}{2\kappa^4 \tau_i} (1 - \exp(-\kappa \tau_i)) (2\kappa \tau_i \exp(-\kappa \tau_i) + 3 \exp(-\kappa \tau_i) - 3)$$

$$+ \left(\frac{\lambda}{\kappa^2} + \frac{2\sigma}{\kappa^3} \right) \left(1 + \frac{\exp(-\kappa \tau_i) - 1}{\kappa \tau_i} \right) + \left(\frac{\lambda}{\kappa} + \frac{\sigma}{\kappa^2} \right) \left(\frac{(\kappa \tau_i + 1) \exp(-\kappa \tau_i) - 1}{\kappa^2 \tau_i} \right)$$

$$\frac{\partial^2 \mathbf{d}}{\partial \sigma \partial \lambda} = \frac{\partial^2 \mathbf{d}}{\partial \lambda \partial \sigma} = \frac{1}{\kappa} \left(\frac{1 - \exp(-\kappa \tau_i)}{\kappa \tau_i} - 1 \right)$$

$$\begin{aligned} \frac{\partial^2 \mathbf{d}}{\partial \kappa^2} &= \left(\frac{\lambda \sigma}{\kappa^2} + \frac{\sigma^2}{\kappa^3} \right) \left(\frac{1 - (\kappa \tau_i + 1) \exp(-\kappa \tau_i)}{\tau_i \kappa^2} \right) \\ &+ \left(\mu - \frac{\lambda \sigma}{\kappa} - \frac{\sigma^2}{2\kappa^2} \right) \cdot \left(\frac{\tau_i \exp(-\kappa \tau_i)}{\kappa} + \frac{2((\kappa \tau_i + 1) \exp(-\kappa \tau_i) - 1)}{\tau_i \kappa^3} \right) \\ &+ \left(\frac{\sigma^2}{4\kappa^4} \exp(-\kappa \tau_i) - \frac{\sigma^2}{\kappa^5 \tau_i} (1 - \exp(-\kappa \tau_i)) \right) \cdot \\ &\cdot (2\kappa \tau_i \exp(-\kappa \tau_i) + 3 \exp(-\kappa \tau_i) - 3) \\ &+ \frac{\sigma^2}{4\kappa^4 \tau_i} (1 - \exp(-\kappa \tau_i)) \cdot \\ &\cdot (2\tau_i \exp(-\kappa \tau_i) - 2\kappa \tau_i^2 \exp(-\kappa \tau_i) - 3\tau_i \exp(-\kappa \tau_i)) \\ &- \left(\frac{2\lambda \sigma}{\kappa^3} + \frac{3\sigma^2}{\kappa^4} \right) \left(1 + \frac{\exp(-\kappa \tau_i) - 1}{\kappa \tau_i} \right) \\ &+ \left(\frac{\lambda \sigma}{\kappa^2} + \frac{\sigma^2}{\kappa^3} \right) \left(\frac{1 - (\kappa \tau_i + 1) \exp(-\kappa \tau_i)}{\kappa^2 \tau_i} \right) \end{aligned}$$

$$\frac{\partial^2 \mathbf{d}}{\partial \kappa \partial \lambda} = \frac{\partial^2 \mathbf{d}}{\partial \lambda \partial \kappa} = \frac{\sigma}{\kappa^3 \tau_i} ((\kappa \tau_i + 2) \exp(-\kappa \tau_i) + \kappa \tau_i - 2)$$

$$\frac{\partial^2 \mathbf{W}_i}{\partial \psi_i \partial \psi_j} = \mathbf{0} \quad \forall \quad j = 1, \dots, 4.$$

3.7 Segundas derivadas del vector de estado y covarianzas, adelantados y actualizados

Algunas de las derivadas de la sección 3.6 valen cero. Si sustituimos estos casos en (2.3.19), (2.3.20), (2.3.21) y (2.3.22) para $\psi_1 = \mu$, $\psi_2 = \sigma$, $\psi_3 = \kappa$ y $\psi_4 = \lambda$ obtenemos:

$$\frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \mu^2} = \mathbf{A} \frac{\partial^2 \hat{\mathbf{x}}_{t-1|t-1}}{\partial \mu^2}$$

$$\frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \mu \partial \sigma} = \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \sigma \partial \mu} = \mathbf{A} \frac{\partial^2 \hat{\mathbf{x}}_{t-1|t-1}}{\partial \mu \partial \sigma}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \mu \partial \kappa} &= \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \kappa \partial \mu} = \\ &= \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{x}_{t-1|t-1}}{\partial \mu} + \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \mu \partial \kappa} + \frac{\partial^2 \mathbf{b}}{\partial \mu \partial \kappa} \end{aligned}$$

$$\frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \mu \partial \lambda} = \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \lambda \partial \mu} = \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \mu \partial \lambda}$$

$$\frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \sigma^2} = \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \sigma^2}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \sigma \partial \kappa} &= \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \kappa \partial \sigma} = \\ &= \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{x}_{t-1|t-1}}{\partial \sigma} + \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \sigma \partial \kappa} + \frac{\partial^2 \mathbf{b}}{\partial \sigma \partial \kappa} \end{aligned}$$

$$\frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \sigma \partial \lambda} = \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \lambda \partial \sigma} = \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \sigma \partial \lambda} + \frac{\partial^2 \mathbf{b}}{\partial \sigma \partial \lambda}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \kappa^2} &= \frac{\partial^2 \mathbf{A}}{\partial \kappa^2} \mathbf{x}_{t-1|t-1} + 2 \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{x}_{t-1|t-1}}{\partial \kappa} \\ &+ \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \kappa^2} + \frac{\partial^2 \mathbf{b}}{\partial \kappa^2} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \kappa \partial \lambda} &= \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \lambda \partial \kappa} = \\ &= \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{x}_{t-1|t-1}}{\partial \lambda} + \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \kappa \partial \lambda} + \frac{\partial^2 \mathbf{b}}{\partial \kappa \partial \lambda} \end{aligned}$$

$$\frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \lambda^2} = \mathbf{A} \frac{\partial^2 \mathbf{x}_{t-1|t-1}}{\partial \lambda^2}$$

$$\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu^2} = \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \mu^2} \mathbf{A}'$$

$$\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \sigma} = \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \mu} = \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \mu \partial \sigma} \mathbf{A}'$$

$$\begin{aligned} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \kappa} &= \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa \partial \mu} = \\ &= \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \mu} \mathbf{A}' + \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \mu \partial \kappa} \mathbf{A}' + \mathbf{A} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \mu} \frac{\partial \mathbf{A}'}{\partial \kappa} \end{aligned}$$

$$\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \lambda} = \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda \partial \mu} = \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \mu \partial \lambda} \mathbf{A}'$$

$$\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma^2} = \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \sigma^2} \mathbf{A}' + \frac{\partial^2 \mathbf{V}}{\partial \sigma^2}$$

$$\begin{aligned} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \kappa} &= \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa \partial \sigma} = \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \sigma} \mathbf{A}' + \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \sigma \partial \kappa} \mathbf{A}' \\ &+ \mathbf{A} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \sigma} \frac{\partial \mathbf{A}'}{\partial \kappa} + \frac{\partial^2 \mathbf{V}}{\partial \sigma \partial \kappa} \end{aligned}$$

$$\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \lambda} = \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda \partial \sigma} = \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \sigma \partial \lambda} \mathbf{A}'$$

$$\begin{aligned} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa^2} &= \\ &= \frac{\partial^2 \mathbf{A}}{\partial \kappa^2} \mathbf{P}_{t-1|t-1} \mathbf{A}' + \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \kappa^2} \mathbf{A}' + \mathbf{A} \mathbf{P}_{t-1|t-1} \frac{\partial^2 \mathbf{A}'}{\partial \kappa^2} + \frac{\partial^2 \mathbf{V}}{\partial \kappa^2} \\ &+ 2 \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \kappa} \mathbf{A}' + 2 \frac{\partial \mathbf{A}}{\partial \kappa} \mathbf{P}_{t-1|t-1} \frac{\partial \mathbf{A}'}{\partial \kappa} + 2 \mathbf{A} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \kappa} \frac{\partial \mathbf{A}'}{\partial \kappa} \\ \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa \partial \lambda} &= \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda \partial \kappa} = \\ &= \frac{\partial \mathbf{A}}{\partial \kappa} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \lambda} \mathbf{A}' + \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \kappa \partial \lambda} \mathbf{A}' + \mathbf{A} \frac{\partial \mathbf{P}_{t-1|t-1}}{\partial \lambda} \frac{\partial \mathbf{A}'}{\partial \kappa} \end{aligned}$$

$$\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda^2} = \mathbf{A} \frac{\partial^2 \mathbf{P}_{t-1|t-1}}{\partial \lambda^2} \mathbf{A}'$$

$$\begin{aligned} \frac{\partial^2 \hat{\mathbf{x}}_{t|t}}{\partial \mu^2} &= \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \mu^2} + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\frac{\partial^2 \mathbf{v}_t}{\partial \mu^2} - \frac{\partial^2 \mathbf{F}_t}{\partial \mu^2} \mathbf{F}_t^{-1} \mathbf{v}_t \right) \\ &+ \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu^2} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{v}_t \end{aligned}$$

$$\begin{aligned}
& + \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \lambda} \right) \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{v}_t}{\partial \kappa} - \frac{\partial \mathbf{F}_t}{\partial \kappa} \mathbf{F}_t^{-1} \mathbf{v}_t \right) \\
& + \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \mathbf{C}' + \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'}{\partial \kappa} - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \kappa} \right) \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{v}_t}{\partial \lambda} - \frac{\partial \mathbf{F}_t}{\partial \lambda} \mathbf{F}_t^{-1} \mathbf{v}_t \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \hat{\mathbf{x}}_{t|t}}{\partial \lambda^2} &= \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \lambda^2} + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\frac{\partial^2 \mathbf{v}_t}{\partial \lambda^2} - \frac{\partial^2 \mathbf{F}_t}{\partial \lambda^2} \mathbf{F}_t^{-1} \mathbf{v}_t \right) \\
& + \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda^2} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{v}_t \\
& + 2 \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \lambda} \right) \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{v}_t}{\partial \lambda} - \frac{\partial \mathbf{F}_t}{\partial \lambda} \mathbf{F}_t^{-1} \mathbf{v}_t \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{P}_{t|t}}{\partial \mu^2} &= \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu^2} - \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu^2} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
& + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \mu^2} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu^2} \\
& - 2 \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \mu} \right) \mathbf{F}_t^{-1} \left(-\frac{\partial \mathbf{F}_t}{\partial \mu} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{P}_{t|t}}{\partial \mu \partial \sigma} &= \frac{\partial^2 \mathbf{P}_{t|t}}{\partial \sigma \partial \mu} = \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \sigma} - \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \sigma} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
& + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \mu \partial \sigma} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \sigma} \\
& - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \mu} \right) \mathbf{F}_t^{-1} \left(-\frac{\partial \mathbf{F}_t}{\partial \sigma} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \right) \\
& - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \sigma} \right) \mathbf{F}_t^{-1} \left(-\frac{\partial \mathbf{F}_t}{\partial \mu} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{P}_{t|t}}{\partial \mu \partial \kappa} &= \frac{\partial^2 \mathbf{P}_{t|t}}{\partial \kappa \partial \mu} = \\
& = \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \kappa} - \left(\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \kappa} \mathbf{C}' + \frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \frac{\partial \mathbf{C}'}{\partial \kappa} \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
& + \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \mu \partial \kappa} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1}
\end{aligned}$$

$$\begin{aligned}
& -\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\left(\frac{\partial\mathbf{C}}{\partial\kappa}\frac{\partial\mathbf{P}_{t|t-1}}{\partial\mu}+\mathbf{C}\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\mu\partial\kappa}\right) \\
& -\left(\frac{\partial\mathbf{P}_{t|t-1}}{\partial\mu}\mathbf{C}'-\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial\mathbf{F}_t}{\partial\mu}\right)\mathbf{F}_t^{-1}\cdot \\
& \cdot\left(-\frac{\partial\mathbf{F}_t}{\partial\kappa}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1}+\frac{\partial\mathbf{C}}{\partial\kappa}\mathbf{P}_{t|t-1}+\mathbf{C}\frac{\partial\mathbf{P}_{t|t-1}}{\partial\kappa}\right) \\
& -\left(\frac{\partial\mathbf{P}_{t|t-1}}{\partial\kappa}\mathbf{C}'+\mathbf{P}_{t|t-1}\frac{\partial\mathbf{C}'}{\partial\kappa}-\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial\mathbf{F}_t}{\partial\kappa}\right)\mathbf{F}_t^{-1}\cdot \\
& \cdot\left(-\frac{\partial\mathbf{F}_t}{\partial\mu}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1}+\mathbf{C}\frac{\partial\mathbf{P}_{t|t-1}}{\partial\mu}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\mathbf{P}_{t|t}}{\partial\mu\partial\lambda}&=\frac{\partial^2\mathbf{P}_{t|t}}{\partial\lambda\partial\mu}=\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\mu\partial\lambda}-\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\mu\partial\lambda}\mathbf{C}'\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1} \\
& +\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial^2\mathbf{F}_t}{\partial\mu\partial\lambda}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1}-\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\mathbf{C}\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\mu\partial\lambda} \\
& -\left(\frac{\partial\mathbf{P}_{t|t-1}}{\partial\mu}\mathbf{C}'-\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial\mathbf{F}_t}{\partial\mu}\right)\mathbf{F}_t^{-1}\left(-\frac{\partial\mathbf{F}_t}{\partial\lambda}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1}+\mathbf{C}\frac{\partial\mathbf{P}_{t|t-1}}{\partial\lambda}\right) \\
& -\left(\frac{\partial\mathbf{P}_{t|t-1}}{\partial\lambda}\mathbf{C}'-\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial\mathbf{F}_t}{\partial\lambda}\right)\mathbf{F}_t^{-1}\left(-\frac{\partial\mathbf{F}_t}{\partial\mu}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1}+\mathbf{C}\frac{\partial\mathbf{P}_{t|t-1}}{\partial\mu}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\mathbf{P}_{t|t}}{\partial\sigma^2}&=\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\sigma^2}-\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\sigma^2}\mathbf{C}'\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1} \\
& +\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial^2\mathbf{F}_t}{\partial\sigma^2}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1}-\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\mathbf{C}\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\sigma^2} \\
& -2\left(\frac{\partial\mathbf{P}_{t|t-1}}{\partial\sigma}\mathbf{C}'-\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial\mathbf{F}_t}{\partial\sigma}\right)\mathbf{F}_t^{-1}\left(-\frac{\partial\mathbf{F}_t}{\partial\sigma}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1}+\mathbf{C}\frac{\partial\mathbf{P}_{t|t-1}}{\partial\sigma}\right)
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial^2\mathbf{P}_{t|t}}{\partial\sigma\partial\kappa}=\frac{\partial^2\mathbf{P}_{t|t}}{\partial\kappa\partial\sigma}= \\
& =\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\sigma\partial\kappa}-\left(\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\sigma\partial\kappa}\mathbf{C}'+\frac{\partial\mathbf{P}_{t|t-1}}{\partial\sigma}\frac{\partial\mathbf{C}'}{\partial\kappa}\right)\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1} \\
& +\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\frac{\partial^2\mathbf{F}_t}{\partial\sigma\partial\kappa}\mathbf{F}_t^{-1}\mathbf{C}\mathbf{P}_{t|t-1} \\
& -\mathbf{P}_{t|t-1}\mathbf{C}'\mathbf{F}_t^{-1}\left(\frac{\partial\mathbf{C}}{\partial\kappa}\frac{\partial\mathbf{P}_{t|t-1}}{\partial\sigma}+\mathbf{C}\frac{\partial^2\mathbf{P}_{t|t-1}}{\partial\sigma\partial\kappa}\right)
\end{aligned}$$

$$\begin{aligned}
& - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \sigma} \right) \mathbf{F}_t^{-1} \cdot \\
& \cdot \left(- \frac{\partial \mathbf{F}_t}{\partial \kappa} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \frac{\partial \mathbf{C}}{\partial \kappa} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \right) \\
& - \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \mathbf{C}' + \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'}{\partial \kappa} - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \kappa} \right) \mathbf{F}_t^{-1} \cdot \\
& \cdot \left(- \frac{\partial \mathbf{F}_t}{\partial \sigma} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{P}_{t|t}}{\partial \sigma \partial \lambda} &= \frac{\partial^2 \mathbf{P}_{t|t}}{\partial \lambda \partial \sigma} = \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \lambda} - \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \lambda} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
&+ \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \sigma \partial \lambda} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \lambda} \\
&- \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \lambda} \right) \mathbf{F}_t^{-1} \left(- \frac{\partial \mathbf{F}_t}{\partial \sigma} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \right) \\
&- \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \sigma} \right) \mathbf{F}_t^{-1} \left(- \frac{\partial \mathbf{F}_t}{\partial \lambda} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{P}_{t|t}}{\partial \kappa^2} &= \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa^2} \\
&- \left(\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa^2} \mathbf{C}' + 2 \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \frac{\partial \mathbf{C}'}{\partial \kappa} + \mathbf{P}_{t|t-1} \frac{\partial^2 \mathbf{C}'}{\partial \kappa^2} \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
&+ \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \kappa^2} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
&- \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \cdot \left(\frac{\partial^2 \mathbf{C}}{\partial \kappa^2} \mathbf{P}_{t|t-1} + 2 \frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} + \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa^2} \right) \\
&- 2 \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \mathbf{C}' + \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'}{\partial \kappa} - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \kappa} \right) \cdot \\
&\cdot \mathbf{F}_t^{-1} \left(- \frac{\partial \mathbf{F}_t}{\partial \kappa} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \frac{\partial \mathbf{C}}{\partial \kappa} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \right)
\end{aligned}$$

$$\frac{\partial^2 \mathbf{P}_{t|t}}{\partial \kappa \partial \lambda} = \frac{\partial^2 \mathbf{P}_{t|t}}{\partial \lambda \partial \kappa} =$$

$$\begin{aligned}
&= \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa \partial \lambda} - \left(\frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa \partial \lambda} \mathbf{C}' + \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \frac{\partial \mathbf{C}'}{\partial \kappa} \right) \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
&+ \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \kappa \partial \lambda} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
&- \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \left(\frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} + \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa \partial \lambda} \right) \\
&- \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \lambda} \right) \mathbf{F}_t^{-1} \cdot \\
&\cdot \left(-\frac{\partial \mathbf{F}_t}{\partial \kappa} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \frac{\partial \mathbf{C}}{\partial \kappa} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \right) \\
&- \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \mathbf{C}' + \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'}{\partial \kappa} - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \kappa} \right) \mathbf{F}_t^{-1} \cdot \\
&\cdot \left(-\frac{\partial \mathbf{F}_t}{\partial \lambda} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{P}_{t|t}}{\partial \lambda^2} &= \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda^2} - \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda^2} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
&+ \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial^2 \mathbf{F}_t}{\partial \lambda^2} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} \\
&- \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda^2} - 2 \left(\frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \mathbf{C}' - \mathbf{P}_{t|t-1} \mathbf{C}' \mathbf{F}_t^{-1} \frac{\partial \mathbf{F}_t}{\partial \lambda} \right) \cdot \\
&\cdot \mathbf{F}_t^{-1} \left(-\frac{\partial \mathbf{F}_t}{\partial \lambda} \mathbf{F}_t^{-1} \mathbf{C} \mathbf{P}_{t|t-1} + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \right)
\end{aligned}$$

Las condiciones iniciales para las recursiones anteriores son:

$$\frac{\partial^2 \hat{\mathbf{x}}_{0|0}}{\partial \psi_i \partial \psi_j} = 0 \text{ si } \psi_i \text{ o } \psi_j \text{ son } \mu.$$

$$\frac{\partial^2 \hat{\mathbf{x}}_{0|0}}{\partial \sigma^2} = \frac{\partial^2 \hat{\mathbf{x}}_{0|0}}{\partial \lambda^2} = 0$$

$$\frac{\partial^2 \hat{\mathbf{x}}_{0|0}}{\partial \sigma \partial \kappa} = \frac{\partial^2 \hat{\mathbf{x}}_{0|0}}{\partial \kappa \partial \sigma} = \frac{\lambda}{\kappa^2}$$

$$\frac{\partial^2 \mathbf{x}_{0|0}}{\partial \sigma \partial \lambda} = \frac{\partial^2 \mathbf{x}_{0|0}}{\partial \lambda \partial \sigma} = -\frac{1}{\kappa}$$

$$\frac{\partial^2 \mathbf{x}_{0|0}}{\partial \kappa^2} = -\frac{2\lambda\sigma}{\kappa^3}$$

$$\frac{\partial^2 \mathbf{x}_{0|0}}{\partial \kappa \partial \lambda} = \frac{\partial^2 \mathbf{x}_{0|0}}{\partial \lambda \partial \kappa} = \frac{\sigma}{\kappa^2}$$

$$\frac{\partial^2 \mathbf{P}_{0|0}}{\partial \psi_i \partial \psi_j} = 0 \text{ si } \psi_i \text{ o } \psi_j \text{ son } \mu \text{ o } \lambda.$$

$$\frac{\partial^2 \mathbf{P}_{0|0}}{\partial \sigma^2} = \frac{1}{\kappa}$$

$$\frac{\partial^2 \mathbf{P}_{0|0}}{\partial \sigma \partial \kappa} = \frac{\partial^2 \mathbf{P}_{0|0}}{\partial \kappa \partial \sigma} = -\frac{\sigma}{\kappa^2}$$

$$\frac{\partial^2 \mathbf{P}_{0|0}}{\partial \kappa^2} = \frac{\sigma^2}{\kappa^3}$$

3.8 Segundas derivadas del error de pronóstico y su matriz de covarianzas

Algunas de las derivadas de la sección 3.6 valen cero. Si sustituimos estos casos en (2.3.17) y (2.3.18) para $\psi_1 = \mu$, $\psi_2 = \sigma$, $\psi_3 = \kappa$ y $\psi_4 = \lambda$ obtenemos:

$$\frac{\partial^2 \mathbf{v}_t}{\partial \mu^2} = -\mathbf{C} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \mu^2}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \mu \partial \sigma} = \frac{\partial^2 \mathbf{v}_t}{\partial \sigma \partial \mu} = -\mathbf{C} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \mu \partial \sigma}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \mu \partial \kappa} = \frac{\partial^2 \mathbf{v}_t}{\partial \kappa \partial \mu} = -\frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \mathbf{x}_{t|t-1}}{\partial \mu} - \mathbf{C} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \mu \partial \kappa} - \frac{\partial^2 \mathbf{d}}{\partial \mu \partial \kappa}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \mu \partial \lambda} = \frac{\partial^2 \mathbf{v}_t}{\partial \lambda \partial \mu} = -\mathbf{C} \frac{\partial^2 \mathbf{x}_{t|t-1}}{\partial \mu \partial \lambda}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \sigma^2} = -\mathbf{C} \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \sigma^2} - \frac{\partial^2 \mathbf{d}}{\partial \sigma^2}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \sigma \partial \kappa} = \frac{\partial^2 \mathbf{v}_t}{\partial \kappa \partial \sigma} = -\frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \sigma} - \mathbf{C} \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \sigma \partial \kappa} - \frac{\partial^2 \mathbf{d}}{\partial \sigma \partial \kappa}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \sigma \partial \lambda} = \frac{\partial^2 \mathbf{v}_t}{\partial \lambda \partial \sigma} = -\mathbf{C} \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \sigma \partial \lambda} - \frac{\partial^2 \mathbf{d}}{\partial \sigma \partial \lambda}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \kappa^2} = -\frac{\partial^2 \mathbf{C}}{\partial \kappa^2} \hat{\mathbf{x}}_{t|t-1} - 2 \frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \kappa} - \mathbf{C} \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \kappa^2} - \frac{\partial^2 \mathbf{d}}{\partial \kappa^2}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \kappa \partial \lambda} = \frac{\partial^2 \mathbf{v}_t}{\partial \lambda \partial \kappa} = -\mathbf{C} \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \kappa \partial \lambda} - \frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \hat{\mathbf{x}}_{t|t-1}}{\partial \lambda} - \frac{\partial^2 \mathbf{d}}{\partial \kappa \partial \lambda}$$

$$\frac{\partial^2 \mathbf{v}_t}{\partial \lambda^2} = -\mathbf{C} \frac{\partial^2 \hat{\mathbf{x}}_{t|t-1}}{\partial \lambda^2}$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \mu^2} = \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu^2} \mathbf{C}'$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \mu \partial \sigma} = \frac{\partial^2 \mathbf{F}_t}{\partial \sigma \partial \mu} = \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \sigma} \mathbf{C}'$$

$$\begin{aligned} & \frac{\partial^2 \mathbf{F}_t}{\partial \mu \partial \kappa} = \frac{\partial^2 \mathbf{F}_t}{\partial \kappa \partial \mu} = \\ & = \frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \mathbf{C}' + \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \kappa} \mathbf{C}' + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \mu} \frac{\partial \mathbf{C}'}{\partial \kappa} \end{aligned}$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \mu \partial \lambda} = \frac{\partial^2 \mathbf{F}_t}{\partial \lambda \partial \mu} = \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \mu \partial \lambda} \mathbf{C}'$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \sigma^2} = \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma^2} \mathbf{C}'$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \sigma \partial \kappa} = \frac{\partial^2 \mathbf{F}_t}{\partial \kappa \partial \sigma} =$$

$$= \frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \mathbf{C}' + \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \kappa} \mathbf{C}' + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \sigma} \frac{\partial \mathbf{C}'}{\partial \kappa}$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \sigma \partial \lambda} = \frac{\partial^2 \mathbf{F}_t}{\partial \lambda \partial \sigma} = \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \sigma \partial \lambda} \mathbf{C}'$$

$$\begin{aligned} \frac{\partial^2 \mathbf{F}_t}{\partial \kappa^2} &= \frac{\partial^2 \mathbf{C}}{\partial \kappa^2} \mathbf{P}_{t|t-1} \mathbf{C}' + \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa^2} \mathbf{C}' + \mathbf{C} \mathbf{P}_{t|t-1} \frac{\partial^2 \mathbf{C}'}{\partial \kappa^2} \\ &+ 2 \frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \mathbf{C}' + 2 \frac{\partial \mathbf{C}}{\partial \kappa} \mathbf{P}_{t|t-1} \frac{\partial \mathbf{C}'}{\partial \kappa} + 2 \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \kappa} \frac{\partial \mathbf{C}'}{\partial \kappa} \end{aligned}$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \kappa \partial \lambda} = \frac{\partial^2 \mathbf{F}_t}{\partial \lambda \partial \kappa} =$$

$$= \frac{\partial \mathbf{C}}{\partial \kappa} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \mathbf{C}' + \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \kappa \partial \lambda} \mathbf{C}' + \mathbf{C} \frac{\partial \mathbf{P}_{t|t-1}}{\partial \lambda} \frac{\partial \mathbf{C}'}{\partial \kappa}$$

$$\frac{\partial^2 \mathbf{F}_t}{\partial \lambda^2} = \mathbf{C} \frac{\partial^2 \mathbf{P}_{t|t-1}}{\partial \lambda^2} \mathbf{C}'$$

Capítulo 4: Resultados

Tras aplicar el filtro de Kalman en el espacio de estados que presentamos en el capítulo 3 y resolver la función de verosimilitud correspondiente utilizando algún método recursivo de optimización para funciones no lineales, de varias variables, en problemas no restringidos,¹ el valor de los hiperparámetros al que se llega es:

$$\begin{aligned}\mu &= 7.84435911 \\ \sigma &= 1.68771177 \\ \kappa &= 0.56696227 \\ \lambda &= -0.99535321\end{aligned}\tag{4.1.1}$$

Los cálculos requeridos los hicimos con Excel 2003. La función de log-verosimilitud, evaluada en estos resultados, es de -1070.14727.

Ahora bien, de (3.1.1) tenemos:

$$\mathbf{y}_t = \mathbf{C} \mathbf{x}_{t|t} + \mathbf{d}\tag{4.1.2}$$

La expresión anterior nos servirá para graficar la ETTI. Recordemos que el vector de estado representa la tasa de interés de corto plazo. Antes de graficar la ETTI, sustituyamos (4.1.1) en las matrices \mathbf{C} y \mathbf{d} dadas por (3.1.4) y (3.1.5), y a su vez esos resultados en (4.1.2). También apliquemos en (4.1.2) los valores $\mathbf{x}_{t|t}$ que el filtro de Kalman arrojó como vector de estado actualizado en t , de manera que obtengamos el vector de rendimientos estimados en t para las cuatro diferentes temporalidades al vencimiento utilizadas:

$$\mathbf{y}_t = \begin{pmatrix} \hat{y}_{28 \text{ días}, t} \\ \hat{y}_{91 \text{ días}, t} \\ \hat{y}_{182 \text{ días}, t} \\ \hat{y}_{364 \text{ días}, t} \end{pmatrix}$$

Los gráficos 4.1 al 4.4 muestran por separado cada uno de los elementos del vector anterior. Las líneas de color rosa corresponden con los registros reales del mercado mexicano de valores y las azules con las estimaciones. Estos gráficos los hicimos con Excel 2003.

¹ Puede consultar el apéndice E.

Gráfico 4.1: Tasas cero de CETES de 28 días

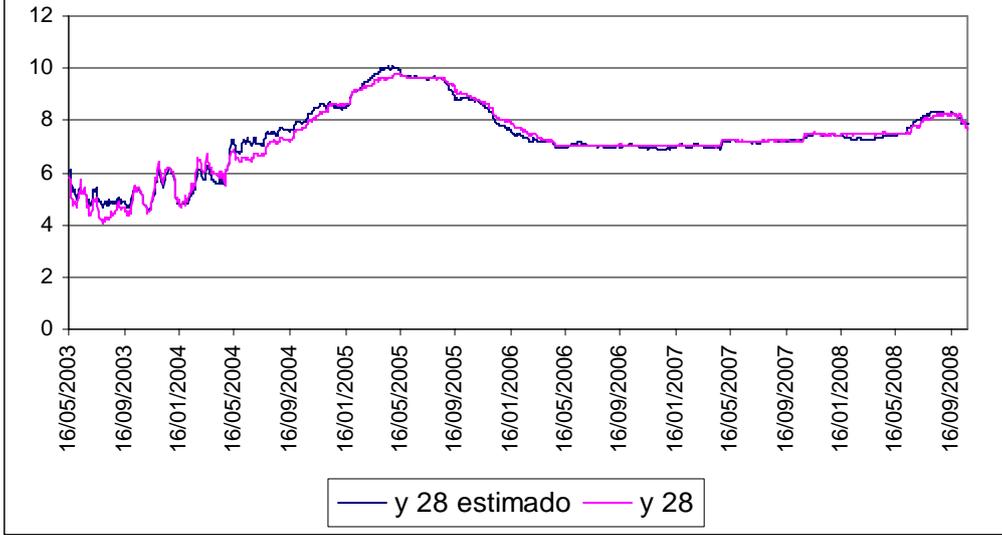


Gráfico 4.2: Tasas cero de CETES de 91 días

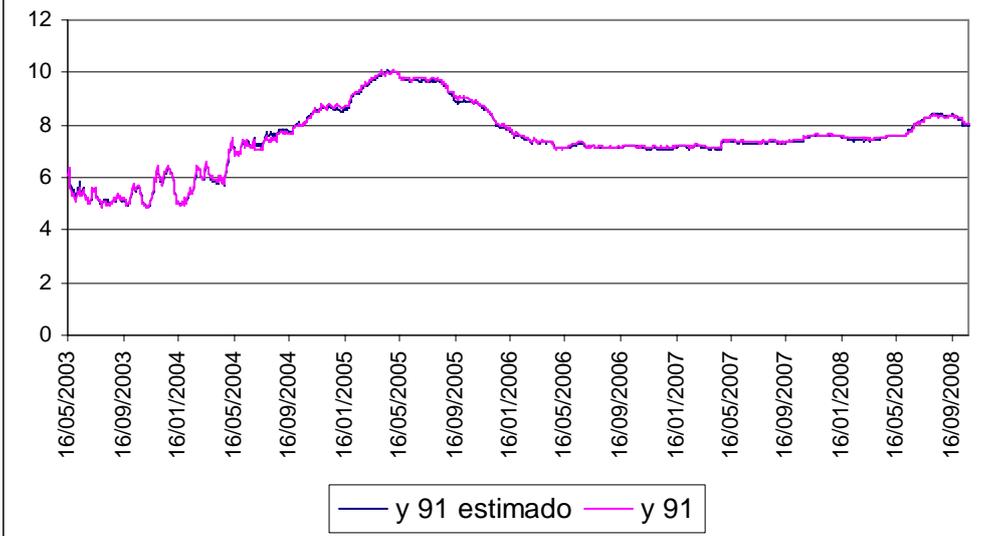


Gráfico 4.3: Tasas cero de CETES de 182 días

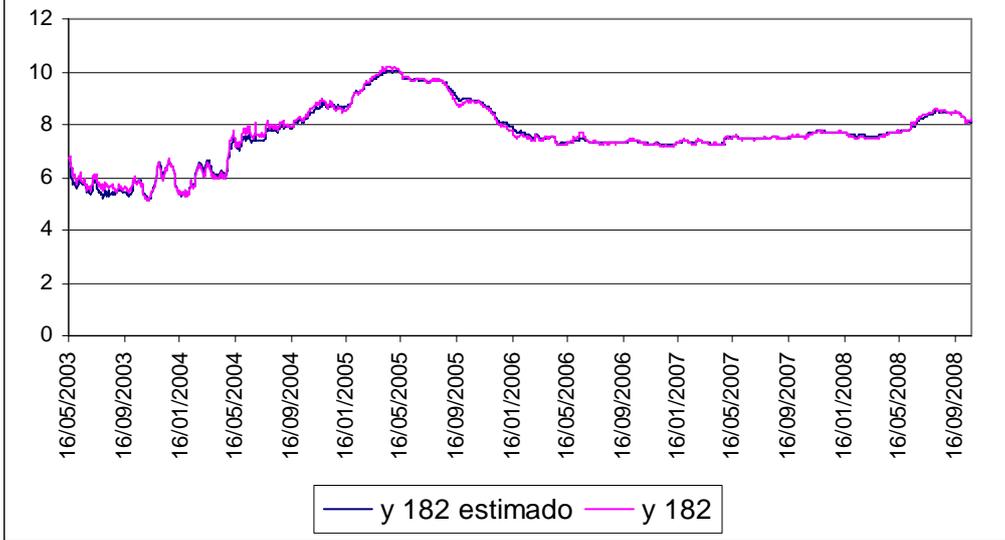
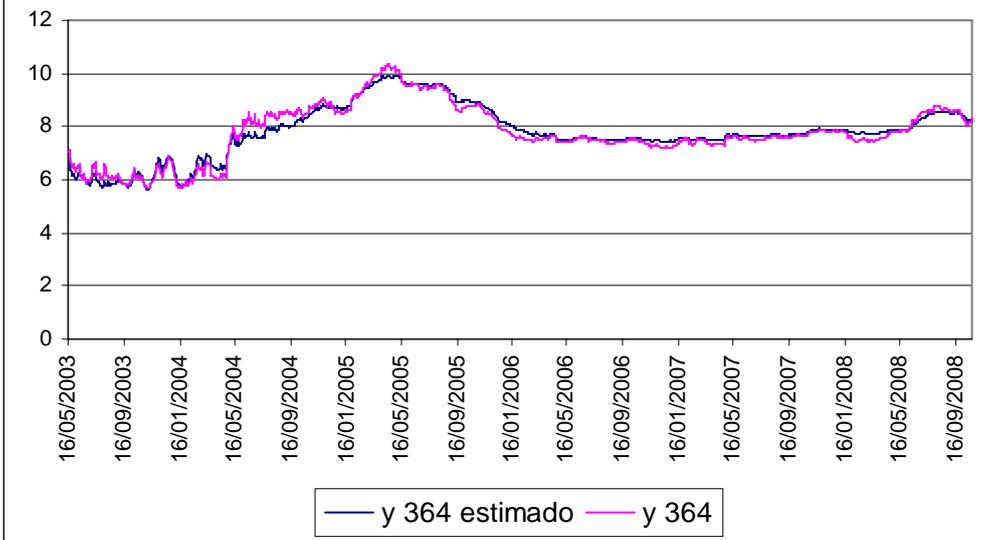


Gráfico 4.4: Tasas cero de CETES de 364 días



Recordemos que en el capítulo 3 obtuvimos (3.1.4) y (3.1.5) partiendo de (3.1.1) y (3.1.3) con $n = 4$ temporalidades al vencimiento, $\tau_i, i = 1, 2, \dots, n$. Para graficar la ETTI tendremos que hacer lo mismo pero usando $n = 364$ valores para τ_i , estos son:

$\frac{1}{364}, \frac{2}{364}, \dots, \frac{364}{364}$. Primero hay que sustituir (4.1.1) en los valores $F(\tau_i)$ y $G(\tau_i)$ dados por (1.6.5) y (1.6.7), y luego esos resultados en (3.1.3). Además, en la r_t de (3.1.3) sustituiremos los valores $\hat{\mathbf{x}}_{t|t}$ que el filtro de Kalman arrojó como vector de estado actualizado en t . Finalmente, usando Matlab 7.0 obtenemos la estructura de temporalidad de tipos de interés de nuestro mercado, del 16 de mayo del 2003 al 21 de octubre de 2008, en los siguientes cuatro gráficos. Los gráficos 4.5A, 4.5B, 4.5C y 4.5D son en realidad el mismo, visto desde cuatro ángulos distintos.

Gráfico 4.5A

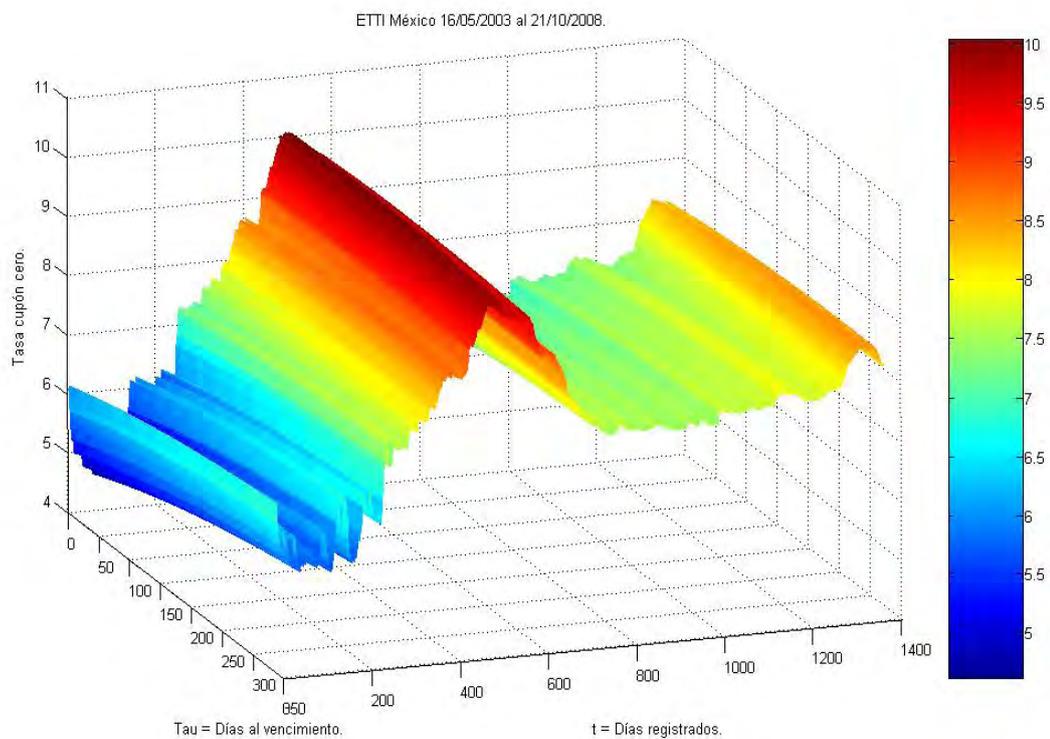


Gráfico 4.5B

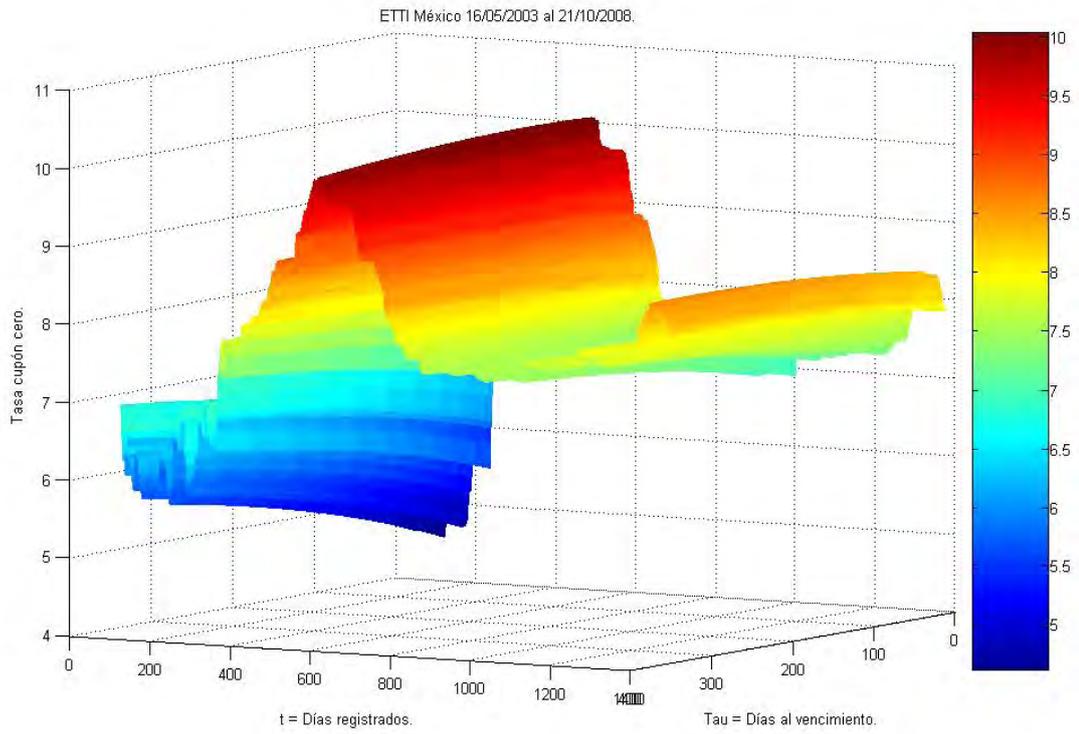


Gráfico 4.5C

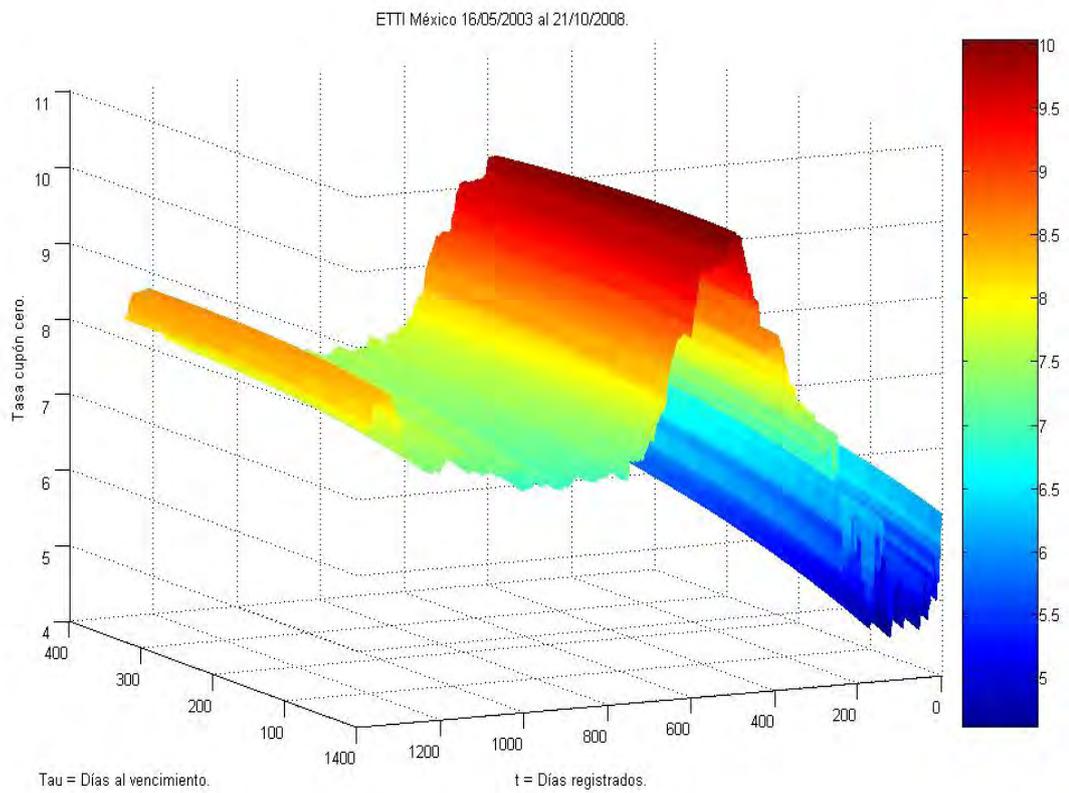
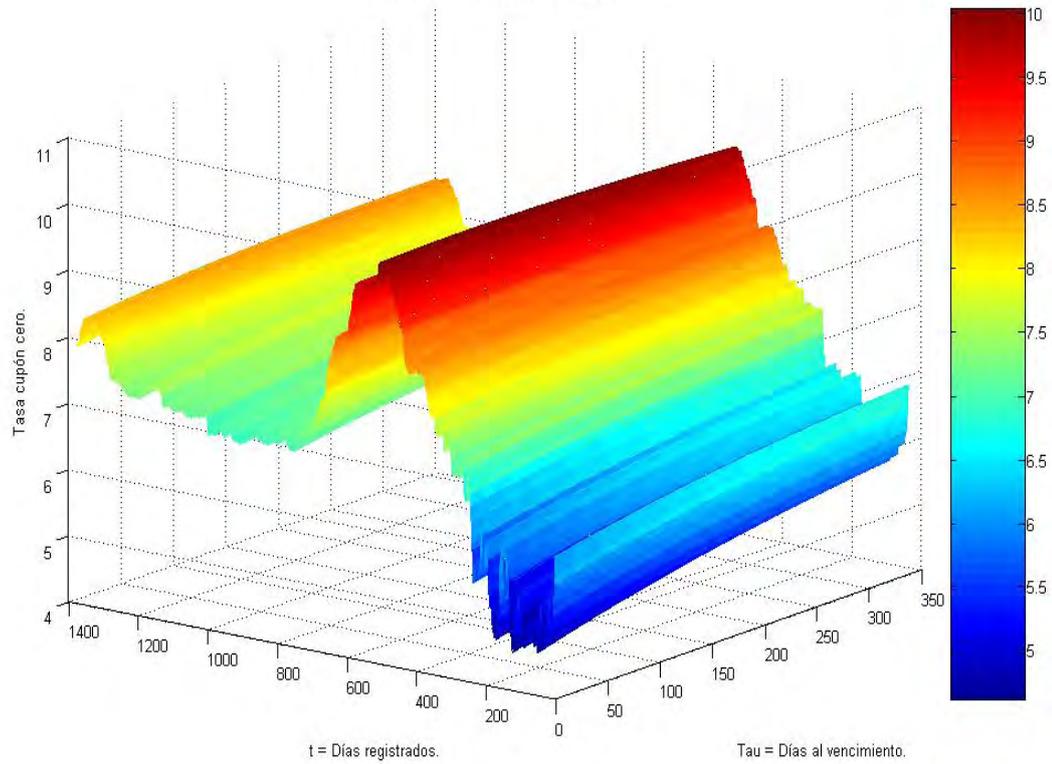
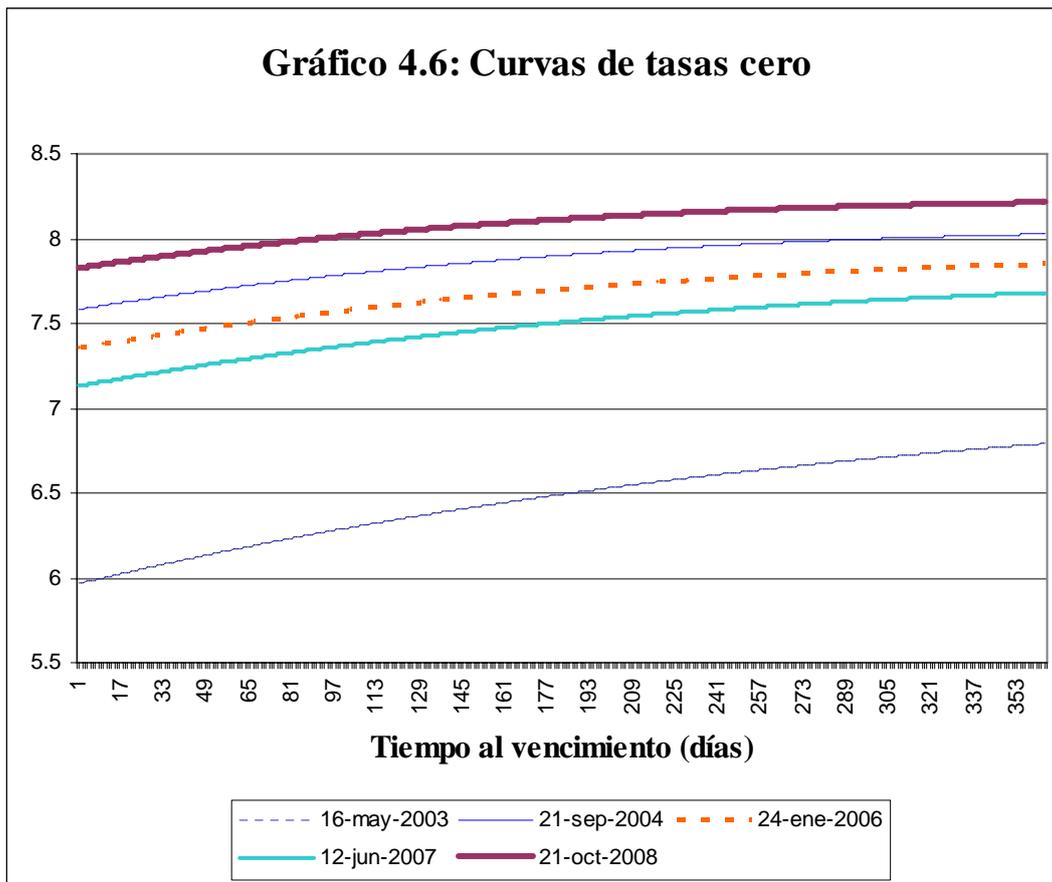


Gráfico 4.5D

ETTI México 16/05/2003 al 21/10/2008.



Por último, en el gráfico 4.6 presentamos algunas curvas de tasas cero. Es claro que tanto los gráficos 4.1 al 4.4 como el gráfico 4.6 están implícitos en el 4.5.



Podemos notar que en los gráficos 4.5A a 4.5D y en el 4.6, los plazos al vencimiento se expresaron en días y no en años. Esto lo hicimos por factibilidad para presentar nuestra información, pues los plazos que nos conciernen no superan un año. Para los cálculos sin embargo, el tiempo al vencimiento se midió en años, tal como se explicó al inicio del capítulo 3. Los días registrados por su parte, en los gráficos 4.1 a 4.4 se presentaron como fechas y en los gráficos 4.5A a 4.5D se señalaron según el lugar que ocupan; ambas formas son frecuentes.

Conclusiones y futuras investigaciones

Hemos construido una representación de espacio de estados para el modelo de estructura de temporalidad de tipos de interés de Vasicek (1977), cuyo único factor es la tasa de interés de corto plazo. Como vimos, las matrices del sistema del espacio de estados están en función de los llamados hiperparámetros, que en nuestro caso vienen siendo los parámetros del factor único del modelo, que como sabemos es la tasa de interés instantánea.

El espacio de estados en cuestión es un sistema de dos ecuaciones: 1) El vector de estado y 2) el vector de observaciones. El vector de estado representa los valores de las tasas de interés de corto plazo de Vasicek. Sus valores son desconocidos. El vector de observaciones está en función del vector de estado. Las salidas del vector de observaciones son las tasas de rendimiento de los bonos cupón cero o tasas cupón cero y son conocidas dado que fueron tomadas directamente de los registros del mercado mexicano de valores.

Presentamos las ecuaciones recursivas del filtro de Kalman, las cuales sirvieron para encontrar, a partir de las salidas conocidas del vector de observaciones, en cada punto de la serie de tiempo, los valores previamente desconocidos del vector de estado adelantado y su matriz de covarianza, los del vector de estado actualizado y su matriz de covarianza y los del error de pronóstico o vector de innovación y su matriz de covarianzas. Tales resultados sirvieron para construir una función de máxima verosimilitud en forma de descomposición del error de pronóstico del filtro.

Tras aplicar cualquier método numérico recursivo de optimización no lineal sin restricciones que nos ayudase a resolver la función de verosimilitud, encontramos el valor de los hiperparámetros del espacio de estados presentado, que son los parámetros de la tasa corta de Vasicek. Como resultado pudimos estimar las tasas cupón cero, de México, del 16/05/2003 al 21/10/2008 para cualquier tiempo al vencimiento entre 28 y 364 días, como una función de la tasa de interés instantánea estimada, cumpliendo así el objetivo general de esta tesis.

Algunas ideas para investigaciones futuras relacionadas con esta tesis son:

- Analizar las **propiedades** de controlabilidad, observabilidad, estabilidad y detectabilidad del espacio de estados propuesto en esta tesis.
- Proponer un espacio de estados para el modelo de ETTI de Vasicek, al cual puedan aplicarse las ecuaciones recursivas de Kalman en la modalidad de **pronóstico** o bien en la modalidad de **suavización**, en vez de la modalidad de filtrado, que fue el caso de esta tesis. La diferencia radica en la forma de procesar el conjunto de datos disponible.
- Construir un espacio de estados para **otro modelo de ETTI** como por ejemplo el CIR (1985) o Vasicek extendido (1990), que pueda resolverse aplicando el filtro de Kalman.
- Mostrar un espacio de estados para cualquier modelo de ETTI, como el Vasicek por ejemplo, utilizando un movimiento **browniano fraccional** para simular el ruido del proceso de tasa de interés instantánea y aplicar Kalman.
- Encontrar el valor de los hiperparámetros del espacio de estados propuesto en esta tesis o de cualquier otro mencionado anteriormente, a través de la estimación sobre el **dominio de frecuencias**. Lo que aplicamos en esta tesis fue estimación sobre el dominio del tiempo. Manejar el dominio de frecuencias requiere conocer temas como transformada Z, transformada de Laplace, de Fourier, de Fourier inversa o Wavelets. Dependiendo del problema puede ser más conveniente usar un dominio u otro.
- Proponer **nuevos modelos** estructurales a partir de los antedichos.

Recordemos que la ETTI es la referencia obligatoria para introducir el componente de mercado en un determinado análisis y tiene las importantes aplicaciones económico-financieras que hemos mencionado anteriormente.

Apéndices

Apéndice A: Procesos estocásticos estacionarios

Este apéndice define los procesos estocásticos estacionarios.¹

Un proceso estocástico es un mecanismo que genera observaciones y_t sobre un período de tiempo $t = 1, 2, \dots, T$. Un proceso estocástico puede generar un conjunto infinito de series de tiempo sobre dicho período de tiempo. Cada uno de los conjuntos individuales de observaciones se conoce como una realización. Para que un proceso estocástico sea *estacionario* o *débilmente estacionario* debe satisfacerse $\forall t$ las condiciones:

$$E(y_t) = \mu$$

$$E((y_t - \mu)^2) = \text{var}(y_t) = \gamma(0)$$

$$E((y_t - \mu)(y_{t-k} - \mu)) = \gamma(k) \quad k = 1, 2, 3, \dots$$

La esperanza es tomada sobre todas las posibles realizaciones del proceso.

La característica esencial de un proceso estacionario es que sus características no cambian a los largo del tiempo. La media permanece en un nivel constante $\mu \forall t$. Igualmente, la varianza $\gamma(0)$ y las autocovarianzas $\gamma(k)$ tampoco dependen del tiempo. Consecuentemente la media, varianza y autocovarianza pueden estimarse a partir de una realización particular llamada respectivamente *media muestral*, *varianza muestral* y *autocovarianza muestral*:

$$\bar{\mu} = \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

$$\hat{\gamma}(0) = c(0) = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2 \quad (\text{A.1})$$

$$\hat{\gamma}(k) = c(k) = \frac{1}{T} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}) \quad k = 1, 2, 3, \dots \quad (\text{A.2})$$

Las propiedades dinámicas de un proceso estacionario pueden resumirse graficando $\gamma(k)$ contra k . El conjunto entero de las $\gamma(k)$ es conocido como función de autocovarianza. Dado que $\gamma(k) = \gamma(-k)$ para series reales, es innecesario extender la gráfica sobre valores negativos de k . Las autocovarianzas pueden estandarizarse dividiendo por el proceso de varianza. Esto genera las *autocorrelaciones*:

¹ Harvey, *Forecasting, structural time series models and the Kalman filter*, pp. 49-50.

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \quad k = 1, 2, 3, \dots$$

Una gráfica de $\rho(k)$ contra k muestra la función de autocorrelación. Note que $\rho(0)$ siempre valdrá 1.

Las autocorrelaciones muestrales se definen:

$$r(k) = \frac{c(k)}{c(0)} \quad k = 0, 1, 2, \dots$$

Una gráfica de $r(k)$ contra k es conocida como *correlograma*. Las autocorrelaciones muestrales están sujetas a la variabilidad de la muestra y aunque el correlograma tenderá a reflejar las propiedades de la función de autocorrelación teórica, no las reproducirá exactamente. Note que si sabemos que la media del proceso es $\mu = 0$ no es necesario calcular la media muestral para sustituirla en $c(0)$ y $c(k)$ de (A.1) y (A.2).

Un proceso es *estrictamente estacionario* si la distribución de probabilidad conjunta de un conjunto de observaciones $r(t_1), r(t_2), \dots, r(t_r)$ es idéntica a la distribución de probabilidad conjunta de las observaciones $r(t_1 + k), r(t_2 + k), \dots, r(t_r + k) \forall k$. Si el proceso es estrictamente estacionario entonces es débilmente estacionario. Lo opuesto generalmente no es cierto. No obstante una serie débilmente estacionaria con distribución gaussiana sí es estacionaria en el sentido estricto.

Apéndice B: Otros conceptos de procesos estocásticos

En este apéndice encontraremos las definiciones de proceso de Markov, movimiento browniano, movimiento browniano estándar, martingala, la propiedad de isometría y el lema de Ito y el teorema de Girsanov.¹

Suposición: Todas las variables aleatorias y procesos estocásticos se definen en un espacio de probabilidad completo dado $(\Omega, \mathcal{A}, \mathbb{P})$.

Concepto B.1. Proceso de Markov: Conforme pasa el tiempo observamos la evolución del objeto descrito por el proceso estocástico. En cualquier t' los valores previos $\{x_t\}_{t \in [0, t']}$ donde $x_t \in S$ serán conocidos. Constituyen la historia del proceso hasta el momento t' . Los valores futuros son estocásticos.

Conforme pasa el tiempo podemos revisar la distribución de probabilidad que atribuimos al proceso $\forall t, u \in \mathfrak{R}^+$ en cualquier punto futuro del tiempo. Dado el

¹ Constantin Tudor, *Procesos Estocásticos*, 3ª ed. (México: Sociedad Matemática Mexicana, 2002), pp. 178 y 355 - 356.

tiempo presente t y algún punto futuro $u > t$. La distribución del valor de x_u está caracterizada por las probabilidades $P(x_u \in A)$ para subconjuntos A -medibles del espacio de estados S . Si $\forall t, u \in \mathfrak{R}^+$ con $t < u$ y $\forall A \subseteq S$ se tiene que

$$P[x_u \in A | \{x_s\}_{s \in [0, t]}] = P(x_u \in A | x_t)$$

entonces $\{x_t\}_{t \geq 0}$ se denomina un **proceso de Markov**. Esta condición implica que en este tipo de proceso únicamente su estado actual es relevante para predecir el estado futuro; la historia pasada del proceso y la forma en que el presente ha emergido del pasado son irrelevantes.

Definición B.2: Una filtración $\mathbf{F} = (\mathfrak{F}_t)_{0 \leq t \leq T}$ es una sucesión de subsigma álgebras de A tales que $\mathfrak{F}_s \subseteq \mathfrak{F}_t$ para $s \leq t$.

\mathbf{F} es usada para modelar un flujo de información. Conforme pasa el tiempo, un observador dispone de más y mejor información. En otros términos, conforme t crece, \mathfrak{F}_t como partición de Ω se vuelve mejor.

Definición B.3: Un movimiento browniano $(W_t)_{t \geq 0}$ es un proceso estocástico continuo valuado en los números reales, con las siguientes propiedades definitorias:

1. **Incrementos independientes:** $\forall s \leq t$, $W_t - W_s$ es independiente de la sub- σ álgebra generada por $\{W_u, u \leq s\}$, denotada por $\mathfrak{F}_s = \sigma(W_u, u \leq s)$.
2. **Incrementos estacionarios:** $\forall s \leq t$, $W_t - W_s$ y $W_{t-s} - W_0$ tienen la misma ley de probabilidad.
3. **Continuidad del trayecto:** $W_t(\omega)$ es una función continua de t .

A partir de que N. Wiener en 1931 estableció los fundamentos matemáticos del movimiento browniano como un proceso estocástico, a este proceso también se le conoce como *proceso de Wiener*.

Teorema B.4: Si $(W_t)_{t \geq 0}$ es un movimiento browniano, entonces $W_t - W_0$ es una variable aleatoria normal con media ct y varianza $\sigma^2 t$, donde c y σ son números reales constantes.

Un proceso de Wiener es estándar si:

- $W_0 = 0$
- $E(W_t) = 0$
- $E(W_t^2) = t$

Generalmente cuando se dice movimiento browniano se refiere uno al movimiento browniano estándar.

Definición B.5: La colección $\mathbf{F} = \{ \mathfrak{F}_t, t \in [0, T] \}$ con $\mathfrak{F}_t = \sigma \{ W_s, s \leq t, \mathfrak{N} \}$ donde \mathfrak{N} es el conjunto de todos los conjuntos P -insignificantes de ω , llamado la filtración natural de W o la filtración generada por W .

Definición B.6: Un proceso estocástico continuo valuado en los números reales es un movimiento \mathfrak{F}_t -browniano si satisface:

- $\forall t \geq 0, W_t$ es \mathfrak{F}_t -medible.
- $\forall s \leq t, W_t - W_s$ es independiente de la σ -álgebra \mathfrak{F}_s
- $\forall s \leq t, W_t - W_s$ y $W_{t-s} - W_0$ tienen la ley de probabilidad.

Un movimiento \mathfrak{F}_t -browniano es un movimiento browniano respecto su filtración natural.

Definición B.7: Un proceso estocástico $(M_t)_{t \geq 0}$ adaptado a $(\mathfrak{F}_t)_{t \geq 0}$ es una martingala si para cualquier t es integrable, es decir, $E(|M_t|) < +\infty$ y para cualquier $s \leq t$ $E(M_t | \mathfrak{F}_s) = M_s$ casi seguramente.

La definición C.6 implica que $E(M_t) = E(M_0)$ para cualquier t .

Lema B.8. Lema de Ito:¹

Si una variable S sigue el proceso de Wiener

$$dS = \mu S dt + \sigma S dW_t$$

y $F = F(S, t)$ es suave entonces

$$dF = \sigma S \frac{\partial F}{\partial S} dW_t + \left(\frac{\partial F}{\partial t} + \mu S \frac{\partial F}{\partial S} + \frac{1}{2} \sigma^2 \frac{\partial^2 F}{\partial S^2} \right) dt$$

Concepto B.9. Propiedad de isometría: Supongamos un proceso de Wiener W_t . Como $W_{t_2} - W_{t_1}, \dots, W_{t_{n+1}} - W_{t_n}$ son independientes y con distribución $N(0, t_2 - t_1), \dots, N(0, t_{n+1} - t_n)$ respectivamente, resulta que $\int_T f(t) dW_t$ tiene distribución normal con media cero y varianza:

$$E \left[\left| \int_T f(t) dW_t \right|^2 \right] = \sum_{j=1}^n a_j^2 (t_{j+1} - t_j) = \int_T |f(t)|^2 dt$$

¹ Es considerada una versión estocástica del teorema fundamental del cálculo.

Definición B.10: Supongamos que el proceso Z^f es martingala y fijemos $T > 0$. Definimos la probabilidad $\tilde{\mathbb{P}}_T$ sobre \mathfrak{F}_T por:

$$\tilde{\mathbb{P}}_T(A) = E \left[1_A Z_T^f \right] \quad \forall A \in \mathfrak{F}_T \quad (\text{B.1})$$

Como Z^f es martingala se sigue:

$$\tilde{\mathbb{P}}_t(A) = \tilde{\mathbb{P}}_T(A) \quad \forall A \in \mathfrak{F}_t, \quad 0 \leq t \leq T$$

Teorema B.11. Teorema de Girsanov: Sea $f \in P^{1,d}$ tal que el proceso Z^f es martingala. Entonces el proceso $\{\tilde{W}_t\}_{0 \leq t \leq T}$ definido por

$$\tilde{W}_t = W_t - \int_0^t f(s) ds, \quad 0 \leq t \leq T$$

es un movimiento \mathfrak{F}_t – browniano d – dimensional con respecto a la probabilidad $\tilde{\mathbb{P}}_T$ definida en (B.1).

Apéndice C: Correspondencia del filtro de Kalman con el método de mínimos cuadrados recursivos

A continuación explicaremos cómo la estimación por el algoritmo de mínimos cuadrados recursivos es un caso particular especial del filtro de Kalman.

Supóngase que existen $p-1$ regresores, variables independientes o variables de regresión x_1, x_2, \dots, x_{p-1} y n observaciones o mediciones $(x_{i1}, x_{i2}, \dots, x_{i(p-1)}, y_i)$ para $i = 1, 2, \dots, n$, $n > p$ y que el modelo que relaciona esos regresores con la respuesta o variable independiente y es:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{i(p-1)} + \varepsilon_i$$

donde ε_i es un término de error aleatorio en el modelo de pronóstico.

Este modelo es un sistema de n ecuaciones y puede expresarse en notación matricial como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

El estimador de mínimo error cuadrático medio de $\boldsymbol{\beta}$ usando n observaciones es:

$$\hat{\boldsymbol{\beta}}(n) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

La matriz de covarianzas de $\hat{\boldsymbol{\beta}}(n)$ es:

$$\text{cov}(\hat{\boldsymbol{\beta}}(n)) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Los supuestos del método son: $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma^2$, $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ para $i \neq j$ y $E(x_i \varepsilon_j) = 0$.

La ley recursiva del método de mínimos cuadrados permite actualizar $\hat{\boldsymbol{\beta}}(n)$ cada vez que añadimos una observación adicional y consta de las ecuaciones:

$$\hat{\boldsymbol{\beta}}(n+1) = \hat{\boldsymbol{\beta}}(n) + \text{cov}(\hat{\boldsymbol{\beta}}(n+1)) \mathbf{x}_{n+1} (y_{n+1} - \mathbf{x}'_{n+1} \hat{\boldsymbol{\beta}}(n))$$

$$\text{cov}(\hat{\boldsymbol{\beta}}(n+1)) = \text{cov}(\hat{\boldsymbol{\beta}}(n)) - \frac{\text{cov}(\hat{\boldsymbol{\beta}}(n)) \mathbf{x}_{n+1} \mathbf{x}'_{n+1} \text{cov}(\hat{\boldsymbol{\beta}}(n))}{1 + \mathbf{x}'_{n+1} \text{cov}(\hat{\boldsymbol{\beta}}(n)) \mathbf{x}_{n+1}}$$

Otro enfoque muy común para el método de mínimos cuadrados recursivos consiste en que la estimación en cada instante es actualizada con un incremento proporcional al error residual obtenido en el instante anterior. Se utiliza el vector de ganancia¹ \mathbf{k}_{n+1} como factor de ponderación del error residual preliminar.

$$\mathbf{k}_{n+1} = \frac{\text{cov}(\hat{\boldsymbol{\beta}}(n)) \mathbf{x}_{n+1}}{1 + \mathbf{x}'_{n+1} \text{cov}(\hat{\boldsymbol{\beta}}(n)) \mathbf{x}_{n+1}} \quad (\text{C.1})$$

$$\hat{\boldsymbol{\beta}}(n+1) = \hat{\boldsymbol{\beta}}(n) + \mathbf{k}_{n+1} (y_{n+1} - \mathbf{x}'_{n+1} \hat{\boldsymbol{\beta}}(n)) \quad (\text{C.2})$$

$$\text{cov}(\hat{\boldsymbol{\beta}}(n+1)) = \text{cov}(\hat{\boldsymbol{\beta}}(n)) - \mathbf{k}_{n+1} \mathbf{x}'_{n+1} \text{cov}(\hat{\boldsymbol{\beta}}(n)) \quad (\text{C.3})$$

con las condiciones iniciales:

$$\text{cov}(\hat{\boldsymbol{\beta}}(0)) = C\mathbf{I}$$

$$\hat{\boldsymbol{\beta}}(0) = \boldsymbol{\beta}^0$$

donde C es una constante grande, por ejemplo $C \approx 10^6$.

Ahora considérese el siguiente caso especial de espacio de estados invariante en el tiempo:

¹ La notación \mathbf{k}_{n+1} proviene de que a este término suelen llamarle *ganancia de Kalman* debido a la relación que, como veremos, el método de mínimos cuadrados recursivos guarda con el llamado filtro de Kalman.

$$\boldsymbol{\beta}_k = \mathbf{I} \cdot \boldsymbol{\beta}_{k-1} + 0 + 0 \quad (\text{C.4})$$

$$y_k = \mathbf{x}'_k \boldsymbol{\beta}_k + 0 + \varepsilon_k \quad (\text{C.5})$$

Vemos que (C.4) y (C.5)¹ son un caso especial de (2.1.1) y (2.1.2) en el cual:

- \mathbf{u}_k es nulo
- \mathbf{v}_k es nulo y por ende también su matriz de covarianzas \mathbf{V}_k
- \mathbf{x}_k se llama $\boldsymbol{\beta}_k$
- \mathbf{y}_k es un escalar denotado y_k
- \mathbf{w}_k es un escalar nombrado ε_k tal que $E(\varepsilon_k^2) = 1$
- $\mathbf{A}_k = \mathbf{I}$
- \mathbf{C}_k es un vector fila llamado \mathbf{x}'_k

Ahora observe que:

$$\begin{aligned} \mathbf{W}_k &= \text{cov}(\varepsilon_k) = E\left((\varepsilon_k - E(\varepsilon_k))(\varepsilon_k - E(\varepsilon_k))' \right) = \\ &= E\left((\varepsilon_k - 0)(\varepsilon_k - 0)' \right) = E(\varepsilon_k \varepsilon_k') = E(\varepsilon_k^2) = 1 \end{aligned}$$

Considerando las peculiaridades mencionadas, las ecuaciones del filtro de Kalman son:

$$\hat{\boldsymbol{\beta}}_{k|k-1} = \hat{\boldsymbol{\beta}}_{k-1|k-1}$$

$$\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1|k-1}$$

$$\mathbf{K}_k = \frac{\mathbf{P}_{k|k-1} \mathbf{x}_k}{1 + \mathbf{x}'_k \mathbf{P}_{k|k-1} \mathbf{x}_k} \quad (\text{C.6})$$

$$\hat{\boldsymbol{\beta}}_{k|k} = \hat{\boldsymbol{\beta}}_{k|k-1} + \mathbf{K}_k (y_k - \mathbf{x}'_k \hat{\boldsymbol{\beta}}_{k|k-1}) \quad (\text{C.7})$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{x}'_k \mathbf{P}_{k|k-1} \quad (\text{C.8})$$

En las tres ecuaciones anteriores ya podemos reconocer el algoritmo de mínimos cuadrados recursivos. Para $\sigma^2 = 1$ las ecuaciones (C.1), (C.2) y (C.3) son (C.6), (C.7) y (C.8).

¹ La ecuación (C.5) podría no ser vista con la estructura del llamado *filtrado adaptativo* pues \mathbf{x}'_k es el vector de entradas al filtro, tomando el lugar de la matriz \mathbf{C}_k del sistema en vez de la entrada \mathbf{u}_k al sistema.

Apéndice D: Inversión de matrices mediante la descomposición de Cholesky

Matrices definidas no negativas y matrices definidas positivas

Una matriz *definida no negativa*, denotada por $S \geq 0$, es una matriz simétrica

$$\underbrace{S}_{n \times n} = \underbrace{A'}_{n \times m} \underbrace{A}_{m \times n} \quad (D.1)$$

que cumple que dado cualquier vector columna x no nulo de dimensión n puede construirse otro vector

$$y = Ax$$

tal que el escalar

$$x'Sx = x'(A'A)x = (x'A')(Ax) = (xA)'(Ax) = y'y = \sum_{i=1}^m y_i^2 \geq 0$$

Cuando una matriz $S_{n \times n}$ cumple que para cualquier vector columna x no nulo de dimensión n siempre se tiene

$$x'Sx > 0$$

se dice entonces que esa matriz S es definida positiva y se denota $S > 0$.

También podemos caracterizar a una matriz definida no negativa como aquella cuyos valores propios nunca son negativos.¹ Análogamente, una matriz es definida positiva si sus valores propios siempre son positivos.² Recuerde que los valores propios o raíces características de una matriz cuadrada M son las raíces de su polinomio característico:

$$p(\lambda) = \det(M - \lambda I)$$

Cabe mencionar el siguiente cuadro, que sugiere unas fórmulas sencillas para saber el signo de una forma cuadrática.

Tabla D.1: Fórmulas para determinar el signo de matrices cuadráticas

Menores principales conducentes de una matriz A cuadrada:	
Determinante de orden 1: Hasta primera fila y primera columna.	$A_1 = a_{11} $

¹ Esto es, son cero o positivos.

² Esto es, son mayores que cero.

Determinante de orden 2: Hasta segunda fila y segunda columna.	$\mathbf{A}_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$
Determinante de orden 3: Hasta tercera fila y tercera columna.	$\mathbf{A}_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$
Etcétera.	
Condición necesaria y suficiente para \mathbf{A} definida positiva:	$\mathbf{A}_1 > 0, \mathbf{A}_2 > 0, \mathbf{A}_3 > 0, \mathbf{A}_4 > 0, \dots$
Condición necesaria y suficiente para \mathbf{A} definida negativa:	$\mathbf{A}_1 < 0, \mathbf{A}_2 > 0, \mathbf{A}_3 < 0, \mathbf{A}_4 > 0, \dots$
Condición necesaria para \mathbf{A} semidefinida:	$ \mathbf{A} = 0$
Condición suficiente para \mathbf{A} semidefinida positiva:	$\mathbf{A}_1 > 0, \mathbf{A}_2 > 0, \mathbf{A}_3 > 0, \dots, \mathbf{A}_r > 0$ $\mathbf{A}_{r+1} = 0, \dots, \mathbf{A}_n = 0$ Donde: $r = \text{rango}(\mathbf{A})$
Condición suficiente para \mathbf{A} semidefinida negativa:	$\mathbf{A}_1 < 0, \mathbf{A}_2 > 0, \mathbf{A}_3 < 0, \mathbf{A}_4 > 0, \dots$ hasta \mathbf{A}_r $\mathbf{A}_{r+1} = 0, \dots, \mathbf{A}_n = 0$ Donde: $r = \text{rango}(\mathbf{A})$
Para \mathbf{A} nula:	Puros ceros en \mathbf{A}

Descomposición de matrices definidas positivas en matrices triangulares o descomposición de Cholesky

Para toda $\underbrace{\mathbf{S}}_{n \times n} > 0$ simétrica siempre existe una matriz triangular superior o inferior $\underbrace{\mathbf{T}}_{n \times n}$ tal que

$$\mathbf{S} = \mathbf{T}'\mathbf{T}$$

Resolviendo el sistema de ecuaciones resultante de la expresión anterior se encuentra que los elementos de \mathbf{T} son:

$$t_{ij} = \begin{cases} 0 & \text{si } i > j \\ + \sqrt{s_{ii} - \sum_{k=1}^{i-1} t_{ki}^2} & \text{si } i = j \\ \frac{s_{ij} - \sum_{k=1}^{i-1} t_{ki} t_{kj}}{t_{ii}} & \text{si } i < j \end{cases}$$

Habrá que calcular dichos elementos por filas, de izquierda a derecha y de arriba a abajo.

No se puede aplicar con éxito la descomposición de Cholesky en matrices que solamente son definidas no negativas, pues surgen problemas como raíces cuadradas de números negativos. Un proceso exitoso de descomposición de Cholesky nos indica que la matriz simétrica que estamos tratando es definida positiva. No obstante, existen situaciones límite como cuando tratamos de obtener raíces cuadradas de números negativos muy próximos a cero y dependemos de la precisión de nuestra computadora.

Inversión de una matriz triangular

La inversa de una matriz triangular superior \mathbf{T} también es triangular superior. Sea:

$$\mathbf{T}\mathbf{T}^{-1} = \mathbf{I}$$

Resolviendo el sistema de ecuaciones resultante de la ecuación anterior se encuentra que los elementos de \mathbf{T}^{-1} son:

$$t_{ij}^{(-1)} = \begin{cases} 0 & \text{si } i > j \\ \frac{1}{t_{ii}} & \text{si } i = j \\ -\frac{\sum_{k=i+1}^j t_{ik} t_{kj}^{(-1)}}{t_{ii}} & \text{si } i < j \end{cases}$$

Tendremos que calcular dichos elementos por columnas de izquierda a derecha y de abajo hacia arriba: Se empieza con el elemento ubicado en la diagonal principal.

La condición necesaria y suficiente para que \mathbf{T}^{-1} exista es que \mathbf{T} no tenga ningún elemento diagonal nulo. Esto está en acuerdo con la condición de que el determinante de \mathbf{T} no sea cero.

Inversión de una matriz A

Suponga que queremos invertir la matriz regular $\underbrace{\mathbf{A}}_{n \times n}$. Dado que $\mathbf{S} = \mathbf{A}'\mathbf{A} > 0$ existirá su descomposición de Cholesky, $\mathbf{S} = \mathbf{T}'\mathbf{T}$ y su inversa:

$$\mathbf{S}^{-1} = (\mathbf{T}'\mathbf{T})^{-1} = \mathbf{T}^{-1}(\mathbf{T}')^{-1} \quad (\text{D.2})$$

Por (D.1) y (D.2) obtenemos

$$\mathbf{S}^{-1} = (\mathbf{A}'\mathbf{A})^{-1} = \mathbf{A}^{-1}(\mathbf{A}')^{-1}$$

De (D.2) y la igualdad anterior se puede ver que:

$$\mathbf{A}^{-1}(\mathbf{A}')^{-1} = \mathbf{T}^{-1}(\mathbf{T}')^{-1}$$

Multiplicando \mathbf{A}' en cada miembro de la ecuación anterior por la derecha finalmente obtenemos:

$$\mathbf{A}^{-1} = \mathbf{T}^{-1}(\mathbf{T}')^{-1}\mathbf{A}'$$

Por una propiedad de la inversa de una matriz, la expresión previa es:

$$\mathbf{A}^{-1} = \mathbf{T}^{-1}(\mathbf{T}^{-1})'\mathbf{A}'$$

Resumen del algoritmo

Recapitulando, el algoritmo basado en la descomposición de Cholesky, para invertir una matriz $\underbrace{\mathbf{A}}_{n \times n}$ de determinante no nulo, es el siguiente:

1. Especificar la matriz \mathbf{A} que queremos invertir.
2. Construir la matriz $\mathbf{S} = \mathbf{A}'\mathbf{A}$
3. Obtener la matriz \mathbf{T} de la descomposición de Cholesky de \mathbf{S} y pasar al paso 4. Si no podemos obtener \mathbf{T} es porque \mathbf{S} tiene cuando menos un valor propio nulo: La matriz \mathbf{S} cumple lo necesario para ser no negativa pero no para ser definida positiva y \mathbf{A} no es invertible.
4. Computar la matriz \mathbf{T}^{-1}
5. Calcular $\mathbf{A}^{-1} = \mathbf{T}^{-1}(\mathbf{T}^{-1})'\mathbf{A}'$

Apéndice E: Métodos de optimización para funciones diferenciables no lineales de varias variables en problemas no restringidos

Aplicabilidad de los métodos para funciones cuadráticas, en funciones diferenciables de tipo general

Recordemos que una función $f(\mathbf{x})$ es cuadrática si se puede representar como

$$f(\mathbf{x}) = a + \mathbf{b}'\mathbf{x} + \mathbf{x}'\mathbf{A}\mathbf{x}$$

donde a es un escalar, \mathbf{b} un vector y \mathbf{A} una matriz simétrica.

Dada una función diferenciable de varias variables, la serie de expansión de Taylor en torno al óptimo $\boldsymbol{\psi}^*$ dice que:

$$f(\boldsymbol{\psi}) \approx f(\boldsymbol{\psi}^*) + (\boldsymbol{\psi} - \boldsymbol{\psi}^*)' \mathbf{S}(\boldsymbol{\psi}^*) + \frac{1}{2} (\boldsymbol{\psi} - \boldsymbol{\psi}^*)' \mathbf{H}(\boldsymbol{\psi}^*) (\boldsymbol{\psi} - \boldsymbol{\psi}^*)$$

Dado que $\boldsymbol{\psi}^*$ satisface la condición necesaria de optimalidad, $\mathbf{S}(\boldsymbol{\psi}^*) = 0$, se tiene:

$$f(\boldsymbol{\psi}) \approx f(\boldsymbol{\psi}^*) + \frac{1}{2} (\boldsymbol{\psi} - \boldsymbol{\psi}^*)' \mathbf{H}(\boldsymbol{\psi}^*) (\boldsymbol{\psi} - \boldsymbol{\psi}^*) \quad (\text{E.1})$$

Dado que también se cumple la condición suficiente de optimalidad de que $\mathbf{H}(\boldsymbol{\psi}^*)$ sea definida positiva o definida negativa, $f(\boldsymbol{\psi})$ se comporta como una función cuadrática en el entorno del punto $\boldsymbol{\psi}^*$. De aquí se infiere, que cualquier método que se comporte eficientemente con una forma cuadrática, lo hará también con una función de tipo general diferenciable, ya que por la serie de expansión de Taylor cualquier función diferenciable de tipo general puede aproximarse por la forma cuadrática (E.1).¹

Método de Newton

La primera versión de este método data de 1669. Aquí presentamos un algoritmo con diversas mejoras.

Sea $\boldsymbol{\psi}$ es un vector de dimensiones $p \times 1$.

1. Sugerir un $\boldsymbol{\psi}^0$ inicial y fijar escalares δ y ε , pequeños y positivos, como 1×10^{-4} .
2. Para minimizar f se requiere que $\mathbf{H}(\boldsymbol{\psi}^k)$ sea definida positiva. Para maximizar f se requiere que $\mathbf{H}(\boldsymbol{\psi}^k)$ sea definida negativa. Una forma de volver definida positiva a una matriz es sumarle $c \cdot \mathbf{I}$, donde \mathbf{I} es la matriz identidad y c es un escalar positivo. El escalar más pequeño que hay que añadir ha de ser ligeramente mayor que el menor autovalor de la matriz que queremos hacer definida positiva. Similarmente, para volver definida negativa a una matriz se le suma $-c \cdot \mathbf{I}$.
3. Calcular la dirección de búsqueda:

$$\mathbf{r}^k = -\mathbf{H}^{-1}(\boldsymbol{\psi}^k) \mathbf{S}(\boldsymbol{\psi}^k)$$

4. Escoger la longitud de búsqueda $0 < \alpha^k < 1$ que ayude a maximizar $f(\boldsymbol{\psi}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección ascendente, o bien a minimizar $f(\boldsymbol{\psi}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección descendente.

¹ Juan Prawda W., *Métodos y modelos de investigación de operaciones* (México : Limusa-Noriega Editores, 2000), vol. 1, p. 661.

5. Actualizar:

$$\hat{\Psi}^{k+1} = \hat{\Psi}^k + \alpha^k \mathbf{r}^k$$

6. Si $\|\hat{\Psi}^{k+1} - \hat{\Psi}^k\| < \varepsilon(1 + \|\hat{\Psi}^k\|)$ ir al paso 7. Si no, volver al paso 2.

7. Si $\|\mathbf{S}(\hat{\Psi}^{k+1})\| < \delta(1 + \|\mathbf{S}(\hat{\Psi}^{k+1})\|)$ finalizar el algoritmo y reportar éxito. Si no, reportar convergencia a un punto no óptimo.

En funciones cuadráticas el método de Newton converge en una sola iteración.

Los siguientes tres métodos que veremos en este apéndice evitan el laborioso cálculo de la matriz hessiana y eso los hace muy prácticos. Los algoritmos de DFP y BFGS son considerados los más eficientes para calcular puntos óptimos en funciones diferenciables de varias variables. El de Fletcher-Reeves no es tan eficiente como esos dos pero requiere aun menos memoria de cómputo.¹

Método de Davidon-Fletcher-Powell (DFP) o de direcciones conjugadas (1963)

Sea ψ es un vector de dimensiones $p \times 1$. Este método ofrece una estimación simétrica $\mathbf{D}(\psi)$ de la inversa del hessiano. La estimación se conserva definida negativa o definida positiva a lo largo de las iteraciones según propongamos al comienzo del algoritmo. Esto garantiza que con cada iteración nos iremos acercando cada vez más a nuestro propósito² de maximizar, o bien de minimizar, por lo cual obtendremos convergencia hacia algún punto, óptimo o no.

1. Sugerir un $\hat{\Psi}^0$ inicial y fijar escalares δ y ε , pequeños y positivos.
2. Sugerir una $\mathbf{D}(\hat{\Psi}^0)$ definida negativa si queremos una dirección ascendente o definida positiva si queremos una dirección descendente. Si no hay información disponible, seleccionar $\mathbf{D}(\hat{\Psi}^0) = -\mathbf{I}$ para una definida negativa o $\mathbf{D}(\hat{\Psi}^0) = \mathbf{I}$ para una definida positiva. Entiéndase por \mathbf{I} la matriz identidad.
3. Calcular la dirección de búsqueda $\mathbf{r}^k = -\mathbf{D}(\hat{\Psi}^k)\mathbf{S}(\hat{\Psi}^k)$.
4. Escoger la longitud de búsqueda $0 < \alpha^k < 1$ que ayude a maximizar $f(\hat{\Psi}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección ascendente, o bien a minimizar $f(\hat{\Psi}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección descendente.
5. Actualizar $\hat{\Psi}^{k+1} = \hat{\Psi}^k + \alpha^k \mathbf{r}^k$.
6. Actualizar la estimación de la inversa del hessiano con:

¹ Prawda W., *Métodos y modelos de investigación de operaciones*, vol. 1, p. 675.

² *Ibíd.*, p. 669.

$$\mathbf{u}^{k+1} = \boldsymbol{\psi}^{k+1} - \boldsymbol{\psi}^k$$

$$\mathbf{s}^{k+1} = \mathbf{S}(\boldsymbol{\psi}^{k+1}) - \mathbf{S}(\boldsymbol{\psi}^k)$$

$$\mathbf{D}(\boldsymbol{\psi}^{k+1}) = \mathbf{D}(\boldsymbol{\psi}^k) + \frac{\mathbf{u}^{k+1} (\mathbf{u}^{k+1})'}{(\mathbf{u}^{k+1})' \mathbf{s}^{k+1}} - \frac{\mathbf{D}(\boldsymbol{\psi}^k) \mathbf{s}^{k+1} (\mathbf{s}^{k+1})' \mathbf{D}(\boldsymbol{\psi}^k)}{(\mathbf{s}^{k+1})' \mathbf{D}(\boldsymbol{\psi}^k) \mathbf{s}^{k+1}}$$

7. Si $\|\boldsymbol{\psi}^{k+1} - \boldsymbol{\psi}^k\| < \varepsilon (1 + \|\boldsymbol{\psi}^k\|)$ ir al paso 8. Si no, volver al paso 3.
8. Si $\|\mathbf{S}(\boldsymbol{\psi}^{k+1})\| < \delta (1 + \|\mathbf{S}(\boldsymbol{\psi}^{k+1})\|)$ finalizar el algoritmo y reportar éxito. Si no, reportar convergencia a un punto no óptimo.

En una función cuadrática este método converge al óptimo cuando mucho en p iteraciones y la estimación $\mathbf{D}(\boldsymbol{\psi}^{k+1})$ es idéntica que $\mathbf{H}^{-1}(\boldsymbol{\psi}^{k+1})$ desde $k = p - 1$.

Método de Fletcher-Reeves o del gradiente conjugado (1964)

Sea $\boldsymbol{\psi}$ un vector de dimensiones $p \times 1$.

1. Sugerir un $\boldsymbol{\psi}^0$ inicial y fijar escalares δ y ε , pequeños y positivos. Además calcular la primera dirección de búsqueda:

$$\mathbf{r}^0 = -\mathbf{S}(\boldsymbol{\psi}^0)$$

2. Escoger la longitud de búsqueda $0 < \alpha^k < 1$ que ayude a maximizar $f(\boldsymbol{\psi}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección ascendente, o bien a minimizar $f(\boldsymbol{\psi}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección descendente.
3. Actualizar:

$$\boldsymbol{\psi}^{k+1} = \boldsymbol{\psi}^k + \alpha^k \mathbf{r}^k$$

4. Si $\|\boldsymbol{\psi}^{k+1} - \boldsymbol{\psi}^k\| < \varepsilon (1 + \|\boldsymbol{\psi}^k\|)$ ir al paso 5. Si no, ir al paso 6.
5. Si $\|\mathbf{S}(\boldsymbol{\psi}^{k+1})\| < \delta (1 + \|\mathbf{S}(\boldsymbol{\psi}^{k+1})\|)$ finalizar el algoritmo y reportar éxito. Si no, reportar convergencia a un punto no óptimo.
6. Calcular el factor de desviación:

$$\beta^k = \frac{\mathbf{S}'(\boldsymbol{\psi}^{k+1}) \mathbf{S}(\boldsymbol{\psi}^{k+1})}{\mathbf{S}'(\boldsymbol{\psi}^k) \mathbf{S}(\boldsymbol{\psi}^k)}$$

7. Actualizar la dirección de búsqueda:

$$\mathbf{r}^{k+1} = -\mathbf{S}(\hat{\boldsymbol{\psi}}^{k+1}) + \beta^k \mathbf{r}^k$$

8. Volver al paso 2.

Para una función cuadrática, con este algoritmo es posible encontrar el óptimo en p iteraciones.

El *método de Polak-Ribiere* es muy parecido a éste pero usa el factor de desviación:

$$\beta^k = \frac{(\mathbf{S}(\hat{\boldsymbol{\psi}}^{k+1}) - \mathbf{S}(\hat{\boldsymbol{\psi}}^k))' \mathbf{S}(\hat{\boldsymbol{\psi}}^{k+1})}{\mathbf{S}'(\hat{\boldsymbol{\psi}}^k) \mathbf{S}(\hat{\boldsymbol{\psi}}^k)}$$

Método de Broyden-Fletcher-Goldfarb-Shanno (BFGS) o de secante con actualización (1970)

Sea $\boldsymbol{\psi}$ es un vector de dimensiones $p \times 1$. Este método ofrece una estimación simétrica $\hat{\mathbf{H}}(\boldsymbol{\psi})$ del hessiano. La estimación se conserva definida negativa o definida positiva a lo largo de las iteraciones según proponamos al comienzo del algoritmo. Esto garantiza que con cada iteración nos iremos acercando cada vez más a nuestro propósito de maximizar, o bien de minimizar, por lo cual obtendremos convergencia hacia algún punto, óptimo o no.

1. Sugerir un $\hat{\boldsymbol{\psi}}^0$ inicial y fijar escalares δ y ε , pequeños y positivos.
2. Sugerir una $\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}}^0)$ definida negativa si queremos una dirección ascendente o definida positiva si queremos una dirección descendente. Si no hay información disponible, seleccionar $\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}}^0) = -\mathbf{I}$ para una definida negativa o $\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}}^0) = \mathbf{I}$ para una definida positiva. Entiéndase por \mathbf{I} la matriz identidad.
3. Calcular la dirección de búsqueda $\mathbf{r}^k = -\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\psi}}^k) \mathbf{S}(\hat{\boldsymbol{\psi}}^k)$.
4. Escoger la longitud de búsqueda $0 < \alpha^k < 1$ que ayude a maximizar $f(\hat{\boldsymbol{\psi}}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección ascendente, o bien a minimizar $f(\hat{\boldsymbol{\psi}}^k + \alpha^k \mathbf{r}^k)$ si queremos una dirección descendente.
5. Actualizar $\hat{\boldsymbol{\psi}}^{k+1} = \hat{\boldsymbol{\psi}}^k + \alpha^k \mathbf{r}^k$.
6. Actualizar la estimación del hessiano con:

$$\mathbf{u}^{k+1} = \hat{\boldsymbol{\psi}}^{k+1} - \hat{\boldsymbol{\psi}}^k$$

$$\mathbf{s}^{k+1} = \mathbf{S}(\hat{\boldsymbol{\psi}}^{k+1}) - \mathbf{S}(\hat{\boldsymbol{\psi}}^k)$$

$$\tilde{\mathbf{H}}(\boldsymbol{\psi}^{k+1}) = \tilde{\mathbf{H}}(\boldsymbol{\psi}^k) + \frac{\mathbf{s}^{k+1}(\mathbf{s}^{k+1})'}{(\mathbf{s}^{k+1})' \mathbf{u}^{k+1}} - \frac{\tilde{\mathbf{H}}(\boldsymbol{\psi}^k) \mathbf{u}^{k+1} (\mathbf{u}^{k+1})' \tilde{\mathbf{H}}(\boldsymbol{\psi}^k)}{(\mathbf{u}^{k+1})' \tilde{\mathbf{H}}(\boldsymbol{\psi}^k) \mathbf{u}^{k+1}}$$

7. Si $\|\boldsymbol{\psi}^{k+1} - \boldsymbol{\psi}^k\| < \varepsilon (1 + \|\boldsymbol{\psi}^k\|)$ ir al paso 8. Si no, volver al paso 3.
8. Si $\|\mathbf{S}(\boldsymbol{\psi}^{k+1})\| < \delta (1 + \|\mathbf{S}(\boldsymbol{\psi}^{k+1})\|)$ finalizar el algoritmo y reportar éxito. Si no, reportar convergencia a un punto no óptimo.

En una función cuadrática este método converge al óptimo cuando mucho en p iteraciones y la estimación $\tilde{\mathbf{H}}(\boldsymbol{\psi}^{k+1})$ es idéntica que $\mathbf{H}(\boldsymbol{\psi}^{k+1})$ desde $k = p - 1$.

Referencias

1. Anderson, B. D. O. y J. B. Moore. *Optimal filtering*. Estados Unidos: Prentice-Hall, 1979.
2. Bayazit, Dervis. "Yield curve estimation and prediction with Vasicek model." Tesis de licenciatura, Middle East Technical University, Estados Unidos, 2004.
3. Castillo R., Augusto y Alejandro Valenzuela D. Un modelo de valoración de bonos con rescate anticipado. *Estudios de Administración*, vol. 12, num. 1, 2005, pp. 1-39. Universidad Católica de Chile.
4. Chatterjee, Somnath. *Application of the Kalman filter for estimating continuous time term structure models: The case of UK and Germany*. Reino Unido: University of Glasgow, 2005.
5. Duffie, Darrell y Rui Kan. A yield factor model of interest rates. *Mathematical Finance*, vol. 6, num. 4, 1996, pp. 379 - 406.
6. Galisteo R., Mercedes. "Dinámica de la estructura temporal de tipos de interés: Modelo de tres factores." Tesis doctoral, Universidad de Barcelona, España, 2002.
7. Gómez del V., María de L. "Nuevos planteamientos en modelos unifactoriales de la estructura temporal de los tipos de interés." Tesis doctoral, Universidad de Valladolid, España, 2004.
8. Harvey, Andrew C. *Forecasting, structural time series models and the Kalman filter*. Estados Unidos: Cambridge University Press, 1989.
9. Jamieson B., David. *Affine Term Structure Models: Theory and Implementation*. Canadá: Bank of Canada, 2001. Documento de trabajo 2001-15.
10. Lund, Jesper. "Dynamic models of the term structure of interest rates." Tesis de maestría, Aarhus University, Dinamarca, 1993.
11. ----- *Review of Continuous Time Term Structure Models. Part I: equilibrium models*. Dinamarca: Nykredit Markets, 1997.
12. Prawda W., Juan. *Métodos y modelos de investigación de operaciones*. 2 vols. México : Limusa-Noriega Editores, 2000.
13. Sotoca L., Sonia. "Aplicaciones Económicas del filtro de Kalman y algunas variaciones numéricas: El Filtro de Chandrasekhar." Tesis doctoral, Universidad Complutense de Madrid, España, 1992.
14. Tudor, Constantin. *Procesos Estocásticos*. 3ª ed. México: Sociedad Matemática Mexicana, 2002.
15. Valderrama B., Mariano J., et. al. *Filtrado de Kalman, Aplicaciones en Economía e Ingeniería*. España: EUB, 1996.
16. Zivot, Eric. *State Space Models and the Kalman Filter*. Estados Unidos: University of Washington, 2006.