

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO



Título

**EL MODELO DE PRECIOS DE COMMODITY DE
SCHWARTZ-SMITH Y FILTROS DE KALMAN
CON PANELES DE DATOS DE FUTUROS**

TESIS PARA OPTAR EL GRADO ACADÉMICO DE MAGÍSTER EN
MATEMÁTICAS APLICADAS CON MENCIÓN EN APLICACIONES A
LA ECONOMÍA

AUTOR

Ing. Kendy Brigitte Ocola Agüero

ASESOR

Dr. Johel Victorino Beltrán Ramírez

JURADO

Dr. Jhonatan Samuel Farfán Vargas

Dr. Jorge Richard Chavez Fuentes

Diciembre del 2018

Índice general

Introducción	III
1. El modelo continuo de los precios spot de un commodity	1
1.1. Elementos del modelo	1
1.2. El modelo de precios spot de Schwartz y Smith	7
1.3. Solución del modelo y su estructura probabilística	9
2. El modelo spot ajustado por riesgo neutral y valoración de Futuros	13
2.1. Valoración en mercados incompletos	15
2.2. Contrato Futuro	24
2.3. Modelo spot ajustado al riesgo neutral	30
3. El modelo en la forma espacio estado y propiedades de Markov	38
3.1. Representación espacio estado del modelo	39
3.2. Simulación de precios del modelo	43
3.3. Modelo oculto de Markov	44
4. Filtro de Kalman discreto	52
4.1. Conceptos preliminares	54
4.2. Recursividad de la distribución de filtraje	62
4.3. Filtro de Kalman en el sistema lineal Markov Gausiano	68
5. Estimación paramétrica con Máxima Verosimilitud	83
5.1. Solución por el método del Gradiente	84
5.2. Aplicación al modelo oculto de Markov	86
5.3. Aplicación al modelo de Schwartz y Smith	90
6. Conclusión	94
A. Precios spot simulado	96
Bibliografía	97

Resumen

Este trabajo estudia el modelo estocástico de precios spot de commodity de Schwartz y Smith (2000). Este modelo asume que el precio spot de un commodity S_t es una función de dos factores estocásticos, $\ln(S_t) = \chi_t + \xi_t$, con una dinámica $X_t = (\chi_t, \xi_t)$ descrita por el sistema de ecuaciones diferenciales estocásticas de difusión

$$\begin{aligned}d\chi_t &= -\kappa\chi_t dt + \sigma_\chi dB_t^\chi \\d\xi_t &= \mu_\xi dt + \sigma_\xi dB_t^\xi,\end{aligned}$$

donde el vector (B_t^χ, B_t^ξ) es un \mathbb{P} -proceso Browniano correlacionado con coeficiente $\rho_{\chi\xi}$.

Dado el modelo de precios spot y un conjunto de precios futuros de cierto commodity Y_t , el problema general consiste en calibrar X_t y el conjunto de parámetros $\theta = \{\kappa, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho_{\chi\xi}\}$, considerando de que X_t solo es observable indirectamente a través de Y_t . Para el abordaje de este problema se recurre al método de filtraje estocástico a partir de data observable. El objetivo del filtraje estocástico es calcular la distribución condicional $E[X_t | y_1, \dots, y_t]$ dada una muestra finita (y_1, \dots, y_t) de observaciones discretas de Y_t .

La solución al problema de filtraje no es directo y se basa en tres etapas. Primero, se encuentra la solución de X_t . Segundo, se realiza el cambio de medida de probabilidad \mathbb{P} de nuestro modelo spot por una medida equivalente \mathbb{Q} llamado medida de riesgo neutral, aplicando el Teorema de Girsanov con precios de mercado de riesgo $(\lambda_\chi, \lambda_\xi)$. Con el cambio de medida se obtiene la curva de precios futuros para un T fijo ,aplicando la definición de precios futuros, $F(t, T) = E^\mathbb{Q}(S_T / \mathcal{F}_t)$:

$$Y_t \equiv \log(F_{t,T}) = e^{-\kappa(T-t)}\chi_t + \xi_t + A(t, T),$$

donde $A(t, T)$ es una función determinística.

Luego de determinar la ecuación de futuros del modelo de Schwartz y Smith (2000), en la tercera etapa para n fijo y un panel de datos de futuros $\{F_{t,T_1}, \dots, F_{t,T_n}\}$, se representa el modelo en la forma espacio estado discreto para aplicar el método de Filtro de Kalman en la estimación recursiva del sistema lineal discreto. Con las ecuaciones de filtraje se usa el método de máxima verosimilitud para estimar el conjunto de parámetros θ .

Términos Claves

Valoración de Futuros de Commodities, Teorema de Girsanov, Teoría de Arbitraje en tiempo continuo , Cálculo de Itô, Filtro de Kalman y Modelo Oculto de Markov.

Abstract

This paper studies the stochastic model of commodity spot prices of Schwartz and Smith (2000). This model assumes that the spot price of a commodity S_t is a function of two stochastic factors, $\ln(S_t) = \chi_t + \xi_t$, not directly observable $X_t = (\chi_t, \xi_t)$ described by a system of stochastic diffusion differential equations

$$\begin{aligned}d\chi_t &= -\kappa\chi_t dt + \sigma_\chi dB_t^\chi \\d\xi_t &= \mu_\xi dt + \sigma_\xi dB_t^\xi,\end{aligned}$$

where the vector (B_t^χ, B_t^ξ) is a correlated \mathbb{P} -Brownian process with coefficient $\rho_{\chi\xi}$.

Given the spot price model, the general problem is to calibrate X_t and the set of parameters $\theta = \{\kappa, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho_{\chi\xi}\}$. Since X_t is only indirectly observable through Y_t , the problem of stochastic filtering is addressed. The objective of stochastic filtering is to calculate the conditional distribution $E[X_t | y_1, \dots, y_t]$ given a finite sample (y_1, \dots, y_t) of discrete observations de Y_t .

The solution to the filtering problem is not direct and is based on three stages. First, the solution of X_t is needed to be derived. Second, the probability measure \mathbb{P} of our spot model is changed by an equivalent measure \mathbb{Q} called a neutral risk measure and for this the Girsanov Theorem is applied, using constant risk market prices $(\lambda_\chi, \lambda_\xi)$. From the definition of future prices, $F(t, T) = E^\mathbb{Q}(S_T/\mathcal{F}_t)$, the curve of future prices for a fixed T is derived:

$$Y_t \equiv \log(F_{t,T}) = e^{-\kappa(T-t)}\chi_t + \xi_t + A(t, T),$$

where $A(t, T)$ it is a deterministic function.

After determining the future equation of the model, in the third stage for fixed n and a panel data of future prices $\{F_{t,T_1}, \dots, F_{t,T_n}\}$, the model is represented in the discrete state space form so as to apply the Kalman filter method in the recursive estimation of the discrete linear system. Finally, by means of the maximum likelihood method, the set of parameters θ is estimated .

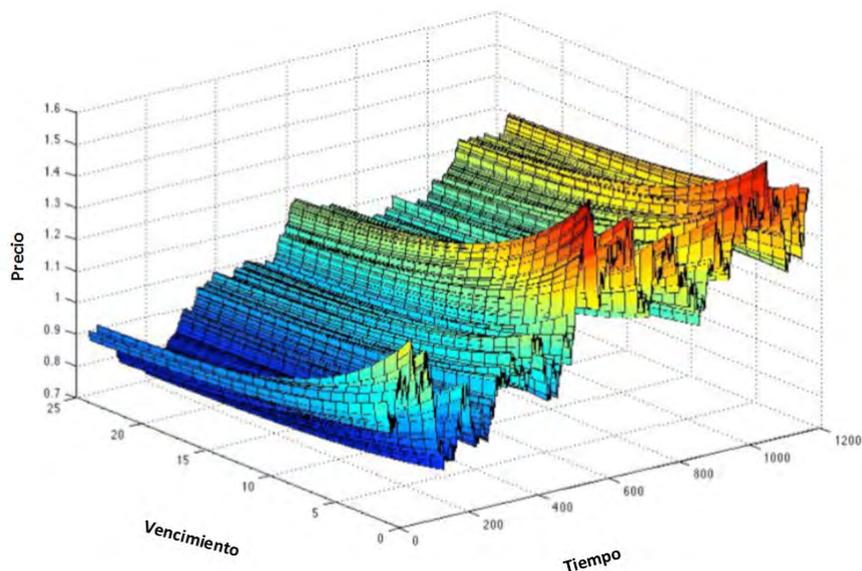
Key Terms

Commodity Futures Valuation, Girsanov's Theorem, Arbitrage theory in continuous time, Calculation of Itô, Kalman Filter and Hidden Markov Model.

Introducción

La modelación estocástica en tiempo continuo de los precios de commodities para valorar derivados financieros tiene como principal antecedente el trabajo de Black y Scholes (1973). No obstante la aplicación del Filtro de Kalman en modelos financieros se inicia con el trabajo de Schwartz (1997) cuando modela la estructura temporal de precios futuros de commodities o curva de futuros. Como se observa en la figura, esta curva refleja la relación que existe entre el tiempo que resta hasta el vencimiento de los diversos contratos y el comportamiento estocástico del precio futuro. El modelo intenta capturar la dinámica de los movimientos de precios de los contratos de corto plazo y de largo vencimiento, introduciendo factores estocásticos que determinarían el precio spot.

Figura 0.1.: Curva de Precios Futuros generado por el Modelo de Schwartz y Smith



En el modelo de Black Scholes adaptado al mercado de commodities, la dinámica del precio spot S_t está dada por

$$\begin{cases} dX_t &= \left(\mu - \frac{\sigma^2}{2} + X_t\right) dt + \sigma dB_t \\ S_t &= \exp(X_t), \end{cases}$$

donde X_t es el vector de factores estocásticos. En este contexto, basta con tomar logaritmos para recuperar la dinámica de los precio spot S_t . Sin embargo, como señala Schwartz (1997),

este modelo implica perfecta correlación entre diferentes contratos futuros contrario a la evidencia empírica del mercado de commodities. El supuesto simplístico de que los precios de commodities siguen un movimiento Browniano geométrico (similar a los precios de los activos financieros como las acciones) no es apropiado con excepción del precio del oro. La razón principal es que los factores económicos de ajuste de oferta y demanda inducen a que los precios de commodities reviertan a su nivel medio histórico¹.

Adicional a este hecho estilizado, en la literatura de modelos spot para la valoración de contratos futuros, el precio spot de los commodities es modelizada como una función de variables de estado no observables llamados factores estocásticos. Como ejemplo tenemos el factor estacional como una variable de estado del precio de la electricidad. Para el precio de commodities almacenables como el cobre, el factor rendimiento por conveniencia de mantener una posición de cierto commodity como inventario vendría ser una variable de estado de relevancia.²

La modelización de la variable rendimiento por conveniencia podemos encontrar en Gibson y Schwartz (1990) quienes plantean un modelo de dos factores con volatilidad constante donde los precios spot y el rendimiento por conveniencia siguen un proceso estocástico conjunto con correlación constante. La variable rendimiento por conveniencia se modela como un dividendo asociado al commodity. Pero debido a que no es muy intuitivo modelar esta variable, Schwartz y Smith (2000) plantean un modelo equivalente pero con simplicidad computacional.

En el modelo de Schwartz y Smith se asume que la volatilidad del precio spot de cierto commodity se descompone en procesos que representan a las variaciones de precios de corto y largo plazo: la idea es que las perturbaciones aleatorias afectan principalmente al precio de corto plazo y en el largo plazo el precio del commodity revierte a su nivel medio. El nivel de precios de largo plazo se modifica ante cambios bruscos en los precios de corto plazo.

El marco de desarrollo del modelo de Schwartz y Smith (2000) se delimita en el sistema establecida por el proceso de difusión X_t , solución de la ecuación diferencial estocástica

$$\begin{cases} dX_t &= (b + AX_t) dt + DdB_t \\ S_t &= \exp(X_t), \end{cases}$$

donde S_t es el precio spot del commodity, X_t es el vector de n estados los cuales generalmente no son observables, B_t es el proceso Browniano de dimensión n , mientras que b , A , D son matrices de apropiado tamaño, los cuales en la aplicación empírica son estimados.

Para obtener la ecuación de la curva de futuros se transforman las variables de estado X_t introduciendo el vector denominado precio de riesgo λ la cual se estima con información de mercado. Esta nueva dinámica de precios se basa en Schwartz (1997) siguiendo el enfoque martingala de valoración de riesgo neutral de la teoría de arbitraje

¹La idea de la reversión a la media es que los precios reaccionan instantáneamente a fluctuaciones de la oferta y demanda, pero una vez que las perturbaciones se desvanecen, los precios tienden a revertir a los niveles medios

²Este rendimiento es positivo cuando el precio del commodity registra un alza luego de un periodo de escasez, proporcionando una potencial prima por la conveniencia de contar con el commodity en ese momento.

$$\begin{cases} dX_t &= (b - \lambda + AX_t) dt + DdB_t \\ S_t &= \exp(X_t). \end{cases}$$

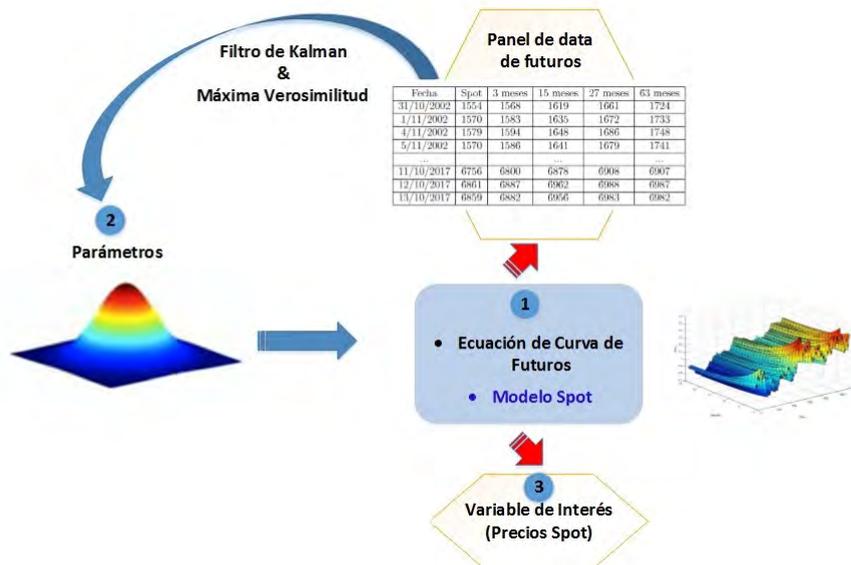
La valoración de los contratos futuros no se realiza con la información histórica de los precios futuros sino con información de precios ajustados por una medida martingala neutral al riesgo. El valor esperado que el inversionista racional espera ganar condicionado a la información disponible hasta el instante t debe ser igual al retorno libre de riesgo. Es decir, $F_{t,T} = E^Q [S_T | \mathcal{F}_t]$ donde $F_{t,T}$ es el precio futuro en el instante t con fecha de vencimiento fijo T , Q es la medida neutral del riesgo y \mathcal{F}_t es la información disponible hasta el tiempo t .

Al asumir que el precio spot está determinado por factores ocultos se plantea el problema de la extracción de información de precios spot a partir de datos observables de precios futuros conocido también como filtraje estocástico, ver figura 0.2. Para ello el modelo se representa en la forma espacio estado discreto para aplicar el método de Filtro de Kalman

$$\begin{cases} X_{\tau_j} &= b_D + A_D X_{\tau_{j-1}} + \omega_{\tau_j} \\ \log F_{\tau_j, T} &= d(T) + c(T) X_{\tau_j} + v_{\tau_j}, \end{cases}$$

donde b_D , A_D , $d(T)$ y $c(T)$ se obtienen de la discretización de la solución explícita del modelo spot y de la curva de futuros, ω es el proceso Browniano discretizado y v es el ruido de los datos de medición.

Figura 0.2.: Problema de Filtraje Estocástico



Objetivos de la tesis

La motivación de la elección del modelo estocástico de precios spot de Schwartz y Smith para la valoración de derivados son dos. Primero, es un modelo estándar en la literatura de modelos spot para la valoración de contratos futuros. El enfoque de descomponer los precios en procesos de corto y largo plazo es aplicado en estudios empíricos, ver Wets y Rios (2015) en el estudio de precios del cobre. Asimismo, el modelo de Schwartz y Smith presenta bondades en la simplicidad computacional para la estimación paramétrica, lo cual permite extender a modelos sofisticados de precios de commodities como el de la electricidad, ver Mastro (2013) y Carmona (2014). Segundo, Sauvageau y Kumral (2016) recientemente realizan un estudio de simulación de precios futuros para un proyecto minero de depósitos de cobre. Lo novedoso es que alternativamente al método tradicional de optimización de Expectation-Maximisation-Algorithm emplean el Algoritmo Genético con Filtros de Kalman, pero no explican con suficiente rigurosidad el problema no trivial de la estimación paramétrica con Filtros de Kalman en el modelo de Schwartz y Smith.

Esta tesis tiene como objetivo general desarrollar el modelo estocástico de precios spot de Schwartz y Smith (2000). Para tal fin, se persiguen tres objetivos específicos.

- Se describe las propiedades matemáticas del modelo mediante la derivación de las soluciones explícitas de las ecuaciones diferenciales estocásticas de los precios spot usando métodos probabilísticos.
- Se aborda el problema de valoración de contratos futuros sobre commodities mediante la derivación de la solución del modelo bajo la medida neutral del riesgo y la definición matemática de precios de contratos futuros.
- Se describe el método del Filtro de Kalman para la estimación de las variables de estado del precio spot del modelo.

La importancia académica de este trabajo es que pone en relieve que la descripción matemática debe ser consistente con las condiciones económicas que presenta el mercado de commodities. Asimismo, presenta la teoría y la implementación práctica del modelo. Este estudio es concordante con la línea de investigación de valoración de opciones reales donde es importante tener información de precios esperados de los precios spot para horizontes de inversión a 30 años. Estudios peruanos, Mendiola y Aguirre (2014) y Tresierra y Carrasco (2016) señalan que la principal dificultad del uso de las opciones reales es la necesidad de desarrollar modelos y herramientas que puedan facilitar su aplicación.

Organización de la tesis

Este trabajo está organizado en cinco capítulos los cuales nos permitirán desarrollar de manera ordenada los tres objetivos específicos planteados.

El capítulo 1 presenta el modelo de los precios spot del commodity según el mercado económico propuesto por Schwartz y Smith (2000). Inicia con la descripción general del proceso estocástico de los precios de un commodity como una combinación lineal finita de factores ocultos. Fijado el contexto, se describe los supuestos del modelo de Schwartz y Smith cuyo

mercado se basa en dos factores estocásticos correlacionados (proceso Browniano Geométrico y Ornstein-Uhlenbeck). El capítulo concluye con la explicación de la solución explícita del modelo empleando el cálculo estocástico de Itô para procesos de Difusión. Este enfoque nos facilita describir la estructura probabilística asociada a la solución conjunta del modelo.

El capítulo 2 muestra la fórmula de valoración de contratos futuros en el mercado financiero de Schwartz y Smith (2000). Se comienza presentando la valoración de contratos en cierto tipo de mercados incompletos llamado modelo de factores no observables. Siguiendo a Schwartz (1997), el enfoque adoptado es el método de las ecuaciones en derivadas parciales donde se muestra cómo es introducido el precio de mercado del riesgo denotado por λ . Gracias al Teorema de Feynman-Kac y el Teorema de Girsanov, la ecuación de valoración es equivalente al de valoración de riesgo neutral. Es decir, se concluye que existe una relación de equivalencia entre λ y \mathbb{Q} . La idea medular del enfoque de valoración es que dentro del modelo solo se requiere que λ debe ser el mismo para todos los contratos, pero su elección es arbitrario. Para diferentes λ conseguiremos diferentes precios de contrato libre de arbitraje. A diferencia de mercados completos, con el modelo no es posible elegir el “correcto” λ , sino que es determinado por los agentes del mercado. Por ejemplo, mientras que Schwartz y Smith asumen que λ es constante, Espend (2011) asume que λ presenta la propiedad de reversión a la media. Sentando las bases de valoración en mercado incompleto, se describe las propiedades que caracterizan el precio de contratos futuros. Se muestra que el proceso de precios futuros es una martingala. En contraste, Burger et.al. (2007, pag 94) señalan que el enfoque de la medida neutral del riesgo en el contexto de precios spot del commodity implica que el proceso de precios spot descontados no necesariamente tiene que ser martingala. Finalmente este capítulo concluye con el ajuste de nuestro modelo spot bajo la medida neutral al riesgo para encontrar la curva de precios futuros.

El capítulo 3 aborda el problema de la representación espacio estado del modelo en estudio. Comienza con la discretización de las variables de estado lo cual es directo como resultado de que la solución de nuestro modelo es explícita y no hay necesidad de aplicar métodos probabilísticos de aproximaciones de una ecuación diferencial estocástica. Luego se estudia estructura probabilística del modelo espacio estado lineal discreto bajo la teoría de procesos de Markov, en particular se presenta las propiedades del modelo oculto de Markov. La motivación del estudio de este modelo es que nuestras variables de estado no son observables directamente. Se concluye este capítulo con la simulación de los precios del modelo.

En el capítulo 4 se describe los fundamentos teóricos que sustenta el algoritmo del filtro de Kalman en el contexto de modelo oculto de Markov, asumiendo que el espacio de parámetros son conocidos. Basado en Catlin (1989), se plantea el problema de filtraje estocástico, el cual se basa en calcular la distribución condicional de X_t a partir de una muestra finita de observaciones. En el contexto de modelo oculto de Markov, siguiendo a Cappé (2009), se presenta la definición de función de verosimilitud de las observaciones y la distribución de suavizamiento dado que de la marginalización de esta medida se obtiene la distribución de filtraje. Luego se presenta la construcción general para la estimación recursiva de las distribuciones de filtraje en el cual se hace uso intensivo de las nociones de distribución condicional. Se muestra la distribución de filtraje para modelos de Markov Gaussianos lineales haciendo uso del teorema instrumental de modelo lineal Gaussiano condicionado. Se discute también que no es trivial mostrar que la distribución de filtraje es también Gaussiano. Se muestra finalmente que para encontrar las ecuaciones de filtraje no es necesario asumir que

los errores del sistema sigan una distribución Gaussiana.

Finalmente, en el capítulo 5 se describe brevemente la estimación paramétrica de la función de máxima verosimilitud de nuestro modelo. Para tal fin, se escoge el método gradiente como el método de optimización. Se concluye con la aplicación de esta herramienta en el contexto de nuestro modelo.



Capítulo 1.

El modelo continuo de los precios spot de un commodity

En esta sección se presenta el modelo de Schwartz y Smith, el cual asume que los precios spot es una función de las variaciones de los precios de corto y largo plazo. Este enfoque es una aproximación estadística usual que descompone la trayectoria del precio en dos factores, uno cíclico cuyos efectos se desvanecen en el tiempo en ausencia de nuevos shocks perturbadores originados por los cambios bruscos en la oferta y demanda de los commodities. El segundo componente representa la tendencia de los precios que captura los efectos permanentes de los shocks determinando la proyección del precio de largo plazo. Una de las principales ventajas de este enfoque es evitar realizar estimaciones de numerosas variables no observables que explicarían la dinámica estocástica de los precios spot.

1.1. Elementos del modelo

Se considera una economía donde se negocian continuamente activos financieros en el intervalo $[0, \bar{T}]$ donde \bar{T} es la fecha de término de toda actividad comercial. La incertidumbre de los precios de los activos financieros es modelizada por medio de un proceso Browniano estándar ¹ de dimensión finita n :

$$B_t = (B_t^1, \dots, B_t^n) : \Omega \rightarrow \mathbb{R}^n,$$

definido sobre un espacio probabilístico completo $(\Omega, \mathcal{F}, \mathbb{P})$, equipado por una familia no-decreciente de σ -álgebras $\mathbb{F} = (\mathcal{F}_t^B)$ donde la filtración generada por (B_t) es

$$\mathcal{F}_t^B = \sigma(B_s : 0 \leq s \leq t) , \quad t \in [0, \bar{T}].$$

Consideraremos para nuestro modelo una filtración aumentada

$$\mathcal{F}_t \triangleq \sigma(\mathcal{F}_t^B \cup N) , \quad t \in [0, \bar{T}] ,$$

¹Los componentes son procesos Brownianos independientes: $dB_t^i \cdot dB_t^j = 0, \forall i \neq j$

donde $N \triangleq \{ A \in \mathcal{F}_{\bar{T}}^B : \mathbb{P}(A) = 0 \}$. El conjunto de filtración (\mathcal{F}_t) satisface las condiciones de ser continua a la derecha, i.e. $\mathcal{F}_t = \mathcal{F}_{t+} = \left(\bigcap_{s>t} \mathcal{F}_s \right)$ y \mathcal{F}_0 es completado con los conjuntos \mathbb{P} -nulos (ver Proposición 2.7.7 de Karatzas y Shreve, 1998). Los procesos de precios de activos riesgosos además de ser continuos, son estrictamente positivos \mathbb{P} -casi ciertamente para todo t y son \mathcal{F}_t -semimartingalas.

Se distinguen dos clases de activos financieros. Por un lado, tenemos a los activos base y por otro lado, a los derivados o reclamos contingentes sobre los cuales se plantea el problema de la valoración. Entre los derivados transados en nuestra economía se encuentran un número finito de contratos futuros sobre un cierto commodity. El proceso de precio spot de este commodity, que denotaremos por S_t , será una función de n factores estocásticos descritos por el vector $X_t = (X_t^1, \dots, X_t^n)$ mediante la siguiente relación

$$S_t = e^{X_t^1 + \dots + X_t^n}, \quad 0 \leq t \leq \bar{T}.$$

Definimos a $(X_t)_{t \geq 0}$ como una familia de variables aleatorias en $(\Omega, \mathcal{F}, \mathbb{P})$ tal que X_t es \mathcal{F}_t -medible, para todo $t \geq 0$; y por lo tanto diremos que $X = (X_t, \mathcal{F}_t)$ es adaptado a (\mathcal{F}_t) . Este proceso pertenece al espacio (D, \mathcal{D}) donde D es el espacio de funciones continuas por la derecha y límites por la izquierda. \mathcal{D} es el σ -álgebra generada por los borelianos de \mathbb{R}^n .

Dada la condición inicial $X_0 = Z$, el proceso X_t es solución de la ecuación diferencial estocástica

$$dX_t^i = \mu_i(X_t^i) dt + \sum_{j=1}^n \sigma_{ij}(X_t^i) dB_t^j, \quad i = 1, \dots, n \quad (1.1)$$

donde Z es una variable \mathcal{F}_0 -medible con $\mathbb{E}[|Z|^2] < \infty$, y para $C > 0$, el vector $\mu : D \rightarrow \mathbb{R}^n$ y la matriz $\sigma : D \rightarrow \mathbb{R}^{n \times n}$ verifican la siguiente condición

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C|x - y|; \quad x, y \in \mathbb{R}^n,$$

con $|\sigma|^2 = \sum |\sigma_{ij}|^2$. Bajo esta condición, por el Teorema 5.2.1 en (Oksendal, 1998), la existencia y unicidad de la solución de 1.1 está garantizada². Además cumple

$$E \left[\int_0^{\bar{T}} |X_t|^2 dt \right] < \infty.$$

La ecuación 1.1 se puede escribir de manera matricial del siguiente modo

²También se muestra $\mathbb{E} \left[\sup_{0 \leq t \leq \bar{T}} |X_t|^2 \right] < C \left[1 + \mathbb{E} \left[|Z|^2 \right] \right]$, para mayor detalle ver Teorema 6.5.3 en Elliott y Ekkehard (2005, pag. 159)

$$\begin{bmatrix} dX_t^1 \\ \vdots \\ dX_t^i \\ \vdots \\ dX_t^n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_n \end{bmatrix} dt + \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{i1} & & \sigma_{in} \\ \vdots & & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix} \begin{bmatrix} dB_t^1 \\ \vdots \\ dB_t^i \\ \vdots \\ dB_t^n \end{bmatrix} \quad (1.2)$$

donde μ_i y σ_i es la fila i del vector $[\mu_i]$ y la matriz volatilidad $[\sigma_{ij}]$ de orden $n \times n$ respectivamente con $1 \leq i \leq n \in \mathbb{N}$.

El modelo con procesos Brownianos estándar tiene su expresión equivalente en términos de procesos Brownianos correlacionados. Esta forma equivalente será utilizada en la representación del modelo de Schwartz y Smith (2000).

Consideramos una matriz (determinística y constante) δ

$$\delta = \begin{bmatrix} \delta_{11} & \cdots & \delta_{1n} \\ \vdots & & \vdots \\ \delta_{i1} & & \delta_{in} \\ \vdots & & \vdots \\ \delta_{n1} & \cdots & \delta_{nn} \end{bmatrix}$$

y definimos el proceso Browniano (\bar{B}_t) como $\bar{B} \triangleq \delta B$, es decir $\bar{B}_t^i = \sum_{j=1}^n \delta_{ij} B_t^j$ con $1 \leq i \leq n$. Se define la matriz de correlación ρ de \bar{B} por

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \rho_{13} & \rho_{23} & 1 & \cdots & \rho_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \rho_{3n} & \cdots & 1 \end{bmatrix}, \quad \rho_{ij} dt = \text{Cov} [d\bar{B}_t^i, d\bar{B}_t^j], \quad (1.3)$$

el cual es un proceso estocástico que toma valores entre -1 y 1 . Se calcula fácilmente que

$$\begin{aligned} \rho_{ij} dt &= E [d\bar{B}_t^i \cdot d\bar{B}_t^j] - E [d\bar{B}_t^i] \cdot E [d\bar{B}_t^j] = E [d\bar{B}_t^i \cdot d\bar{B}_t^j] \\ &= \sum_{k=1}^n \delta_{ik} \delta_{jk} dt, \end{aligned}$$

es decir $\rho = \delta \delta^\top$.

La siguiente proposición nos dice cómo podemos encontrar una matriz δ adecuada a partir de la matriz de correlación ρ para que corresponda $\bar{B} \triangleq \delta B$. Para la prueba ver Proposición 3.12 en Muirhead (2011, pag. 10).

Proposición 1.1.1. (Construcción de procesos Brownianos correlacionados). Sea una matriz $\rho = [\rho_{ij}]_{1 \leq i, j \leq n}$ positiva definida. Un proceso Browniano (\bar{B}_t) con matriz de correlación ρ se construye aplicando una transformación lineal al proceso Browniano (B_t) estándar a través de su descomposición de Cholesky: $\rho = \delta\delta^\top$ con $\delta = [\delta_{ij}]_{1 \leq i, j \leq n}$ tal que

$$\|\delta_i\| = 1, \quad i = 1, \dots, n,$$

comienza de la siguiente forma

$$\delta = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \rho_{12} & \sqrt{1 - \rho_{12}^2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{bmatrix},$$

y los demás elementos δ_{ij} se calculan iterativamente

$$\delta_{i,j} = \begin{cases} \frac{1}{\delta_{j,j}} \left(\rho_{i,j} - \sum_{k=1}^j \delta_{i,j-k} \delta_{i,j-k} \right) & j < i \\ \sqrt{1 - \sum_{k=1}^i \delta_{i,i-k}^2} & j = i \\ 0 & j > i \end{cases}$$

El proceso \bar{B} construido mediante Proposición 1.1.1 se llama vector de proceso Browniano correlacionado con matriz de correlación ρ .

Ejemplo 1.1.2. (Descomposición de procesos Brownianos correlacionados en procesos Brownianos independientes). Sea $(\bar{B}_t^1, \bar{B}_t^2)$ un proceso Browniano con matriz de correlación

$$\rho = \begin{bmatrix} 1 & \rho_* \\ \rho_* & 1 \end{bmatrix}, \quad (-1 \leq \rho_* \leq 1)$$

positiva definida, es decir $d\bar{B}_t^1 d\bar{B}_t^2 = \rho_* dt$. De la Proposición 1.1.1 su correspondiente transformación matricial δ tiene la siguiente forma

$$\delta = \begin{bmatrix} 1 & 0 \\ \rho_* & \sqrt{1 - \rho_*^2} \end{bmatrix}.$$

Con este ejemplo, debe quedar claro que podemos expresar un proceso Browniano

$$\begin{bmatrix} \bar{B}_t^1 \\ \bar{B}_t^2 \end{bmatrix} = \begin{bmatrix} B_t^1 \\ \rho_* B_t^1 + \sqrt{1 - \rho_*^2} B_t^2 \end{bmatrix},$$

correlacionado ρ en términos de un proceso Browniano independiente (B_t^1, B_t^2) .

Observación 1.1.3. Se muestra que \bar{B}_t^2 es un proceso Browniano como resultado de la caracterización de Levy. El cual señala si una local martingala (\bar{B}_t^2) con $t \geq 0$ presenta variación cuadrática $[\bar{B}_t^2, \bar{B}_t^2]_t = t, \forall t \geq 0$, entonces (\bar{B}_t^2) es un movimiento Browniano, ver Karatzas y Shreve (2004, pag. 157).

La Proposición 1.1.1 se puede reescribir en términos de 1.1 de la siguiente manera. Para $i = 1, \dots, n$ definimos

$$\|\sigma_i\| \triangleq \left[\sum_{j=1}^n \sigma_{ij}^2 \right]^{\frac{1}{2}}$$

y asumamos que es diferente de cero. Definamos \bar{B}_t^i por

$$\bar{B}_t^i \triangleq \frac{\sum_{j=1}^n \sigma_{ij} B_t^j}{\|\sigma_i\|} \quad i = 1, \dots, n. \quad (1.4)$$

Puede verse fácilmente que \bar{B}_t^i es un proceso Browniano correlacionado, i.e. $d\bar{B}_t^i d\bar{B}_t^j = \rho_{ij} dt$ donde

$$\rho_{ij} dt = E \left[d\bar{B}_t^i d\bar{B}_t^j \right] = \frac{1}{\|\sigma_i\| \cdot \|\sigma_j\|} \sum_{k=1}^n \sigma_{ik} \sigma_{jk} dt = \frac{\sigma_i \sigma_j^*}{\|\sigma_i\| \cdot \|\sigma_j\|} dt. \quad (1.5)$$

Podemos reescribir la dinámica de X , ver 1.1, como

$$dX_t^i = \mu_i(X_t^i) dt + \|\sigma_i\| d\bar{B}_t^i, \quad i = 1, \dots, n.$$

Resumiendo tenemos el siguiente resultado:

Proposición 1.1.4. (*Proceso Brownianos Correlacionados*). *El sistema en 1.6*

$$dX_t^i = \mu_i(X_t^i) dt + \sum_{j=1}^n \sigma_{ij} dB_t^j, \quad i = 1, \dots, n, \quad (1.6)$$

donde B_1, \dots, B_n son independientes, puede ser equivalentemente representado como

$$dX_t^i = \mu_i(X_t^i) dt + \delta_i(X_t^i) d\bar{B}_t^i, \quad i = 1, \dots, n \quad (1.7)$$

donde $\bar{B}_1, \dots, \bar{B}_n$ presentan matriz correlación ρ . La conexión entre 1.6 y 1.7 está dada por las siguientes expresiones

$$\begin{aligned} \bar{B}_t^i &= \frac{1}{\|\sigma_i\|} \sigma_i B_t, \quad i = 1, \dots, n \\ \delta_i &= \|\sigma_i\|, \quad i = 1, \dots, n \\ \rho_{ij} &= \frac{1}{\|\sigma_i\| \cdot \|\sigma_j\|} \sum_{k=1}^n \sigma_{ik} \sigma_{jk} = \frac{\sigma_i \sigma_j^*}{\|\sigma_i\| \cdot \|\sigma_j\|}, \quad i, j = 1, \dots, n. \end{aligned}$$

Ejemplo 1.1.5. Sea el sistema dado por

$$\begin{aligned} dX_t^1 &= \mu_1 dt + \sigma_{11} dB_t^1 \\ dX_t^2 &= \mu_2 dt + \sigma_{21} \rho_* dB_t^1 + \sigma_{21} \sqrt{1 - \rho_*^2} dB_t^2. \end{aligned} \quad (1.8)$$

Aplicando la Proposición 1.1.4 tenemos que 1.8 se puede rescribir como

$$\begin{aligned} dX_t^1 &= \mu_1 dt + \sigma_{11} d\bar{B}_t^1 \\ dX_t^2 &= \mu_2 dt + \sigma_{21} d\bar{B}_t^2, \end{aligned}$$

donde $\bar{B}_t^1 = B_t^1$ y $\bar{B}_t^2 = \rho_* B_t^1 + \sqrt{1 - \rho_*^2} B_t^2$.

1.2. El modelo de precios spot de Schwartz y Smith

En el modelo de Schwartz y Smith, el proceso (S_t) de precios spot está conformado por dos factores estocásticos. Es decir, la ecuación 1.1 se reduce a $n = 2$ con $X_t = (\chi_t, \xi_t)^\top$ solución del sistema conformado por 1.10b y 1.10c. Para $\rho_{\chi\xi} \in (-1, 1)$, se define

$$B_t^X = B_t^1, \quad y \quad B_t^\xi = \rho_{\chi\xi} B_t^1 + \sqrt{1 - \rho_{\chi\xi}^2} B_t^2, \quad (1.9)$$

de modo que (B_t^X, B_t^ξ) es un proceso Browniano correlacionado (ver el Ejemplo 1.1.2) con elementos de la matriz de varianzas- covarianzas , $C(t)$, dados por

$$\text{Cov} [dB_t^X dB_t^\xi] = \rho_{\chi\xi} dt, \quad \text{Var} [dB_t^X] = \text{Var} [dB_t^\xi] = dt.$$

Luego, el modelo matemático es representado de la siguiente forma

$$\ln(S_t) = \chi_t + \xi_t \quad (1.10a)$$

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dB_t^X \quad (1.10b)$$

$$d\xi_t = \mu_\xi dt + \sigma_\xi dB_t^\xi, \quad (1.10c)$$

con parámetros $\kappa > 0$, $\sigma_\chi > 0$, $\sigma_\xi > 0$ y $\mu_\xi \in \mathbb{R}$.

El factor ξ_t representa las variaciones del precio de largo plazo. Este factor captura el comportamiento del costo marginal de producción del commodity, el cual es afectado por factores de producción, impactos tecnológicos y la escasez de inventario. Matemáticamente, para cierto nivel de precio de equilibrio de largo plazo \bar{S}_t , el factor ξ_t es el logaritmo natural de \bar{S}_t y se asume que sigue la dinámica de un proceso Browniano aritmético, ver 1.10c, donde μ_ξ y σ_ξ denotan la variación esperada y la volatilidad de ξ_t , respectivamente siendo ambas constantes. Por el Lema de Ito para $\bar{S}_t = e^{\xi_t}$, el precio de equilibrio se comporta como un proceso Browniano geométrico con tasa de crecimiento constante $\mu_\xi + \frac{1}{2}\sigma_\xi^2$,

$$d\bar{S}_t = \bar{S}_t \left(\mu_\xi + \frac{1}{2}\sigma_\xi^2 \right) dt + \bar{S}_t \sigma_\xi dB_t^\xi.$$

El factor χ_t se define como las desviaciones del precio spot respecto a su nivel de equilibrio de largo plazo , $\chi_t = \ln(S_t) - \ln(\bar{S}_t)$, representando las desviaciones de corto plazo como resultado de la escasez momentánea de inventario de cierto commodity. Se asume que estas desviaciones son efímeras siguiendo el proceso Ornstein-Uhlenbeck, ver 1.10b, donde χ_t revierte al nivel de precio equilibrio cero con una velocidad κ y volatilidad σ_χ , ambos

constantes³. El modelo asume que las innovaciones aleatorias de la economía generan que ξ_t cambie permanentemente y esta perturbación se transmite via el coeficiente de correlación $\rho_{\chi\xi}$ sobre χ_t , pero este solo tendrá efecto transitorio.

Las condiciones iniciales $X_0 = (\chi_0, \xi_0)$ son dadas y conocidas por el inversionista, con momentos de segundo orden finito, i.e.,

$$\mathbb{E} \chi_0^2 < \infty \quad y \quad \mathbb{E} \xi_0^2 < \infty.$$

Se asume que el vector $(\chi_0, \xi_0)^\top$ y el proceso Browniano (B_t^χ, B_t^ξ) son mutuamente independientes.

El modelo equivalente de 1.10b y 1.10c queda representado de la siguiente manera

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dB_t^1 \tag{1.11a}$$

$$d\xi_t = \mu_\xi dt + \sigma_\xi \rho_{\chi\xi} dB_t^1 + \sigma_\xi \sqrt{1 - \rho_{\chi\xi}^2} dB_t^2 \tag{1.11b}$$

En el siguiente Lema se expresa los precios spot en términos de B_t^1 y B_t^2 .

Lema 1.2.1. *El proceso de precios spot satisface la ecuación diferencial estocástica,*

$$dS_t = (\alpha - \kappa\chi_t) S_t dt + \sigma_1 S_t dB_t^1 + \sigma_2 S_t dB_t^2, \tag{1.12}$$

donde $\alpha = \left(\mu_\xi + \frac{1}{2}\sigma_\xi^2 + \frac{1}{2}\sigma_\chi^2 + \sigma_\chi\sigma_\xi\rho_{\chi\xi} \right)$, $\sigma_1 = \sigma_\xi\rho_{\chi\xi} + \sigma_\chi$ y $\sigma_2 = \sigma_\xi\sqrt{1 - \rho_{\chi\xi}^2}$.

Demostración. Resultado directo de aplicar la fórmula de Ito Multidimensional para $n = 2$ y la función $F(\chi_t, \xi_t) \triangleq \exp(\chi_t + \xi_t) = S_t$.

$$\begin{aligned} dF(\chi_t, \xi_t) &= \left\{ b_1 \frac{\partial F}{\partial \chi} + b_2 \frac{\partial F}{\partial \xi} + \frac{1}{2} \left[(\sigma_{11}^2 + \sigma_{12}^2) \frac{\partial^2 F}{\partial \chi^2} + (\sigma_{21}^2 + \sigma_{22}^2) \frac{\partial^2 F}{\partial \xi^2} \right] \right\} dt \\ &\quad + (\sigma_{21}\sigma_{11} + \sigma_{22}\sigma_{12}) \frac{\partial^2 F}{\partial \chi \partial \xi} dt + \left(\sigma_{11} \frac{\partial F}{\partial \chi} + \sigma_{21} \frac{\partial F}{\partial \xi} \right) dB_t^1 + \left(\sigma_{12} \frac{\partial F}{\partial \chi} + \sigma_{22} \frac{\partial F}{\partial \xi} \right) dB_t^2 \end{aligned}$$

donde de 1.11a y 1.11b se definen $b_1 = -\kappa\chi_t$, $b_2 = \mu_\xi$, $\sigma_{11} = \sigma_\chi$, $\sigma_{12} = 0$, $\sigma_{21} = \sigma_\xi\rho_{\chi\xi}$ y $\sigma_{22} = \sigma_\xi\sqrt{1 - \rho_{\chi\xi}^2}$.

□

³También a χ_t se le interpreta como el rendimiento por conveniencia del commodity, ver Gibson y Schwartz (1990) y Schwartz y Smith (2000, pag. 899).

La ecuación diferencial de S_t en términos de los procesos Brownianos correlacionados se obtiene usando la ecuación 1.9 y el Lema anterior.

Lema 1.2.2. *El proceso de precios spot del modelo 1.10a satisface la ecuación diferencial estocástica,*

$$dS_t = (\tilde{\alpha} - \kappa\chi_t) S_t dt + \sigma_\chi S_t dB_t^\chi + \sigma_\xi S_t dB_t^\xi, \quad (1.13)$$

$$\text{con } \tilde{\alpha} = \left(\mu_\xi + \frac{1}{2}\sigma_\xi^2 + \frac{1}{2}\sigma_\chi^2 + \sigma_\chi\sigma_\xi\rho_{\chi\xi} \right).$$

1.3. Solución del modelo y su estructura probabilística

Para el modelo de Schwartz y Smith, la solución (X_t) es explícita como veremos en el siguiente Lema.

Lema 1.3.1. *Dada la condición inicial (χ_0, ξ_0) , la solución $X_t = (\chi_t, \xi_t)^\top$ del sistema conformado por las ecuaciones 1.11a y 1.11b es*

$$\begin{aligned} \chi_t &= e^{-\kappa t} \chi_0 + \sigma_\chi \int_0^t e^{-\kappa(t-s)} dB_s^1 \\ \xi_t &= \xi_0 + \mu_\xi t + \sigma_\xi \int_0^t \left(\rho_{\chi\xi} dB_s^1 + \sqrt{1 - \rho_{\chi\xi}^2} dB_s^2 \right), \quad t \in [0, \bar{T}]. \end{aligned}$$

Demostración. De las ecuaciones 1.11a y 1.11b, el modelo en notación matricial se expresa

$$dX_t = \left[\begin{pmatrix} 0 \\ \mu_\xi \end{pmatrix} + \begin{pmatrix} -\kappa & 0 \\ 0 & 0 \end{pmatrix} X_t \right] dt + \begin{pmatrix} \sigma_\chi & 0 \\ \sigma_\xi \rho_{\chi\xi} & \sigma_\xi \sqrt{1 - \rho_{\chi\xi}^2} \end{pmatrix} dB_t \quad (1.14)$$

con condición inicial $X_0 = \begin{pmatrix} \chi_0 & \xi_0 \end{pmatrix}^\top$ y

$$X_t \triangleq \begin{pmatrix} \chi_t \\ \xi_t \end{pmatrix} \quad \text{y} \quad B_t \triangleq \begin{pmatrix} B_t^1 \\ B_t^2 \end{pmatrix}.$$

El modelo spot es un sistema lineal de dos ecuaciones diferenciales estocásticas homogéneas, por lo tanto la solución existe (ver Arnold (1992), pag. 128). De la ecuación 1.14, la función matriz de volatilidad $D : [0, \bar{T}] \rightarrow \mathbb{R}^{2 \times 2}$ definido por

$$D \triangleq \begin{pmatrix} \sigma_\chi & 0 \\ \sigma_\xi \rho_{\chi\xi} & \sigma_\xi \sqrt{1 - \rho_{\chi\xi}^2} \end{pmatrix},$$

es trivialmente medible y acotada en $[0, \bar{T}]$ al ser constante. Por lo tanto, el Teorema 8.1.5 en Arnold (1992) asegura que para todo valor inicial c existe una solución única y fuerte. Asimismo, del Corolario 8.2.4 en Arnold (1992) se sigue que esta solución está dada por

$$X_t = e^{tA}c + \int_0^t e^{(t-s)A}b ds + \int_0^t e^{(t-s)A}D dB_s, \quad (1.15)$$

donde las matrices A y b son

$$A \triangleq \begin{pmatrix} -\kappa & 0 \\ 0 & 0 \end{pmatrix}, \quad b \triangleq \begin{pmatrix} 0 \\ \mu_\xi \end{pmatrix}.$$

Para λ_1, λ_2 eigenvalores de la matriz diagonalizable $A = \text{diag}(\lambda_1, \lambda_2) = \text{diag}(-\kappa, 0)$, el cálculo exacto de la matriz exponencial e^{tA} es directo

$$e^{tA} = \text{diag}(e^{t\lambda_1}, e^{t\lambda_2}) = \begin{pmatrix} e^{-\kappa t} & 0 \\ 0 & 1 \end{pmatrix}.$$

Por lo tanto, la solución X_t de la ecuación 1.14 está dada por

$$X_t = \begin{pmatrix} e^{-\kappa t} & 0 \\ 0 & 1 \end{pmatrix} c + \int_0^t \begin{pmatrix} e^{-\kappa(t-s)} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \mu_\xi \end{pmatrix} ds + \int_0^t \begin{pmatrix} e^{-\kappa(t-s)} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_\chi & 0 \\ \sigma_\xi \rho_{\chi\xi} & \sigma_\xi \sqrt{1 - \rho_{\chi\xi}^2} \end{pmatrix} \begin{pmatrix} dB_s^1 \\ dB_s^2 \end{pmatrix}. \quad (1.16)$$

Para concluir, solo resta usar la condición inicial (χ_0, ξ_0) . □

Usando el Lema 1.3.1 y la relación de 1.9, se puede reescribir la solución X_t en términos de los procesos Brownianos correlacionados del siguiente modo

$$\begin{aligned} \chi_t &= e^{-\kappa t} \chi_0 + \sigma_\chi \int_0^t e^{-\kappa(t-s)} dB_s^\chi. \\ \xi_t &= \xi_0 + \mu_\xi t + \sigma_\xi \int_0^t dB_s^\xi. \end{aligned} \quad (1.17)$$

Se observa del Lema 1.3.1, que la solución X_t es escrita en términos de integrales de Ito con integrandos determinísticos. Usando este hecho, las componentes aleatorias del siguiente vector

$$\left[\int_0^t e^{-\kappa(t-s)} dB_s^1, \int_0^t (\rho_{\chi\xi} dB_s^1 + \sqrt{1 - \rho_{\chi\xi}^2} dB_s^2) \right]$$

sigue una distribución normal con media cero y matriz de covarianzas obtenidas de la Isometría de Integrales de Ito; es decir para $I_t = \int_0^t e^{ks} dB_s$ presenta media cero y varianza $\int_0^t e^{2ks} ds$, ver Teorema 4.4.9 en Shreve (2004) . Para la estimación de la covarianza entre χ_t y ξ_t ,

$$\text{Cov} [\chi_t, \xi_t] = \mathbb{E} [(\chi_t - \mathbb{E}(\chi_t)) (\xi_t - \mathbb{E}(\xi_t))]$$

recordemos que el modelo asume $dB_t^\chi dB_t^\xi = \rho_{\chi\xi} dt$ y de la ecuación 1.17, tenemos

$$\begin{aligned} \text{Cov} [\chi_t, \xi_t] &= \mathbb{E} \left[\left(\sigma_\chi \int_0^t e^{-\kappa(t-s)} dB_s^\chi \right) \left(\sigma_\xi \int_0^t dB_s^\xi \right) \right] \\ &= e^{-\kappa t} \sigma_\chi \sigma_\xi \mathbb{E} \left[\left(\int_0^t e^{\kappa s} dB_s^\chi \right) \left(\sigma_\xi \int_0^t dB_s^\xi \right) \right] \\ &= e^{-\kappa t} \sigma_\chi \sigma_\xi \rho_{\chi\xi} \int_0^t e^{\kappa s} ds \\ &= (1 - e^{-\kappa t}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa}. \end{aligned}$$

Por tanto, se puede concluir el siguiente Lema.

Lema 1.3.2. *Dado (χ_0, ξ_0) , el proceso $X_t = (\chi_t, \xi_t)^\top$ del modelo de Schwartz y Smith se distribuye conjuntamente como una Gaussiana bivariada con vector media $m(t)$ y matriz de varianzas-covarianzas $C(t)$: $X_t \sim \mathcal{N}(m(t), C(t))$, dados por*

$$m(t) = \mathbb{E} [\chi_t, \xi_t] = \left[e^{-\kappa t} \chi_0 \quad , \quad \xi_0 + \mu_\xi t \right] \quad (1.18a)$$

$$C(t) = \begin{bmatrix} (1 - e^{-2\kappa t}) \frac{\sigma_\chi^2}{2\kappa} & (1 - e^{-\kappa t}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa} \\ (1 - e^{-\kappa t}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa} & \sigma_\xi^2 t \end{bmatrix}. \quad (1.18b)$$

En consecuencia, tenemos el siguiente Lema.

Lema 1.3.3. *El precio spot sigue una distribución log-normal con esperanza*

$$\begin{aligned} \ln(\mathbb{E}[S_t]) &= e^{-\kappa t} \chi_0 + \xi_0 + \mu_\xi t + \frac{1}{2} \left((1 - e^{-2\kappa t}) \frac{\sigma_\chi^2}{2\kappa} \right. \\ &\quad \left. + \sigma_\xi^2 t + 2(1 - e^{-\kappa t}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa} \right). \end{aligned} \quad (1.19)$$

Demostración. Dado χ_0 y ξ_0 , gracias al Lema 1.3.2 se puede concluir que $\chi_t + \xi_t$ se distribuye como una Gaussiana con media y varianza determinadas por

$$\mathbb{E}[\ln(S_t)] = e^{-\kappa t} \chi_0 + \xi_0 + \mu_\xi t \quad (1.20a)$$

$$\text{Var}[\ln(S_t)] = (1 - e^{-2\kappa t}) \frac{\sigma_\chi^2}{2\kappa} + \sigma_\xi^2 t + 2(1 - e^{-\kappa t}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa}. \quad (1.20b)$$

Eso prueba que S_t tiene distribución log-normal. Para calcular la esperanza de S_t aplicamos la función generadora de momentos de la distribución Gaussiana,

$$\begin{aligned} \mathbb{E}[S_t] &= \mathbb{E}[\exp(\chi_t + \xi_t)] \\ &= \exp\left(\mathbb{E}[\chi_t + \xi_t] + \frac{1}{2} \text{Var}[\chi_t + \xi_t]\right), \end{aligned}$$

y usamos las ecuaciones 1.18a y 1.18b. □

Cuando $t \rightarrow \infty$, se observa de la ecuación 1.19 que los términos $e^{-\kappa t}$ y $e^{-2\kappa t}$ se aproximan a cero y

$$\lim_{t \rightarrow \infty} \ln \mathbb{E}[S_t] - \left(\mu_\xi + \frac{1}{2} \sigma_\xi^2 \right) t = \xi_0 + \frac{\sigma_\chi^2}{4\kappa} + \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa}.$$

Capítulo 2.

El modelo spot ajustado por riesgo neutral y valoración de Futuros

En el enfoque clásico para valorar un derivado se necesita cambiar la actual medida de probabilidad \mathbb{P} a una medida de probabilidad equivalente \mathbb{Q} llamado medida de riesgo neutral. La implicancia de este cambio es que se generan ambientes de neutralidad al riesgo en donde el precio de un derivado no depende de las preferencias al riesgo de los agentes. La tendencia del proceso que originalmente guía el proceso subyacente es ajustado, pero manteniendo las propiedades del movimiento Browniano para hacer más simple el proceso de valoración. Por eso decimos que nuestro modelo spot es ajustado por riesgo y el principal instrumental para realizar este ajuste en modelos Gaussianos es el Teorema de Girsanov. La definición formal de \mathbb{Q} es la siguiente:

Definición 2.0.1. (Medida equivalente). Sea el espacio probabilístico considerado en el capítulo anterior. Si Z es una variable aleatoria que verifica $E Z = 1$, entonces \mathbb{Q} definido por

$$\mathbb{Q}(A) = \int_A Z d\mathbb{P} \quad \text{para todo } A \in \mathcal{F},$$

es también una medida de probabilidad. Si además Z es estrictamente positivo \mathbb{P} -casi ciertamente, entonces las dos medidas son equivalentes.

La medida \mathbb{Q} impone restricciones para que el precio de un derivado sea libre arbitraje. Se dice que en un mercado existe arbitraje cuando se da la siguiente condición.

Definición 2.0.2. (Arbitraje). Sea \mathcal{Y}_t el valor de un portafolio. Diremos que en un modelo existe oportunidades de arbitraje cuando se verifica que $\mathcal{Y}_0 = 0$ y $\mathbb{P}[\mathcal{Y}_t \geq 0] = 1$ y para $T > 0$, $\mathbb{P}[\mathcal{Y}_T > 0] > 0$.

La formalización de la definición de arbitraje está dado por el “Primer Teorema Fundamental de Valoración de Activos”, ver Shreve (2004) en el Teorema 5.4.7, el cual señala que un modelo es libre de arbitraje si y solo si admite una medida \mathbb{Q} .

Todo los conceptos presentados hasta el momento nos será útil en la valoración del valor presente de los pagos de un derivado. En los cálculos de valoración elegiremos una medida

\mathbb{Q} que satisfaga ciertas condiciones -como veremos más adelante en el Teorema 2.3.1- y un proceso de descuento D_T dado por

$$D_T = e^{-\int_0^T r_t dt},$$

donde r_t es un proceso de tasa de interés \mathcal{F}_t -medible. Así, el precio en el tiempo t de un derivado que paga V_T en el tiempo T está dado por la esperanza condicional ajustada al riesgo; es decir

$$V_t = \frac{1}{D_t} E^{\mathbb{Q}} [D_T V_T | \mathcal{F}_t]. \quad (2.1)$$

Un aspecto crucial en la valoración de un derivado bajo los principios de arbitraje consiste en encontrar un portafolio que replique su valor a lo largo de cierto horizonte temporal. Esto es lo que se conoce como cobertura financiera; la definición formal es la siguiente:

Definición 2.0.3. Un derivado con fecha de vencimiento T es replicable si existe un portafolio \mathcal{Y}_t tal que $\mathcal{Y}_T = V_T$, donde V_T es un payoff recibido en T .

Todo derivado con payoff V_T aleatorio \mathcal{F}_t -medible es replicable, y el valor en el tiempo t de cualquier portafolio replicante \mathcal{Y}_t es

$$\mathcal{Y}_t = E^{\mathbb{Q}} \left[e^{-\int_t^T r_u du} V_T | \mathcal{F}_t \right].$$

Debemos notar que si un derivado es replicable, entonces el precio del derivado debe ser el mismo para todo \mathbb{Q} . Se asocia formalmente esta idea con el concepto de mercado completo.

Definición 2.0.4. (Mercado completo). Un modelo libre de arbitraje se denomina completo si todo reclamo contingente es replicable.

La formalización de la definición de mercado completo está dado por el “Segundo Teorema Fundamental de Valoración de Activos”, ver Shreve (2004) en el Teorema 5.4.9, el cual se presenta a continuación:

Teorema 2.0.5. *Sea un modelo que admite cierta medida \mathbb{Q} . El mercado es completo si y solamente si \mathbb{Q} es único.*

Basados en esta breve presentación de la teoría de riesgo neutral, el propósito de este capítulo es derivar el precio futuro del modelo spot estudiado en el capítulo 1. Nos referimos al precio futuro de cierto commodity a la valoración de un contrato que promete la entrega de un commodity en un tiempo de vencimiento prefijado. Supongamos que la fecha de entrega es T , con $0 \leq t \leq T < \infty$, y que el precio acordado a pagarse en la fecha de entrega es $F(t, T)$. En el tiempo T , efectivamente recibiremos un pago (posiblemente negativo)

$$S_T - F(t, T),$$

donde podemos interpretar este pago aleatorio en el tiempo T como un payoff de un derivado sobre el spot. Si suponemos que no es costoso entrar en tal contrato y bajo los supuestos de apropiada integrabilidad de los procesos S_t y $F(t, T)$, entonces obtendremos la relación donde podemos determinar el precio futuro

$$e^{-r(T-t)} E^{\mathbb{Q}} [S_T - F(t, T) | \mathcal{F}_t] = 0,$$

donde hemos asumido que $S_T \in L^1(\mathbb{Q})$, el espacio de variables aleatorias integrables con respecto a \mathbb{Q} y r es una tasa determinística. Es razonable asumir que para la formación del precio futuro consideremos la información disponible en el mercado, o en otras palabras, que $F(\cdot, T)$ sea \mathcal{F}_t -adaptado. Si S_t es un proceso semimartingala, entonces existe (al menos una) medida equivalente \mathbb{Q} tal que

$$F(t, T) = E^{\mathbb{Q}} [S_T | \mathcal{F}_t].$$

Este es la relación de valoración fundamental entre el precio spot y el precio futuro, el cual nos permite obtener la dinámica de un precio futuro libre de arbitraje. En un mercado completo, la medida \mathbb{Q} está únicamente determinada¹. En el caso de un mercado incompleto, la elección de \mathbb{Q} es abierto donde tendremos que determinar el más relevante para la valoración de un reclamo.

A continuación daremos un enunciado preciso de esta idea en el Teorema 2.2.3 y concluir así, en la ecuación 2.25 sobre la relación que hay entre los precios futuros y los factores del modelo de Schwartz y Smith. Esta es la ecuación que usaremos en el algoritmo de filtros de Kalman en el capítulo final. Pero previamente, daremos una breve presentación de la valoración en mercados incompletos porque el modelo de Schwartz y Smith es un tipo de mercado incompleto denominado modelo de factores.

2.1. Valoración en mercados incompletos

En general suele ser difícil mostrar que un modelo de mercado es completo debido en gran parte a la herramienta matemática que es usada para este fin (principalmente el Teorema de Girsanov y el Segundo Teorema Fundamental de Valoración de Activos). Una manera intuitiva de determinar que un mercado es completo es aplicando el Meta-Teorema 8.3.1 en Björk (2009).

¹Se invita al lector revisar Jarrow (1987) para una exposición detallada de la metodología de valoración de activos financieros bajo la teoría de arbitraje con el enfoque martingala para un mercado general con un commodity.

El Meta-Teorema nos permite rápidamente determinar que un mercado es completo y libre de arbitraje, comparando el número de fuentes de incertidumbre con respecto al número de activos negociados en el mercado. Supongamos que tenemos un número fijo de fuentes de incertidumbre R . El añadir más activos a nuestro portafolio nos daría las posibilidades de generar ganancias sin asumir riesgo alguno. Entonces, para tener un mercado libre de arbitraje, el número M de activos debería ser menor en comparación al número de fuentes de incertidumbre. Asimismo, cada nuevo activo nos daría la posibilidad de replicar el contrato futuro; por lo tanto, el concepto de mercados completo requiere que M sea mayor en comparación a R . Estas nociones se enuncia a continuación.

Teorema 2.1.1. (*Meta-Teorema de mercados completos*). Sea M el número de activos transados en el modelo sin considerar el activo libre de riesgo y R el número de fuentes de incertidumbre. Generalmente se cumplen las siguientes relaciones:

- (a) El modelo es libre de arbitraje si y solamente si $M \leq R$.
- (b) El modelo es completo si y solamente si $M \geq R$.
- (c) El modelo es completo y libre de arbitraje si y solamente si $M = R$.

Aplicando el Meta-Teorema, podemos decir que un mercado es incompleto cuando el número de fuentes de incertidumbre es mayor con respecto al número de activos negociados en el mercado. En lo que resta de esta sección nos centraremos en un tipo de mercado incompleto llamado “modelo de factores”, correspondiente a un mercado donde existen objetos que no son transados en el mercado. En Bjork (2009) se encuentra una generalización de este tipo del modelo, del cual el modelo Schwartz y Smith es un caso particular y a continuación presentamos los supuestos que subyacen en este tipo de mercado incompleto:

(A.1) **Como el mercado spot de commodity no es líquido, el inversionista no observa directamente los factores estocásticos X_t que determinan el precio spot S_t .**

- Para el proceso $X_t = (\chi_t, \xi_t)$ no se asume ser el proceso de precios de un activo que se transa en el mercado.
- La dinámica de X_t bajo la medida de probabilidad \mathbb{P} están dados por

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dB_t^1 \quad (2.2)$$

$$d\xi_t = \mu_\xi dt + \sigma_\xi \rho_{\chi\xi} dB_t^1 + \sigma_\xi \sqrt{1 - \rho_{\chi\xi}^2} dB_t^2, \quad (2.3)$$

donde B_t^1 y B_t^2 es el proceso Browniano estándar.

(A.2) **Incluimos en el mercado un activo libre de riesgo.**

El proceso de precios de un bono de corto plazo $\{\mathcal{B}_t : 0 \leq t \leq \bar{T}\}$ con dinámica dado por

$$d\mathcal{B}_t = r\mathcal{B}_t dt, \quad (2.4)$$

donde $\mathcal{B}_0 = 1$ y r se asume que es un proceso acotado y adaptado a \mathcal{F}_t , $r > 0$ sobre $t \in [0, \bar{T}]$. La ecuación 2.4 también puede escribirse de manera continua $\mathcal{B}_t = \exp[r t]$.

(A.3) **Existe un mercado líquido para cada reclamo contingente.**

Aplicando el Meta-Teorema, tenemos que el único activo que se negocia es el bono de corto plazo, es decir $M = 0$. Además, se tiene dos fuentes de incertidumbre $R = 2$ determinados por el proceso Browniano (B_t^1, B_t^2) . Por lo que el modelo de Schwartz y Smith es libre de arbitraje e incompleto. Como consecuencia de que el mercado es incompleto, no podrá obtenerse un precio futuro único bajo los principios de un mercado libre de arbitraje.

Consideraremos un reclamo contingente general escrito en términos del proceso X . Más específicamente, definimos el reclamo \mathcal{Y} con fecha de vencimiento en T por

$$\mathcal{Y} = \Phi(X(T)),$$

donde Φ es una función determinística, y nuestro principal problema es estudiar el proceso de precios $\Pi(t; \mathcal{Y})$ sobre este reclamo y como veremos mas adelante encontraremos una función real diferenciable $F(\cdot)$ tal que $\Pi(t; \mathcal{Y}) = F(t, X_t)$.

Para dar contexto a nuestro problema de estudio, supongamos que queremos comprar cierta cantidad de cobre en el tiempo T y negociamos un contrato cuyo precio depende de X . Supongamos que las unidades monetarias están en dólares, la función del contrato Φ puede tener la siguiente forma

$$\Phi(x) = \begin{cases} 100, & \text{si } \sigma_x > 20\% , \\ 95, & \text{si } \sigma_x \leq 20\% . \end{cases}$$

Una posible interpretación de esta función es que si la volatilidad de X en el tiempo T supera el 20% pagaríamos 100 dólares, en caso contrario 95 dólares. El problema es encontrar un precio razonable para el contrato del ejemplo y nos referimos a “razonable” en el sentido de que no habrá posibilidades de arbitraje cuando entramos a negociar dicho contrato.

A diferencia de un modelo con mercado completo, debemos señalar de manera enfática de que el precio del reclamo \mathcal{Y} no será únicamente determinada por la dinámica bajo la medida de probabilidad \mathbb{P} del proceso subyacente X sino que también de la elección de la medida de probabilidad de riesgo neutral \mathbb{Q} , esto se sustenta en las siguientes razones.

- (a) De la aplicación del Meta-Teorema hemos visto de que nuestro mercado es incompleto. La afirmación de que el mercado no es completo resulta de notar de que en nuestro mercado descrito a priori no existen maneras de determinar portafolios autofinanciables interesantes. Recordemos que el único activo de nuestro mercado es el bono de corto plazo \mathcal{B} y solamente tendríamos la opción de invertir todo nuestro dinero en los bonos y no tendríamos la posibilidad de replicar cualquier derivado de la forma $\Phi(X_T)$. Por lo tanto, concluimos que al no poder replicar nuestro reclamo contingente no podemos esperar que podamos obtener un único proceso de precios libre de arbitraje.

- (b) La estrategia estándar para obtener un precio único para el reclamo contingente sobre X es seguir el esquema para el modelo de Black-Scholes. Asumiríamos que el proceso de precios $\Pi(t; \mathcal{Y})$ es de la forma $\Pi(t; \mathcal{Y}) = F(t, X_t)$ y luego formaríamos un portafolio basado sobre el derivado F y el subyacente X . Eligiríamos las ponderaciones de participación de cada activo en el portafolio de tal manera de que no exista incertidumbre y tengamos un rendimiento de un activo libre de riesgo es decir que iguale a la tasa de interés spot r , y esta última igualdad tendría la forma de una ecuación derivada parcial para la función de valoración F . Este enfoque no tiene sentido de aplicarlo en nuestro caso porque el proceso X de nuestro modelo no es el precio de un activo negociado al ser factores no observables y por lo tanto carece de sentido de hablar de portafolio sobre X .

A pesar de que no obtendremos un precio único para nuestro derivado particular, esto no implica que su precio pueda tomar cualquier forma. Del Meta-teorema hemos visto de que la razón de tener mercado incompleto es porque no tenemos suficiente número de activos subyacentes. Si añadimos dos activos más a nuestro mercado sin agregar un proceso Browniano adicional, nuestro mercado sería completo.

El requerimiento de que nuestro mercado de derivados sea libre de arbitraje implica que los precios de los diferentes derivados (es decir, reclamos con diferentes funciones de contrato o fechas de vencimiento) tendrán que verificar ciertas relaciones de consistencia interna para evitar posibilidades de arbitraje en el mercado de bonos. En el contexto de nuestro ejemplo esto significa que si aún es imposible obtener un precio único para nuestro contrato futuro, digamos “volatilidad de precios 20 %”

$$\Phi(x) = \begin{cases} 100, & \text{si } \sigma_x > 20\% , \\ 95, & \text{si } \sigma_x \leq 20\% , \end{cases}$$

con fecha de vencimiento T , deberá existir el requerimiento de consistencia interna entre el precio de este contrato y el precio del siguiente contrato digamos

$$\Gamma(x) = \begin{cases} 100, & \text{si } \sigma_x > 25\% , \\ 95, & \text{si } \sigma_x \leq 25\% . \end{cases}$$

con fecha de vencimiento T . Si tomamos como referencia el precio de mercado de un derivado en particular, entonces los precios de los demás derivados deberán ser determinados por este precio. Este hecho es consistente con el Meta-Teorema en el sentido de que después de añadir dos contratos tendríamos la relación $R = M = 2$, por lo tanto nuestro nuevo modelo sería completo. Si por ejemplo, elegimos el precio del contrato “volatilidad de precios 20 %” como referencia , entonces esperaríamos de que el contrato de “volatilidad de precios 25 %” sea determinado de manera única en términos del precio de referencia.

Para formalizar esta idea supongamos que queremos tomar una posición de compra sobre un contrato \mathcal{O} a T años. Se asume que existen en el mercado tres contratos futuros \mathcal{Y}, \mathcal{Z} y \mathcal{O} de la forma

$$\begin{aligned}
\mathcal{Y} &= \Phi(X_T), \\
\mathcal{Z} &= \Gamma(X_T), \\
\mathcal{O} &= \Lambda(X_T),
\end{aligned}$$

donde Φ, Γ y Λ son funciones determinísticas de valor real. Nuestro interés es encontrar cómo los precios de estos derivados se relacionan para evitar posibilidades de arbitraje en el mercado de derivados. Para ello asumiremos que nuestro mercado es libre de costos de transacción y al igual que en el caso de mercados completos supondremos cierta estructura en los procesos de los precios.

(A.4) **Asumiremos que existe un mercado líquido y sin costos de transacción para cada reclamo contingente \mathcal{Y}, \mathcal{Z} y \mathcal{O} .**

Los precios de mercado de los reclamos son de la siguiente forma

$$\begin{aligned}
\Pi(t, \mathcal{Y}) &= H(t, X_t), \\
\Pi(t, \mathcal{Z}) &= G(t, X_t), \\
\Pi(t, \mathcal{O}) &= \Upsilon(t, X_t),
\end{aligned}$$

donde H, G y Υ son funciones reales diferenciables.

La dinámica de los proceso de precios $H(t, X_t), G(t, X_t)$ y $\Upsilon(t, X_t)$ se obtienen aplicando la fórmula de Itô para el modelo asumido de X_t . Es decir,

$$dH = \alpha_H H dt + \sigma_{1H} H dB_t^1 + \sigma_{2H} H dB_t^2,$$

$$dG = \alpha_G G dt + \sigma_{1G} G dB_t^1 + \sigma_{2G} G dB_t^2,$$

$$d\Upsilon = \alpha_\Upsilon \Upsilon dt + \sigma_{1\Upsilon} \Upsilon dB_t^1 + \sigma_{2\Upsilon} \Upsilon dB_t^2,$$

donde

$$\alpha_H = \frac{\frac{\partial H}{\partial t} - \kappa \chi_t \frac{\partial H}{\partial \chi} + \mu_\xi \frac{\partial H}{\partial \xi} + \frac{1}{2} \left[\sigma_\chi^2 \frac{\partial^2 H}{\partial \chi^2} + \sigma_\xi^2 \frac{\partial^2 H}{\partial \xi^2} \right] + (\sigma_\xi \rho \sigma_\chi) \frac{\partial^2 H}{\partial \chi \partial \xi}}{H},$$

$$\sigma_{1H} = \frac{\sigma_\chi \frac{\partial H}{\partial \chi} + \sigma_\xi \rho \frac{\partial H}{\partial \xi}}{H},$$

$$\sigma_{2H} = \frac{\sigma_\xi \sqrt{1 - \rho^2} \frac{\partial H}{\partial \xi}}{H}. \quad (2.5)$$

De manera similar se obtienen $\alpha_G, \alpha_\Upsilon, \sigma_{1G}, \sigma_{2G}, \sigma_{1\Upsilon}, \sigma_{2\Upsilon}$. Para mayor detalle de los cálculos, ver el siguiente Lema.

Lema 2.1.2. *Sea X_t dado por las relaciones 2.2 y 2.3. Se sigue que la dinámica de H está dada por*

$$\begin{aligned} dH = & \left\{ \frac{\partial H}{\partial t} - \kappa \chi_t \frac{\partial H}{\partial \chi} + \mu_\xi \frac{\partial H}{\partial \xi} + \frac{1}{2} \left[\sigma_\chi^2 \frac{\partial^2 H}{\partial \chi^2} + \sigma_\xi^2 \frac{\partial^2 H}{\partial \xi^2} \right] + (\sigma_\xi \rho \sigma_\chi) \frac{\partial^2 H}{\partial \chi \partial \xi} \right\} dt \\ & + \left(\sigma_\chi \frac{\partial H}{\partial \chi} + \sigma_\xi \rho \frac{\partial H}{\partial \xi} \right) dB_t^1 + \left(\sigma_\xi \sqrt{1 - \rho^2} \frac{\partial H}{\partial \xi} \right) dB_t^2 \end{aligned}$$

Demostración. Este es resultado directo de aplicar la fórmula de Ito para la función $F(\cdot) = H(t, \chi_t, \xi_t)$

$$\begin{aligned} dH(\chi_t, \xi_t) = & \left\{ \frac{\partial H}{\partial t} + b_1 \frac{\partial H}{\partial \chi} + b_2 \frac{\partial H}{\partial \xi} + \frac{1}{2} \left[(\sigma_{11}^2 + \sigma_{12}^2) \frac{\partial^2 H}{\partial \chi^2} + (\sigma_{21}^2 + \sigma_{22}^2) \frac{\partial^2 H}{\partial \xi^2} \right] \right\} dt \\ & + (\sigma_{21} \sigma_{11} + \sigma_{22} \sigma_{12}) \frac{\partial^2 H}{\partial \chi \partial \xi} dt + \left(\sigma_{11} \frac{\partial H}{\partial \chi} + \sigma_{21} \frac{\partial H}{\partial \xi} \right) dB_t^1 + \left(\sigma_{12} \frac{\partial H}{\partial \chi} + \sigma_{22} \frac{\partial H}{\partial \xi} \right) dB_t^2 \end{aligned}$$

donde de 2.2 y 2.3 se definen $b_1 = -\kappa \chi_t, b_2 = \mu_\xi, \sigma_{11} = \sigma_\chi, \sigma_{12} = 0, \sigma_{21} = \sigma_\xi \rho \sigma_\chi$ y $\sigma_{22} = \sigma_\xi \sqrt{1 - \rho^2}$.

□

Conocida la dinámica de nuestros tres contratos, formamos un portafolio autofinanciable² V conformado por H, G y Υ con participaciones θ_H, θ_G y θ_Υ , respectivamente. De acuerdo al Lema 6.4 en Bjork (2009), la tasa de retorno de este portafolio está dado por la siguiente relación

$$dV = V \left\{ \theta_H \cdot \frac{dH}{H} + \theta_G \cdot \frac{dG}{G} + \theta_\Upsilon \cdot \frac{d\Upsilon}{\Upsilon} \right\},$$

y reemplazando las expresiones obtenidas del Lema 2.1.2 obtenemos

²Se dice que un portafolio es autofinanciable si el cambio en el valor de dicho portafolio depende solo de los cambios en los precios de sus componentes. Es decir, que no entran ni salen flujos netos del portafolio.

$$dV = V \{ \theta_H \cdot \alpha_H + \theta_G \cdot \alpha_G + \theta_\Upsilon \cdot \alpha_\Upsilon \} dt + V \{ \theta_H \cdot \sigma_{1H} + \theta_G \cdot \sigma_{1G} + \theta_\Upsilon \cdot \sigma_{1\Upsilon} \} dB_t^1 + \\ V \{ \theta_H \cdot \sigma_{2H} + \theta_G \cdot \sigma_{2G} + \theta_\Upsilon \cdot \sigma_{2\Upsilon} \} dB_t^2.$$

Para que este portafolio sea libre de riesgo, eligiremos θ_H , θ_G y θ_Υ tal que los coeficientes de dB_t^1 y dB_t^2 sean iguales a cero, es decir

$$\begin{aligned} \theta_H + \theta_G + \theta_\Upsilon &= 1, \\ \theta_H \cdot \sigma_{1H} + \theta_G \cdot \sigma_{1G} + \theta_\Upsilon \cdot \sigma_{1\Upsilon} &= 0, \\ \theta_H \cdot \sigma_{2H} + \theta_G \cdot \sigma_{2G} + \theta_\Upsilon \cdot \sigma_{2\Upsilon} &= 0. \end{aligned} \tag{2.6}$$

Reemplazando la solución del sistema dada por 2.6, obtenemos la dinámica de la ecuación de portafolios libre de riesgo

$$dV = V \{ \theta_H \cdot \alpha_H + \theta_G \cdot \alpha_G + \theta_\Upsilon \cdot \alpha_\Upsilon \} dt$$

el cual está dado por

$$dV = V \left\{ \frac{\sigma_{2\Upsilon}}{[\sigma_{2\Upsilon} - \sigma_{2H}]} + \frac{\sigma_{2\Upsilon} + \sigma_{1\Upsilon} [\sigma_{2H} - \sigma_{2\Upsilon}]}{[\sigma_{1G} - \sigma_{1\Upsilon}] [\sigma_{2H} - \sigma_{2\Upsilon}] - [\sigma_{2G} - \sigma_{2\Upsilon}] [\sigma_{1H} - \sigma_{1\Upsilon}]} \right\} \times \\ \left[\frac{\sigma_{2G} - \sigma_{2\Upsilon}}{[\sigma_{2\Upsilon} - \sigma_{2H}]} \right] \cdot \alpha_H + \frac{\sigma_{2\Upsilon} + \sigma_{1\Upsilon} [\sigma_{2H} - \sigma_{2\Upsilon}]}{[\sigma_{1G} - \sigma_{1\Upsilon}] [\sigma_{2H} - \sigma_{2\Upsilon}] - [\sigma_{2G} - \sigma_{2\Upsilon}] [\sigma_{1H} - \sigma_{1\Upsilon}]} \cdot \alpha_G + \\ \frac{[\sigma_{2\Upsilon} - \sigma_{2H}] - \sigma_{2\Upsilon} - \theta_G \cdot [\sigma_{2G} - \sigma_{2\Upsilon} + \sigma_{2\Upsilon} - \sigma_{2H}]}{[\sigma_{2\Upsilon} - \sigma_{2H}]} \cdot \alpha_\Upsilon.$$

De esta manera, hemos construido un portafolio libre de riesgo. Luego por la Proposición 7.6 en (Bjork(2009), pag. 97), para evitar la posibilidad de arbitraje es necesario que la tasa de retorno sobre este portafolio sea igual a la tasa instantánea (o de corto plazo) libre de riesgo: $\theta_H \cdot \alpha_H + \theta_G \cdot \alpha_G + \theta_\Upsilon \cdot \alpha_\Upsilon = r$ para todo t , es decir

$$\theta_H (\alpha_H - r) + \theta_G (\alpha_G - r) + \theta_\Upsilon (\alpha_\Upsilon - r) = 0.$$

De la Proposición 15.1 en (Bjork(2009), pag. 214), se sigue que en un mercado de derivados libre de arbitraje siempre existe un proceso universal $\lambda(t)$, el cual se enuncia a continuación.

Proposición 2.1.3. *Sea la dinámica del contrato H dada por el Lema 2.1.2. Si el mercado de derivados es libre de arbitraje, entonces existe un proceso universal $\lambda(t)$ tal que, con probabilidad uno, y para todo t , se cumple*

$$\alpha_H(t) - r = \sigma_{1H}\lambda_1 + \sigma_{2H}\lambda_2 \quad (2.7)$$

sin considerar la elección específica del derivado H .

Notemos que la existencia de λ de la Proposición 2.1.3 se obtiene de notar de que tenemos un sistema de ecuaciones lineales homogéneas,

$$\begin{bmatrix} \alpha_H - r & \sigma_{1H} & \sigma_{2H} \\ \alpha_G - r & \sigma_{1G} & \sigma_{2G} \\ \alpha_\Upsilon - r & \sigma_{1\Upsilon} & \sigma_{2\Upsilon} \end{bmatrix}^\top \begin{bmatrix} \theta_H \\ \theta_G \\ \theta_\Upsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Si definimos a la matriz J por

$$J := \begin{bmatrix} \alpha_H - r & \sigma_{1H} & \sigma_{2H} \\ \alpha_G - r & \sigma_{1G} & \sigma_{2G} \\ \alpha_\Upsilon - r & \sigma_{1\Upsilon} & \sigma_{2\Upsilon} \end{bmatrix},$$

entonces el supuesto de ausencia de arbitraje implicaría que J debe ser singular y asumiendo que la matriz σ de orden 2×2

$$\sigma := \begin{bmatrix} \sigma_{1H} & \sigma_{2H} \\ \sigma_{1G} & \sigma_{2G} \end{bmatrix}$$

es invertible (con probabilidad uno para cada t), deducimos que la primera columna de J puede escribirse como una combinación lineal de las otras columnas. Por lo tanto, se garantiza la existencia del proceso multiplicador $\lambda_1(t, X_t)$, $\lambda_2(t, X_t)$ tal que

$$\alpha_H - r = \sigma_{1H}\lambda_1 + \sigma_{2H}\lambda_2,$$

$$\alpha_G - r = \sigma_{1G}\lambda_1 + \sigma_{2G}\lambda_2,$$

y

$$\alpha_{\Gamma} - r = \sigma_{1\Gamma}\lambda_1 + \sigma_{2\Gamma}\lambda_2.$$

Cabe mencionar que la interpretación económica de $\lambda_1(t, X_t)$, $\lambda_2(t, X_t)$ es la de “precio de mercado del riesgo”. En forma vectorial,

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

es solución del sistema

$$\alpha_H - r = \sigma_{1H}\lambda_1 + \sigma_{2H}\lambda_2,$$

$$\alpha_G - r = \sigma_{1G}\lambda_1 + \sigma_{2G}\lambda_2,$$

y del supuesto de que σ es no singular vemos que λ está únicamente determinada

$$\lambda = \begin{bmatrix} \sigma_{1H} & \sigma_{2H} \\ \sigma_{1G} & \sigma_{2G} \end{bmatrix}^{-1} \begin{bmatrix} \alpha_H - r \\ \alpha_G - r \end{bmatrix}.$$

Observación 2.1.4. En contraste con el modelo de mercado completo, en nuestro caso la matriz de volatilidad σ y el vector de retornos no son dados a priori, por lo tanto λ no está determinado dentro del modelo.

Finalmente, de la Proposición 2.1.3 y del Lema 2.1.2 se obtiene la siguiente ecuación de valoración de nuestro contrato.

Proposición 2.1.5. (*Ecuación de Valoración*). Si el mercado es libre arbitraje, entonces el proceso de precios libre de arbitraje $\Pi(t, \Phi)$ en nuestro modelo está dado por $\Pi(t, \Phi) = H(t, X_t)$, tal que H resuelve el problema de acotamiento

$$\begin{aligned} \frac{\partial H}{\partial t} + (-\kappa\chi_t - \sigma_\chi\lambda_1) \frac{\partial H}{\partial \chi} + (\mu_\xi - \sigma_\xi\rho_{\chi\xi}\lambda_1 - \sigma_\xi\sqrt{1 - \rho_{\chi\xi}^2}\lambda_2) \frac{\partial H}{\partial \xi} + \\ \frac{1}{2} \left[\sigma_\chi^2 \frac{\partial^2 H}{\partial \chi^2} + \sigma_\xi^2 \frac{\partial^2 H}{\partial \xi^2} + 2(\sigma_\xi\rho_{\chi\xi}\sigma_\chi) \frac{\partial^2 H}{\partial \chi \partial \xi} \right] - Hr = 0, \end{aligned}$$

$$H(T, x) = \Phi(x),$$

donde λ_1, λ_2 son universales en el sentido de que no dependen de la elección específica del derivado, H .

Como resultado del Teorema de Feynman-Kac³, ver Teorema 6.4.1 en Shreve (2004) y del Teorema de Girsanov, podemos reescribir la Proposición 2.1.5 en términos de la medida de riesgo neutral, \mathbb{Q} .

Proposición 2.1.6. (*Valoración de Riesgo Neutral*). Si el mercado es libre de arbitraje, entonces existe una medida martingala equivalente \mathbb{Q} , tal que la función de valoración H en la Proposición 2.1.5 puede representarse como

$$H(t, x) = e^{-r(T-t)} E_{t,x}^{\mathbb{Q}} [\Phi(X(T))],$$

donde la dinámica bajo la medida \mathbb{Q} del proceso X están dados

$$\begin{aligned} d\chi_t &= \{-\kappa\chi_t - \sigma_\chi\lambda_1\} dt + \sigma_\chi d\tilde{B}_t^1 \\ d\xi_t &= \{\mu_\xi - \sigma_\xi\rho_{\chi\xi}\lambda_1 - \sigma_\xi\sqrt{1 - \rho_{\chi\xi}^2}\lambda_2\} dt + \sigma_\xi\rho_{\chi\xi}d\tilde{B}_t^1 + \sigma_\xi\sqrt{1 - \rho_{\chi\xi}^2}d\tilde{B}_t^2 \end{aligned}$$

donde λ_1 y λ_2 son universales y el vector $(\tilde{B}_t^1, \tilde{B}_t^2)$ es un \mathbb{Q} -proceso Browniano estándar.

Notemos que existe una correspondencia de uno a uno entre la medida \mathbb{Q} y el precio de mercado de riesgo λ . Por lo tanto, elegir λ es equivalente a elegir \mathbb{Q} .

2.2. Contrato Futuro

Para definir el precio de un contrato futuro, consideraremos el mercado financiero de la sección anterior, supuestos (A.1)-(A.2), donde el proceso subyacente está dado por el factor estocástico X_t . Se asume que el mercado es libre de arbitraje; es decir, existe al menos una medida martingala de riesgo neutral \mathbb{Q} .

En este contexto, consideraremos también un reclamo \mathcal{Y} con fecha de vencimiento en T de la forma $\mathcal{Y} = \Phi(X_T)$, y asumiremos que estamos en el tiempo t . Si compramos \mathcal{Y} y pagamos hoy día, entonces sabemos que el precio libre de arbitraje está dado por

$$\Pi(t, \mathcal{Y}) = G(t, X_t, T, \mathcal{Y}),$$

donde de la Proposición 2.1.6, la función de valoración G está dado por

$$G(t, X_t, T, \mathcal{Y}) = e^{-r(T-t)} E_{t,x}^{\mathbb{Q}} [\mathcal{Y}]$$

y el flujo de pagos son como sigue:

³La fórmula de Feynman-Kac establece un vínculo entre las ecuaciones diferenciales parciales parabólicas y los procesos estocásticos.

- (a) En el tiempo t pagamos $\Pi(t, \mathcal{Y})$ al vendedor del contrato.
- (b) En el tiempo T recibimos \mathcal{Y} del vendedor del contrato.

Veremos a continuación de que el contrato futuro es una variante de este contrato general, pero previamente describiremos brevemente la formación de precios forward para comprender cómo funciona el mercado de futuros. Un contrato Forward es un acuerdo para comprar o vender un activo en un tiempo futuro a un precio determinado. El siguiente ejemplo nos permite ilustrar sus características más relevantes⁴.

Un comprador de oro busca cubrir su exposición al alza en el “London Bullion Market Association – LBMA” (Bolsa de Metales Preciosos de Londres), y pacta en $t = 0$ un Forward de oro en \$1,250.00 por onza. Si al momento del vencimiento del contrato, digamos $T = 30$ días, el nivel de mercado del oro cierra en \$1,300.00 por onza en el LBMA, entonces se le compensaría financieramente en \$50.00 por onza (i.e. \$1,300.00 – \$1,250.00). Si en cambio, el precio del oro en el LBMA hubiese caído y cerrado en \$1,220.00, entonces el comprador del Forward de oro hubiese tenido que compensar financieramente a la contraparte de esta operación en \$30 (\$1,220.00 – \$1,250.00).

Los contratos Futuros, a diferencia de los Forwards, se negocian en un Mercado Centralizado de Derivados y son ‘Marcados al Mercado’ de manera diaria lo cual genera una compensación financiera que tiene que cubrirse al día siguiente; esto se conoce como “Llamada de Mantenimiento” (i.e. Maintenance Call) o “Llamada de Margen” (i.e. Margin Call). Formalmente, la definición es la siguiente:

Definición 2.2.1. (Contrato Futuro). Sea \mathcal{Y} un reclamo contingente con fecha de vencimiento en T . Un contrato futuro sobre \mathcal{Y} con fecha de entrega en T , es un activo financiero con las siguientes propiedades:

- (a) Para cada instante $t \in [0, T]$, existe en el mercado cotizaciones continuas de $F(t, T, \mathcal{Y})$ conocido como precios futuros para \mathcal{Y} en t , con fecha de entrega en T .
- (b) En el tiempo T , el que toma la posición de compra del contrato paga $F(T, T, \mathcal{Y})$ y recibe el reclamo \mathcal{Y} .
- (c) Durante un intervalo de tiempo arbitrario $(t, t + \Delta t]$, el que toma la posición de compra del contrato recibe (o paga, si es negativo) el importe de $F(t, T, \mathcal{Y}) - F(t + \Delta t, T, \mathcal{Y})$.
- (d) Para cada $t \in [0, T)$, el precio spot de obtener el contrato futuro $\Pi(t)$ es por definición igual a cero.

El cronograma de pagos descrito en la Definición 2.2.1 es conocido como ‘making to market’ o en español ‘ajustados al valor del mercado.’⁵ Para entender, consideremos un intervalo de tiempo $[0, T]$ el cual lo dividimos en subintervalos usando las particiones $0 = t_0 < t_1 < t_2 < \dots < t_n = T$. Nos referimos a cada subintervalo $[t_k, t_{k+1})$ como un “día”, como se muestra en el cuadro 2.1:

⁴Ejemplo tomado de <https://gerens.pe/blog/enfrentar-precios-commodities-instrumentos-financieros-derivados/> con fecha 17/06/2018.

⁵Está organizado de tal manera que el titular del contrato futuro (tanto en posición corta o larga) debe mantener una cierta cantidad de dinero con el broker como un margen de seguridad en el caso de que se incumpla con el contrato.

Cuadro 2.1.: Precios Futuros del Cobre

Fecha	Spot	3 meses	15 meses	27 meses	63 meses
31/10/2002	1554	1568	1619	1661	1724
1/11/2002	1570	1583	1635	1672	1733
4/11/2002	1579	1594	1648	1686	1748
5/11/2002	1570	1586	1641	1679	1741
...		
11/10/2017	6756	6800	6878	6908	6907
12/10/2017	6861	6887	6962	6988	6987
13/10/2017	6859	6882	6956	6983	6982

Si suponemos que la tasa de interés es constante dentro de cada día, entonces el proceso de descuento al cual lo denotaremos por D está dado por $D(0) = 1$ y para $k = 0, 1, \dots, n - 1$,

$$D(t_{k+1}) = \exp \left\{ - \int_0^{t_{k+1}} r_u du \right\} = \exp \left\{ - \sum_{j=0}^k r_{t_j} (t_{j+1} - t_j) \right\},$$

el cual es $\mathcal{F}(t_k)$ -medible. De acuerdo a la fórmula de valoración bajo riesgo neutral, ver Ecuación 5.2.30 en Shreve (2004), el bono de corto plazo que paga 1 dólar en la fecha de vencimiento T su precio en el tiempo t_k es

$$\mathcal{B}_{t_k} = \frac{1}{D(t_k)} E^Q [D(T) | \mathcal{F}_{t_k}].$$

Si mantenemos posición de compra entre t_k y t_{k+1} , entonces en el tiempo t_k recibiremos un pago de

$$F(t_{k+1}, T) - F(t_k, T).$$

El proceso estocástico $F(t, T)$ es construido de tal manera de que $F(t, T)$ es \mathcal{F}_t -medible para cada t y

$$F(T, T, \mathcal{Y}) = \mathcal{Y}. \tag{2.8}$$

Entonces, si compramos un contrato futuro en el tiempo cero y si mantenemos la posición de compra hasta la fecha de vencimiento T , la suma de pagos recibidos es

$$\sum_{k=1}^n (F(t_k, T) - F(t_{k-1}, T)) = F(T, T) - F(t_0, T) = \mathcal{Y} - F(t_0, T),$$

tal como se muestra en el siguiente cuadro.

Cuadro 2.2.: Flujo de Pagos de una Compra de Futuros

Día t	Precios futuros	Pagos realizados en el Día t
0	$F(t_0, T)$	
1	$F(t_1, T)$	$(F(t_1, T) - F(t_0, T))$
2	$F(t_2, T)$	$(F(t_2, T) - F(t_1, T))$
...		...
$n-1$	$F(t_{n-1}, T)$	$(F(t_{n-1}, T) - F(t_{n-2}, T))$
$n=T$	$F(T, T)$	$(F(T, T) - F(t_{n-1}, T))$
	Sumatoria de pagos	$F(T, T) - F(t_0, T)$

Además de verificar la 2.8, el precio futuro es elegido de tal manera de que para cada t_k el valor del pago a ser recibido en el tiempo t_{k+1} , y para todo $t_j > t_k$, es cero. Esto significa que en cualquier momento podemos entrar o cerrar nuestra posición en el contrato sin incurrir en costos adicionales a los pagos incurridos. La condición de que el valor en el tiempo t_k del pago a recibirse en el tiempo t_{k+1} es cero puede escribirse como

$$\begin{aligned}
 0 &= \frac{1}{D(t_k)} E^{\mathbb{Q}} [D(t_{k+1}) (F(t_{k+1}, T) - F(t_k, T)) | \mathcal{F}_{t_k}] \\
 &= \frac{D(t_{k+1})}{D(t_k)} \left\{ E^{\mathbb{Q}} [F(t_{k+1}, T) | \mathcal{F}_{t_k}] - F(t_k, T) \right\}, \tag{2.9}
 \end{aligned}$$

donde se ha usado el hecho de que $D(t_{k+1})$ es \mathcal{F}_{t_k} medible. De 2.9 vemos que

$$E^{\mathbb{Q}} [F(t_{k+1}, T) | \mathcal{F}_{t_k}] = F(t_k, T), \quad k = 0, 1, \dots, n-1. \tag{2.10}$$

Esto muestra que $F(t_k, T)$ debe ser una martingala discreta bajo la medida \mathbb{Q} . Pero también necesitamos que se cumpla que $F(T, T) = \mathcal{Y}$, a partir del cual concluimos de que los precios futuros deben ser dados por la siguiente fórmula

$$F(t_k, T) = E^{\mathbb{Q}} [\mathcal{Y} | \mathcal{F}_{t_k}], \quad k = 0, 1, \dots, n. \tag{2.11}$$

En efecto, bajo la condición de que $F(T, T) = \mathcal{Y}$, vemos que 2.10 y 2.11 son equivalentes.

Finalmente, de la expresión de $F(t, T)$ dada en 2.11 debemos notar que el valor en el tiempo t_k del pago recibido en el tiempo t_j es cero para cada $j \geq k+1$. Esto es resultado de que $D(t_j)$ es $\mathcal{F}_{t_{j-1}}$ medible y de la propiedad de martingala para $F(t, T)$, se muestra que

$$\frac{1}{D(t_k)} E^{\mathbb{Q}} [D(t_j) (F(t_j, T) - F(t_{j-1}, T)) | \mathcal{F}_{t_k}]$$

se puede expresar como

$$\begin{aligned}
&= \frac{1}{D(t_k)} E^{\mathbb{Q}} \left[E^{\mathbb{Q}} \left[D(t_j) (F(t_j, T) - F(t_{j-1}, T)) \mid \mathcal{F}_{t_{j-1}} \right] \mid \mathcal{F}_{t_k} \right] \\
&= \frac{1}{D(t_k)} E^{\mathbb{Q}} \left[D(t_j) E^{\mathbb{Q}} \left[F(t_j, T) \mid \mathcal{F}_{t_{j-1}} \right] - D(t_j) F(t_{j-1}, T) \mid \mathcal{F}_{t_k} \right] \\
&= \frac{1}{D(t_k)} E^{\mathbb{Q}} \left[D(t_j) F(t_{j-1}, T) - D(t_j) F(t_{j-1}, T) \mid \mathcal{F}_{t_k} \right] = 0.
\end{aligned}$$

El siguiente ejemplo ilustra esta idea. Un inversionista compra contratos futuros por un valor de \$150.00. Si al final del día el precio futuro es \$156.00, el inversionista obtiene una ganancia de \$6.00; luego el contrato se rescribiría a un precio futuro de \$156.00 y el valor del contrato resultaría en cero. Si al final del segundo día el precio futuro es \$150.00, el inversionista pierde \$6.00 y el contrato se rescribiría a un precio de \$150.00. El valor acumulado que obtiene el tenedor del contrato futuro durante los dos días sería $\$(6.00 \times e^r - 6.00)$ donde r denota la tasa de interés⁶. En contraste, si los precios futuros caen en \$6.00 en el primer día, y luego retornan a su nivel inicial en el segundo día, el valor acumulado de los dos días sobre la posición de compra de futuros sería $\$(-6.00e^r + 6.00)$.

Estas consideraciones nos permiten realizar la siguiente definición para el caso continuo (es decir, cuando se asume que r_t es un proceso estocástico adaptado, no necesariamente constante en intervalos de tiempo de la forma $[t_k, t_{k+1})$), para mayor detalle ver Shreve (2004, pag. 244).

Definición 2.2.2. (Precio Futuro). Sea \mathcal{Y} un reclamo contingente con fecha de vencimiento en T y asumamos de que el precio de mercado se obtiene de cierta medida martingala de riesgo neutral \mathbb{Q} tal que $E^{\mathbb{Q}}[\mathcal{Y}] < \infty$. El proceso de precios futuros de un reclamo contingente \mathcal{Y} cuyo valor en el tiempo T es \mathcal{Y}_T , está definido por

$$F(t, T, \mathcal{Y}) = E_{t, X_t}^{\mathbb{Q}}(\mathcal{Y} \mid \mathcal{F}_t), \quad \text{para cada } 0 \leq t \leq T. \quad (2.12)$$

Una posición de compra sobre contrato de futuro es un acuerdo de recibir como flujo de caja las variaciones de los precios futuros (el cual puede ser positivo o negativo) durante el tiempo en que mantenemos nuestra posición. Este flujo de caja puede verse como dividendos pagados continuamente en el tiempo. En el siguiente Teorema se caracteriza las propiedades del precio futuro.

Teorema 2.2.3. *El proceso de precios futuros es una martingala bajo la medida neutral al riesgo \mathbb{Q} , cumple $F(T, T, \mathcal{Y}) = \mathcal{Y}$, y el valor de una posición sobre un contrato futuro sobre un intervalo de tiempo es siempre cero.*

Demostración. El argumento usual de iteración de la esperanza condicionada muestra que $F(t, T, \mathcal{Y})$ dada por 2.12 es una \mathbb{Q} -martingala que cumple la condición terminal $F(T, T, \mathcal{Y}) = \mathcal{Y}$. A decir verdad, esta es la única \mathbb{Q} -martingala que verifica esta condición terminal.

⁶Si r fuese cero, el valor de mantener una posición de futuro para t días sería simplemente el cambio de precios sobre ese periodo, tal como se indica en la propiedad (c)

Si la filtración \mathcal{F}_t , $0 \leq t \leq T$, es generada por un proceso Browniano B_t , $0 \leq t \leq T$, entonces por el Corolario 5.3.2 del Teorema de Representación Martingala en Shreve (2004) podemos asumir de que $F(t) = F(t, T, \mathcal{Y})$ es un proceso de Itô

$$F(t, T) = F(0, T) + \int_0^t \tilde{\Gamma}_u d\tilde{B}_u, \quad 0 \leq t \leq T,$$

para algún integrando proceso adaptado $\tilde{\Gamma}$ (es decir, $dF(t, T) = \tilde{\Gamma}_t d\tilde{B}_t$). Sean dados $0 \leq t_0 \leq t_1 \leq T$ y consideremos que en el tiempo t donde $t_0 \leq t \leq t_1$ mantenemos la posición de compra de $\Delta(t)$ contratos futuros. No nos cuesta nada cambiar de posición en los contratos futuros, pero como los contratos futuros generan flujo de caja, podemos invertir este flujo en un bono de corto plazo libre de riesgo a una tasa de interés constante o a una tasa r_t . Nuestra ganancia de realizar esta operación de trading es un proceso estocástico $G(t)$ que cumple

$$dG(t) = \Delta(t) dF(t, T) + r_t G(t) dt = \Delta(t) \tilde{\Gamma}_t d\tilde{B}_t + r_t G(t) dt$$

y por lo tanto el valor presente de nuestra ganancia

$$d(D(t) G(t)) = D(t) \Delta(t) \tilde{\Gamma}_t d\tilde{B}_t.$$

Asumamos que en el tiempo t_0 nuestra ganancia es $G_{t_0} = 0$. En el tiempo t_1 , nuestra ganancia $G(t_1)$ cumplirá

$$D(t_1) G(t_1) = \int_{t_0}^{t_1} D(u) \Delta(u) \tilde{\Gamma}_u d\tilde{B}_u.$$

Como las integrales de Itô son martingalas tenemos

$$\begin{aligned} E^{\mathbb{Q}} [D(t_1) G(t_1) | \mathcal{F}_{t_0}] &= E^{\mathbb{Q}} \left[\int_0^{t_1} D(u) \Delta(u) \tilde{\Gamma}_u d\tilde{B}_u - \int_0^{t_0} D(u) \Delta(u) \tilde{\Gamma}_u d\tilde{B}_u \mid \mathcal{F}_{t_0} \right] \\ &= E^{\mathbb{Q}} \left[\int_0^{t_1} D(u) \Delta(u) \tilde{\Gamma}_u d\tilde{B}_u \mid \mathcal{F}_{t_0} \right] - \int_0^{t_0} D(u) \Delta(u) \tilde{\Gamma}_u d\tilde{B}_u \\ &= 0 \end{aligned}$$

De acuerdo a la fórmula de valoración de riesgo neutral, el valor en el tiempo t_0 de una ganancia de G_{t_1} obtenida en el tiempo t_1 es

$$\frac{1}{D(t_0)} E^{\mathbb{Q}} [D(t_1) G(t_1) | \mathcal{F}_{t_0}],$$

y se ha mostrado que este valor es cero. □

2.3. Modelo spot ajustado al riesgo neutral

Enunciamos a continuación algunos Teoremas que nos serán de ayuda para la representación de nuestro modelo spot ajustado a la medida \mathbb{Q} , como veremos en el Lema 2.3.5 a partir del cual se deriva la curva de futuros mediante la ecuación 2.25.

El teorema de Girsanov construye explícitamente una medida de probabilidad que permite transformar un 'movimiento Browniano con tendencia' en un movimiento Browniano sin tendencia, este último definido en un espacio de probabilidad equivalente. A continuación se menciona la relación que existe entre el Teorema de Girsanov, la exponencial estocástica de Doleans y la condición de Novikov.

Teorema 2.3.1. (*Girsanov Multidimensional*). *Dado un proceso Browniano estándar $B_t = (B_t^1, \dots, B_t^n)$ en un espacio de probabilidad completo $(\Omega, \mathcal{F}, \mathbb{P})$ y sea $\{\mathcal{F}_t\}$, con $0 \leq t \leq \bar{T}$, la filtración para este movimiento Browniano. Supongamos que existe un proceso $\Theta(t) = (\Theta_1(t), \dots, \Theta_n(t))$ adaptado de dimensión n (al cual lo llamaremos Girsanov kernel). Se define el proceso Z_t sobre $[0, \bar{T}]$ como la exponencial martingala estocástica de Doleans,*

$$Z(t) = \exp \left\{ \int_0^t \Theta(u) \cdot dB_u - \frac{1}{2} \int_0^t \|\Theta(u)\|^2 du \right\}, \quad t \leq \bar{T},$$

tal que $\Theta(t)$ satisface la condición de Novikov

$$E \left[\exp \left(\frac{1}{2} \int_0^{\bar{T}} \|\Theta(u)\|^2 du \right) \right] < \infty. \quad (2.13)$$

Si asumimos que $E[Z(T)] = 1$ y definimos la nueva medida de probabilidad \mathbb{Q} sobre \mathcal{F}_T por

$$d\mathbb{Q} = Z(T) d\mathbb{P},$$

entonces,

$$d\tilde{B}_t := dB_t + \Theta(t) dt \quad (2.14)$$

es un proceso Browniano con respecto a la medida \mathbb{Q} .

El Teorema requiere una condición técnica de Novikov dada por 2.13 para garantizar que $E Z(T) = 1$, i.e. \mathbb{Q} sea una medida de probabilidad.

De otro lado, el Girsanov kernel $\Theta(t)$ se relaciona con el precio de mercado del riesgo $\lambda(t)$ que se introdujo en la sección anterior en 2.7. El siguiente Teorema recoge esa idea.

Teorema 2.3.2. (*Precio de Mercado del Riesgo*). *En ausencia de arbitraje, existirá un proceso vector λ_t que verifica 2.7. El precio de mercado del riesgo λ_t está relacionado con el kernel de Girsanov a través de $\lambda_t = -\Theta(t)$. Es decir, se cumple*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp \left\{ - \int_0^t \lambda_u dB_u - \frac{1}{2} \|\lambda_u\|^2 du \right\}.$$

En mercados incompletos existen varias posibilidades de procesos de precios de mercado de riesgo y de medidas neutrales al riesgo consistentes con el principio de mercado libre de arbitraje. Por lo tanto, en nuestro caso $\{\Theta, \lambda, \mathbb{Q}\}$ no se determinan por la restricción de ausencia de arbitraje, sino por la oferta y demanda del mercado; es decir por las preferencias de los agentes.

En la siguiente Proposición, ver Shreve (2004, pag. 257), se muestra las condiciones bajo las cuales el coeficiente de correlación entre los movimientos Brownianos bajo la medida \mathbb{P} no se modifica con el cambio de medida \mathbb{Q} .

Proposición 2.3.3. (*Correlación bajo el cambio de medida*). Sea \bar{B}_t^i el movimiento Browniano correlacionado definido en 1.4 bajo la medida \mathbb{P} . Suponga que exista $\Theta(t) = (\Theta_1(t), \dots, \Theta_n(t))$ tal que verifica las condiciones del Teorema Girsanov. Se cumple los siguientes resultados:

(a) Para $i = 1, \dots, n$, si definimos $\gamma(t)$ por

$$\gamma_i(t) := \sum_{j=1}^n \frac{\sigma_{ij} \Theta_j}{\|\sigma_i\|},$$

entonces se cumple que

$$d\tilde{W}_t^i := d\bar{B}_t^i + \gamma_i(t) dt$$

es un movimiento Browniano bajo \mathbb{Q} .

(b) De la definición de \bar{B}_t^i , ver 1.5, se tiene que $d\bar{B}_t^i d\bar{B}_t^j = \rho_{ij} dt$. Este es la correlación instantánea entre \bar{B}_t^i y \bar{B}_t^j . Si ρ_{ij} es no aleatoria (aunque pueda depender de t), entonces para cada $t \geq 0$,

$$E \left[d\bar{B}_t^i d\bar{B}_t^j \right] = E \left[d\tilde{W}_t^i d\tilde{W}_t^j \right] = \rho_{ij} dt$$

Para representar nuestro modelo spot bajo la medida neutral al riesgo \mathbb{Q} realizaremos el siguiente supuesto.

(A.5) Los precios de mercado de riesgo $\lambda := (\lambda_1, \lambda_2)$ en el modelo de Schwartz y Smith se asumen que son constantes⁷.

⁷Alternativamente, Espen (2011, pag. 12), plantea otro precio de mercado de riesgo: $\lambda_1 := -\frac{\kappa\beta}{\sigma_x} \chi_t$, $\lambda_2 := -\frac{1}{\sqrt{1-\rho_{\chi\xi}^2}} \left(\frac{\theta}{\sigma_\xi} - \rho_{\chi\xi} \frac{\kappa\beta}{\sigma_x} \chi_t \right)$, donde β y θ representan la escala de velocidad de reversión a la media y el nivel de ajuste, respectivamente

Schwartz y Smith construyen una medida de probabilidad equivalente que transforma un 'movimiento Browniano con tendencia constante' en otro movimiento Browniano sin tendencia. Es decir, se define el proceso estocástico

$$\begin{aligned} Z(t, B_t) &:= \exp \left\{ -\frac{1}{2} \int_0^T (\lambda_1^2 + \lambda_2^2) dt - \int_0^T \lambda_1 dB_t^1 - \int_0^T \lambda_2 dB_t^2 \right\} > 0 \\ &= \exp \left\{ -\frac{1}{2} (\lambda_1^2 + \lambda_2^2) T - \lambda_1 B_T^1 - \lambda_2 B_T^2 \right\}, \end{aligned} \quad (2.15)$$

junto con la derivada de Radon-Nikodym

$$\mathbb{Q}(A) = \int Z(b) d\mathbb{P}(b), \quad A \in \mathcal{F}_T \quad \text{donde} \quad d\mathbb{P}(b) = \frac{1}{\sqrt{2\pi T}} e^{-\frac{b^2}{2T}} db.$$

Para mostrar que λ se encuentra bien definido en el sentido del Teorema de Girsanov, debemos verificar que $Z(t)$ es una \mathbb{P} -martingala con respecto a $\mathbb{F} = \{\mathcal{F}_t\}_{0 \leq t \leq T}$. Para ello, es suficiente mostrar que se cumpla la condición de Novikov

$$\sup_{t \leq T} E \left[e^{\delta(\lambda_1^2 + \lambda_2^2)} \right] < \infty,$$

para algún $\delta > 0$, lo cual verifica de manera trivial al ser λ constante. Como consecuencia, para $B_t := (B_t^1, B_t^2)_{t \in [0, T]}$ se muestra que

$$\begin{aligned} E^{\mathbb{P}} \left[e^{-\lambda B_T - \frac{1}{2} \lambda^2 T} \mid \mathcal{F}_t \right] &= E^{\mathbb{P}} \left[e^{-\lambda(B_T - B_t) - \lambda B_t - \frac{1}{2} \lambda^2 T} \mid \mathcal{F}_t \right] \\ &= e^{-\lambda B_t - \frac{1}{2} \lambda^2 T} E^{\mathbb{P}} \left[e^{-\lambda(B_T - B_t)} \mid \mathcal{F}_t \right] \\ &= e^{-\lambda B_t - \frac{1}{2} \lambda^2 T} E^{\mathbb{P}} \left[e^{-\lambda(B_T - B_t)} \right] \\ &= e^{-\lambda B_t - \frac{1}{2} \lambda^2 T} e^{\frac{1}{2} \lambda^2 (T-t)} \\ &= e^{-\lambda B_t - \frac{1}{2} \lambda^2 t}, \end{aligned}$$

es una \mathbb{P} -martingala. Por lo tanto, el Teorema de Girsanov, nos garantiza de que el proceso \tilde{B}_t con componentes

$$\begin{aligned} \tilde{B}_t^1 &= B_t^1 + \lambda_1 t \\ \tilde{B}_t^2 &= B_t^2 + \lambda_2 t, \end{aligned} \quad (2.16)$$

es un \mathbb{Q} -proceso Browniano con tendencia constante $\lambda = (\lambda_1, \lambda_2)$. En efecto, es fácil ver que $\tilde{B}_t^1 = B_t^1 + \lambda_1 t$ es un proceso Browniano: $\tilde{B}_0^1 = 0$; asimismo, la continuidad de la función

$t \mapsto \tilde{B}_t^1$ es inmediata, ya que son funciones continuas con respecto de t . Por otro lado, la independencia de los incrementos están garantizados, pues

$$\tilde{B}_t^1 - \tilde{B}_s^1 = B_t^1 - B_s^1 + \lambda_1(t - s), \text{ para todo } s \leq t,$$

y $B_t^1 - B_s^1$ es independiente de \mathcal{F}_s con $\lambda_1(t - s)$ una función determinista. Por lo tanto, $\tilde{B}_t^1 - \tilde{B}_s^1$ es también independiente de \mathcal{F}_s ; de la misma manera se muestra que \tilde{B}_t^2 es un proceso Browniano.

Observación 2.3.4. Notemos que bajo \mathbb{P} , el proceso \tilde{B}_T^1 tiene una distribución normal con media $\lambda_1 T$ y varianza T . De 2.16, para $\tilde{b} = b + \lambda_1 T$ se cumple

$$d\mathbb{P}(\tilde{b}) = \frac{1}{\sqrt{2\pi T}} e^{-\frac{(\tilde{b} - \lambda_1 T)^2}{2T}} d\tilde{b},$$

$$\begin{aligned} E^{\mathbb{P}}[\tilde{B}_T^1] &= E^{\mathbb{P}}[B_T^1 + \lambda_1 T] = \lambda_1 T \text{ y} \\ \text{Var}^{\mathbb{P}}[\tilde{B}_T^1] &= \text{Var}^{\mathbb{P}}[B_T^1 + \lambda_1 T] = T. \end{aligned}$$

Sin embargo, bajo \mathbb{Q} , se remueve el término de tendencia de \tilde{B}_T^1 . En efecto, de 2.15 para el caso de $\tilde{B}_T^2 = 0$, se tiene

$$\begin{aligned} d\mathbb{Q}(\tilde{b}) &= Z(t; b) d\mathbb{P}(b) \\ &= Z(t; \tilde{b} - \lambda_1 T) d\mathbb{P}(\tilde{b} - \lambda_1 T) \\ &= \frac{1}{\sqrt{2\pi T}} e^{-\lambda_1(\tilde{b} - \lambda_1 T) - \frac{1}{2}\lambda_1^2 T} e^{-\frac{(\tilde{b} - \lambda_1 T)^2}{2T}} d\tilde{b} \\ &= \frac{1}{\sqrt{2\pi T}} e^{-\lambda_1 \tilde{b} + \frac{1}{2}\lambda_1^2 T} e^{-\frac{(\tilde{b} - \lambda_1 T)^2}{2T}} d\tilde{b} \\ &= \frac{1}{\sqrt{2\pi T}} e^{-\tilde{b}^2/2T} d\tilde{b}. \end{aligned} \tag{2.17}$$

Es decir, bajo \mathbb{Q} el proceso \tilde{B}_t^1 es normal con media cero y varianza T ⁸.

⁸Falta mostrar para $\tilde{B}_T^2 \neq 0$, pero el procedimiento es similar

Modelo Spot ajustado

A continuación procederemos con el ajuste de nuestro modelo spot. De la ecuación 2.16 y de la Proposición 2.1.6, se muestra que los procesos χ_t y ξ_t bajo la medida neutral al riesgo se describen de la siguiente manera

$$\begin{aligned} d\chi_t &= -\kappa\chi_t dt + \sigma_\chi (d\tilde{B}_t^1 - \lambda_1 dt) \\ d\xi_t &= \mu_\xi dt + \sigma_\xi \rho_{\chi\xi} (d\tilde{B}_t^1 - \lambda_1 dt) + \sigma_\xi \sqrt{1 - \rho_{\chi\xi}^2} (d\tilde{B}_t^2 - \lambda_2 dt). \end{aligned}$$

Simplificando,

$$d\chi_t = (-\kappa\chi_t - \sigma_\chi \lambda_1) dt + \sigma_\chi d\tilde{B}_t^\chi \quad (2.18a)$$

$$d\xi_t = (\mu_\xi - \sigma_\xi \lambda_2^*) dt + \sigma_\xi d\tilde{B}_t^\xi, \quad (2.18b)$$

donde $\lambda_2^* \triangleq \lambda_1 \rho_{\chi\xi} + \lambda_2 \sqrt{1 - \rho_{\chi\xi}^2}$, $d\tilde{B}_t^\xi \triangleq \rho_{\chi\xi} d\tilde{B}_t^1 + \sqrt{1 - \rho_{\chi\xi}^2} d\tilde{B}_t^2$ y $\tilde{B}_t^\chi = \tilde{B}_t^1$. Luego de la caracterización de Levy (ver Karatzas y Shreve , pag. 157) y la Proposición 2.3.3 se tiene que \tilde{B}_t^ξ es proceso Browniano y que $d\tilde{B}_t^\chi \cdot d\tilde{B}_t^\xi = \rho_{\chi\xi} dt$.

El siguiente Lema se obtiene de la ecuación 2.18a y 2.18b y muestra la transformación del modelo de Schwartz y Smith bajo la medida neutral al riesgo.

Lema 2.3.5. *En el modelo de Schwartz y Smith, la dinámica de los procesos χ_t y ξ_t bajo \mathbb{Q} se reducen a*

$$d\chi_t = (-\kappa\chi_t - \lambda_\chi) dt + \sigma_\chi d\tilde{B}_t^\chi \quad (2.19a)$$

$$d\xi_t = (\mu_\xi - \lambda_\xi) dt + \sigma_\xi d\tilde{B}_t^\xi \quad (2.19b)$$

donde $\lambda_\chi \triangleq \sigma_\chi \lambda_1$ y $\lambda_\xi \triangleq \sigma_\xi \lambda_2^*$.

Si comparamos la dinámica de los procesos bajo la medida \mathbb{P} con los obtenidos bajo \mathbb{Q} , se observa que el proceso χ_t bajo \mathbb{Q} sigue siendo un proceso Ornstein-Uhlenbeck pero ahora revierte a $-\lambda_\chi/\kappa$, mientras que bajo \mathbb{P} este revertía a 0. El proceso ξ_t bajo \mathbb{Q} sigue siendo un movimiento Browniano aritmético, pero con drift $\mu_\xi^* \triangleq \mu_\xi - \lambda_\xi$.

A continuación se enuncia el siguiente Lema, el cual nos será útil para encontrar la ecuación de la curva de futuros del modelo.

Lema 2.3.6. Dada la condición inicial (χ_t, ξ_t) , entonces para $T > t$ la solución χ_T y ξ_T es

$$\chi_T = e^{-\kappa(T-t)}\chi_t - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) + \sigma_\chi \int_t^T e^{-\kappa(T-s)} d\tilde{B}_s^\chi \quad (2.20a)$$

$$\xi_T = \xi_t + (\mu_\xi - \lambda_\xi)(T-t) + \sigma_\xi (\tilde{B}_T^\xi - \tilde{B}_t^\xi). \quad (2.20b)$$

El vector $(\chi_T, \xi_T)^\top$ se distribuye como un proceso Gaussiano con vector media y matriz covarianza dado por

$$\mathbb{E}^\mathbb{Q}[(\chi_T, \xi_T) | \mathcal{F}_t] = \left[e^{-\kappa(T-t)}\chi_t - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) , \xi_t + (\mu_\xi - \lambda_\xi)(T-t) \right], \quad y$$

$$\begin{aligned} Cov^\mathbb{Q}[\chi_T, \xi_T | \mathcal{F}_t] &= \begin{bmatrix} (1 - e^{-2\kappa(T-t)}) \frac{\sigma_\chi^2}{2\kappa} & (1 - e^{-\kappa(T-t)}) \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} \\ (1 - e^{-\kappa(T-t)}) \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} & \sigma_\xi^2 (T-t) \end{bmatrix}, \\ &= Cov^\mathbb{P}[\chi_T, \xi_T | \mathcal{F}_t]. \end{aligned}$$

Demostración. Se demuestra de manera similar al Lema 1.3.1,

$$\begin{aligned} E^\mathbb{Q}[\chi_T | \chi_t] &= E^\mathbb{Q} \left[e^{-\kappa(T-t)}\chi_t - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) + \sigma_\chi \int_t^T e^{-\kappa(T-s)} d\tilde{B}_s^\chi \mid \chi_t \right] \\ &= E^\mathbb{Q} \left[e^{-\kappa(T-t)}\chi_t - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) \mid \chi_t \right] + \sigma_\chi e^{-\kappa T} E^\mathbb{Q} \left[\int_t^T e^{\kappa s} d\tilde{B}_s^\chi \mid \chi_t \right] \\ &= e^{-\kappa(T-t)}\chi_t - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) \quad y \end{aligned}$$

$$\begin{aligned} E^\mathbb{Q}[\xi_T | \xi_t] &= E^\mathbb{Q} \left[\xi_t + (\mu_\xi - \lambda_\xi)(T-t) + \sigma_\xi (\tilde{B}_T^\xi - \tilde{B}_t^\xi) \mid \xi_t \right] \\ &= \xi_t + (\mu_\xi - \lambda_\xi)(T-t). \end{aligned}$$

La varianza se obtiene de la isometría de Itô:

$$\begin{aligned} Var^\mathbb{Q}[\chi_T | \chi_t] &= Var \left[e^{-\kappa(T-t)}\chi_t - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) + \sigma_\chi \int_t^T e^{-\kappa(T-s)} d\tilde{B}_s^\chi \mid \chi_t \right] \\ &= Var \left[\sigma_\chi e^{-\kappa T} \int_t^T e^{\kappa s} d\tilde{B}_s^\chi \mid \chi_t \right] \\ &= \sigma_\chi^2 e^{-2\kappa T} \int_t^T e^{2\kappa s} ds \\ &= (1 - e^{-2\kappa(T-t)}) \frac{\sigma_\chi^2}{2\kappa} \quad y \end{aligned}$$

$$\begin{aligned}
Var^{\mathbb{Q}} [\xi_T | \xi_t] &= Var \left[\xi_t + (\mu_\xi - \lambda_\xi) (T - t) + \sigma_\xi (\tilde{B}_T^\xi - \tilde{B}_t^\xi) \mid \xi_t \right] \\
&= \sigma_\xi^2 (T - t).
\end{aligned}$$

□

Como consecuencia del Lema 2.3.6 , el logaritmo del precio spot , $\ln (S_t)$ se distribuye como un proceso Gaussiano con vector media y matriz covarianza

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}} [\ln (S_T) \mid \mathcal{F}_t] &= e^{-\kappa(T-t)} \chi_t + \xi_t - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) + \\
&\quad (\mu_\xi - \lambda_\xi) (T - t) \\
&= \mathbb{E}^{\mathbb{P}} [\ln (S_T) \mid \mathcal{F}_t] - \frac{\lambda_\chi}{\kappa} (1 - e^{-\kappa(T-t)}) - \lambda_\xi (T - t) \quad (2.22a)
\end{aligned}$$

$$\begin{aligned}
Var^{\mathbb{Q}} [\ln (S_T) \mid \mathcal{F}_t] &= (1 - e^{-2\kappa(T-t)}) \frac{\sigma_\chi^2}{2\kappa} + \sigma_\xi^2 (T - t) + 2 (1 - e^{-\kappa(T-t)}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa} \\
&= Var^{\mathbb{P}} [\ln (S_T) \mid \mathcal{F}_t] \quad (2.22b)
\end{aligned}$$

Valoración de contratos futuros

Sabemos que se cumple $F(T, T) = S_T$, y de la Definición 2.2.2 nos dice que podemos calcular el precio futuro mediante la ecuación $F(t, T, S_T) = \mathbb{E}^{\mathbb{Q}} (S_T \mid \mathcal{F}_t)$. Entonces,

$$\begin{aligned}
\ln (F_t) &= \ln (\mathbb{E}^{\mathbb{Q}} [S_T \mid \mathcal{F}_t]) \\
&= \mathbb{E}^{\mathbb{Q}} [\ln S_T \mid \mathcal{F}_t] + \frac{1}{2} Var^{\mathbb{Q}} [\ln S_T \mid \mathcal{F}_t], \\
&= e^{-\kappa(T-t)} \chi_t + \xi_t + A(t, T), \quad (2.23)
\end{aligned}$$

en la última ecuación se está usando 2.22a y 2.22b donde

$$\begin{aligned}
A(t, T) &= (\mu_\xi - \lambda_\xi) (T - t) - (1 - e^{-\kappa(T-t)}) \frac{\lambda_\chi}{\kappa} \\
&\quad + \frac{1}{2} \left[(1 - e^{-2\kappa(T-t)}) \frac{\sigma_\chi^2}{2\kappa} + \sigma_\xi^2 (T - t) + 2 (1 - e^{-\kappa(T-t)}) \frac{\rho_{\chi\xi} \sigma_\chi \sigma_\xi}{\kappa} \right], \quad (2.24)
\end{aligned}$$

ver figura 2.1. De 2.23 , la varianza del $\ln (F_t)$ está dado por

$$\sigma_\chi^2 e^{-2\kappa(T-t)} + \sigma_\xi^2 + 2e^{-\kappa(T-t)} \rho_{\chi\xi} \sigma_\chi \sigma_\xi.$$

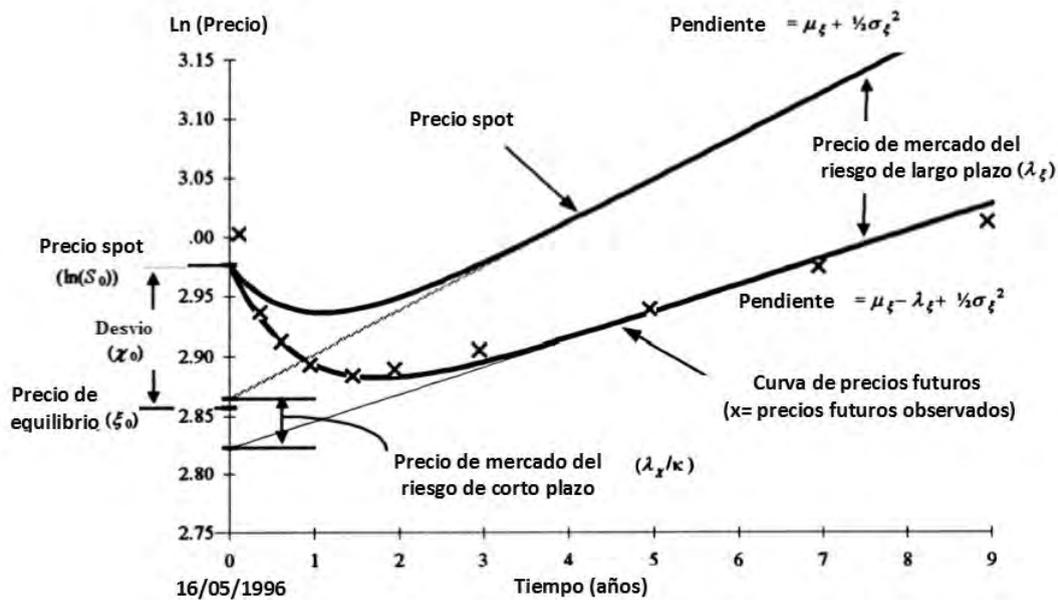
Para contratos próximos a vencer (i.e. $T = 0$) , la varianza es igual a $Cov (\chi_t + \xi_t)$. Cuando T aumenta, la contribución de la varianza de χ_t disminuye, y para $T \rightarrow \infty$, la varianza de $\ln (F_t)$ se aproxima a σ_ξ .

En forma matricial para una serie finita de contratos futuros con fechas de vencimiento T_1, \dots, T_n tendremos

$$\begin{bmatrix} \ln(F_{t,T_1}) \\ \vdots \\ \ln(F_{t,T_j}) \\ \vdots \\ \ln(F_{t,T_n}) \end{bmatrix} = \begin{bmatrix} A(t, T_1) \\ \vdots \\ A(t, T_j) \\ \vdots \\ A(t, T_n) \end{bmatrix} + \begin{bmatrix} e^{-\kappa(T_1-t)} & 1 \\ \vdots & \vdots \\ e^{-\kappa(T_j-t)} & 1 \\ \vdots & \vdots \\ e^{-\kappa(T_n-t)} & 1 \end{bmatrix} \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} \quad (2.25)$$

Cuando apliquemos filtros de Kalman en el capítulo final asumiremos que la data de precios futuros contiene errores de medición debido a cotizaciones de precios asincrónicos, y entonces agregaremos a la ecuación de futuros 2.25 un proceso ruido blanco v_t .

Figura 2.1.: Precios futuros y Esperanza de los precios spot



Capítulo 3.

El modelo en la forma espacio estado y propiedades de Markov

El modelo financiero de Schwartz y Smith es un sistema lineal continuo conformado por dos variables de estado $X_t = (\chi_t, \xi_t)$ y múltiples curvas de futuros $Y_t \equiv \ln(F_{t, T_n, X_t})$ asociados a un número finito de contratos futuros con diferentes fechas de vencimiento T_n . En este capítulo se presenta la notación de este sistema en la forma espacio estado. Veremos que la estructura probabilística que induce este sistema dinámico es la de un modelo oculto de Markov. La motivación de la representación espacio estado se debe a que nos brinda el sistema de ecuaciones en la forma apropiada para el análisis posterior del problema de estimación de filtraje con el Filtro de Kalman.

Definición 3.0.1. (Clasificación del Proceso estocástico). Si los vectores aleatorios son discretos, decimos que el proceso estocástico tiene espacio estado discreto. Si son continuos, se dice que el proceso tiene espacio estado continuo.

La forma espacio estado está conformado por dos sistemas de ecuaciones: (i) la ecuación de medición, la cual relaciona las variables de estado con las variables que pueden ser observadas y las cuales a su vez incluyen un ruido en su medición y (ii) la ecuación de transición, la cual describe la dinámica de las variables no observables, llamado también variables de estados. La formalización de esta definición se presenta a continuación y es el caso general para nuestro modelo lineal.

Definición 3.0.2. (Ecuación de Medición). Representa la relación funcional de las observaciones medibles Y_t con el vector de estado X_t . Se asume que es de la forma lineal tal que

$$Y_t = d + F' X_t + v_t, \quad (3.1)$$

con vector d y matriz F de apropiado tamaño, y un proceso aleatorio v_t . Además se asume para v_t

$$E[v_t] = 0 \quad , \quad E[v_t v_s^T] = \begin{cases} V & \text{para } s = t \\ 0 & \text{otro caso.} \end{cases}$$

para cada t , y las condiciones

$$E[v_t X_s] = 0 \quad y \quad E[v_t \omega_s] = 0, \forall t, s.$$

Definición 3.0.3. (Ecuación de Transición). Es el sistema que describe la evolución de las variables de estado X_t sobre el tiempo t . Se asume la ecuación lineal

$$X_t = c + GX_{t-1} + \omega_t, \tag{3.2}$$

con vector c y matriz de transición G de apropiado tamaño, y un proceso aleatorio ω_t . Además se asume para v_t

$$E[\omega_t] = 0, \quad E[\omega_t \omega_s^T] = \begin{cases} \mathbf{W} & \text{para } s = t \\ 0 & \text{otro caso.} \end{cases}$$

para cada t , y que

$$E[\omega_t X_s] = 0 \quad \text{para } s \leq t.$$

Para, nuestros propósitos de derivación de las ecuaciones recursivas del filtro de Kalman nos centraremos en las ecuaciones 3.1 y 3.2.

3.1. Representación espacio estado del modelo

La representación espacio estado en tiempo continuo se deriva directamente de nuestro modelo spot con excepción de la introducción del error en la ecuación de medición v_t . Estos errores de medición pueden interpretarse como los errores de ajuste en el modelo de los precios observados.

La caracterización de nuestro modelo en la forma espacio estado es la siguiente:

- El estado del sistema dinámico está conformado por $X_t \in \mathbb{R}^2$, tal que dado el estado inicial X_{t_0} , tenemos que X_t puede ser únicamente determinado para $t > t_0$.
- El modelo espacio estado está dado por la ecuación de transición

$$\begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} = \begin{bmatrix} 0 \\ \mu_\xi(t - t_0) \end{bmatrix} + \begin{bmatrix} e^{-\kappa(t-t_0)} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \chi_{t_0} \\ \xi_{t_0} \end{bmatrix} + \begin{bmatrix} \sigma_\chi \int_{t_0}^t e^{-\kappa(t-\tau)} dB_\tau^1 \\ \sigma_\xi \int_{t_0}^t (\rho_{\chi\xi} dB_\tau^1 + \sqrt{1 - \rho_{\chi\xi}^2} dB_\tau^2) \end{bmatrix}, \tag{3.3}$$

y la ecuación de medición

$$\begin{bmatrix} \ln(F_{t,T_1}) \\ \vdots \\ \ln(F_{t,T_j}) \\ \vdots \\ \ln(F_{t,T_n}) \end{bmatrix} = \begin{bmatrix} A(t, T_1) \\ \vdots \\ A(t, T_j) \\ \vdots \\ A(t, T_n) \end{bmatrix} + \begin{bmatrix} e^{-\kappa(T_1-t)} & 1 \\ \vdots & \vdots \\ e^{-\kappa(T_j-t)} & 1 \\ \vdots & \vdots \\ e^{-\kappa(T_n-t)} & 1 \end{bmatrix} \begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} + v_t. \quad (3.4)$$

El espacio estado discreto puede verse como una aproximación del espacio estado continuo. La motivación principal de pasar de tiempo continuo a discreto es que en la calibración de X_t , Schwartz y Smith usan observaciones spot y de precios de contratos futuros discreto. La data discreta consiste en precios diarios de cierre de contratos futuros de cierto commodity como el cobre (\$/TM) que fueron transados en el London Metal Exchange desde 2002-10-31 a 2017-10-13 (3779 observaciones).

Encontrar la apropiada discretización de un proceso estocástico continuo implica realizar un análisis probabilístico riguroso. Una de las maneras más simples de la aproximación discreta de un proceso de Itô es la aproximación de Euler-Muruyama, ver Kloeden (1992) e Iacus (2008). Pero, en nuestro caso tenemos la solución explícita de la ecuación diferencial estocástica y podemos construir la representación discreta exacta de nuestro modelo; no obstante, principal atención merece la discretización del proceso Browniano.

Recordemos las propiedades de los dos movimientos Brownianos dados en 1.9.

$$\begin{aligned} B_t^\chi &= B_t^1 \\ B_t^\xi &= \rho_{\chi\xi} B_t^1 + \sqrt{1 - \rho_{\chi\xi}^2} B_t^2. \end{aligned} \quad (3.5)$$

La media de (B_t^χ, B_t^ξ) es cero y la varianza, $Var[B_t^1] = Var[B_t^\xi] = t$.

De la definición de la covarianza

$$Cov[B_t^1, B_t^\xi] = \rho_{\chi\xi} Cov[B_t^1, B_t^1] + \sqrt{1 - \rho_{\chi\xi}^2} Cov[B_t^1, B_t^2],$$

y de la independencia entre B_t^1 y B_t^2 tenemos

$$Cov[B_t^1, B_t^\xi] = \rho_{\chi\xi} t.$$

Estos dos procesos Brownianos definen las dos integrales estocásticas de Itô de nuestro sistema dado en 3.3:

$$\omega_t^1 := \sigma_\chi \int_{t_0}^t e^{-\kappa(t-\tau)} dB_\tau^1, \quad \omega_t^2 := \sigma_\xi \int_{t_0}^t (\rho_{\chi\xi} dB_\tau^1 + \sqrt{1 - \rho_{\chi\xi}^2} dB_\tau^2).$$

Por el Teorema 4.4.9 en Shreve (2004) y de la Isometría de Itô, tenemos que (ω_t^1, ω_t^2) se encuentran normalmente distribuidos con media cero y matriz de covarianza

$$[\omega_t^1, \omega_t^2] \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sigma_\chi^2}{2\kappa} (1 - e^{-2\kappa(t-\tau)}) & \frac{\sigma_\chi \sigma_\xi \rho_{\chi\xi}}{\kappa} (1 - e^{-\kappa(t-t_0)}) \\ \frac{\sigma_\chi \sigma_\xi \rho_{\chi\xi}}{\kappa} (1 - e^{-\kappa(t-t_0)}) & \sigma_\xi^2 (t - t_0) \end{bmatrix} \right).$$

Para construir la solución discreta de una ecuación diferencial estocástica sobre el intervalo $[0, T]$, se realiza una partición Δt de $[t_0, T]$: conjunto ordenado y finito $\Delta t = \{\tau_0, \tau_1, \dots, \tau_L\}$ tal que $t_0 = \tau_0 < \tau_1 < \dots < \tau_j < \dots < \tau_L = T$, y para intervalos de tiempo equidistantes

$$\tau_j = t_0 + j\Delta t.$$

Sea $\Delta t \equiv$ tamaño de paso y $n_T \equiv$ número de periodos en el conjunto de datos; considerando la Definición 3.0.2 se presenta el siguiente Lema.

Lema 3.1.1. *La Representación Espacio Estado del modelo de Schwartz y Smith es del tipo Gaussiano lineal conformado por la ecuación de transición 3.6 y la ecuación de medición 3.7.*

- La ecuación de transición es de forma lineal

$$\begin{bmatrix} \chi_t \\ \xi_t \end{bmatrix} = \begin{bmatrix} 0 \\ \mu_\xi \Delta t \end{bmatrix} + \begin{bmatrix} e^{-\kappa \Delta t} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \chi_{t-1} \\ \xi_{t-1} \end{bmatrix} + \begin{bmatrix} \omega_t^1 \\ \omega_t^2 \end{bmatrix}, \quad t = 1, \dots, n_T$$

o en forma resumida

$$X_t = c + GX_{t-1} + \omega_t, \quad (3.6)$$

donde ω_t es un vector aleatorio normalmente distribuido con

$$E[\omega_t] = 0, \quad E[\omega_t \omega_s^T] = \begin{cases} \mathbf{W} & \text{para } s = t \\ 0 & \text{otro caso,} \end{cases}$$

con $\mathbf{W} = Cov[(\chi_{\Delta t}, \xi_{\Delta t})]$ coincide con 1.18b:

$$Cov[\omega_t] = \begin{bmatrix} (1 - e^{-2\kappa \Delta t}) \frac{\sigma_\chi^2}{2\kappa} & (1 - e^{-\kappa \Delta t}) \frac{\sigma_\chi \sigma_\xi \rho_{\chi\xi}}{\kappa} \\ (1 - e^{-\kappa \Delta t}) \frac{\sigma_\chi \sigma_\xi \rho_{\chi\xi}}{\kappa} & \sigma_\xi^2 \Delta t \end{bmatrix}.$$

- La ecuación de medición es de forma lineal con respecto a la variable de estado X_t :

$$\begin{bmatrix} \ln(F_{t,T_1}) \\ \vdots \\ \ln(F_{t,T_j}) \\ \vdots \\ \ln(F_{t,T_n}) \end{bmatrix} = \begin{bmatrix} A(t, T_1) \\ \vdots \\ A(t, T_j) \\ \vdots \\ A(t, T_n) \end{bmatrix} + \begin{bmatrix} e^{-\kappa(T_1-t)} & 1 \\ \vdots & \vdots \\ e^{-\kappa(T_j-t)} & 1 \\ \vdots & \vdots \\ e^{-\kappa(T_n-t)} & 1 \end{bmatrix} \begin{bmatrix} X_t \\ \xi_t \end{bmatrix} + v_t, \quad t = 1, \dots, n_T$$

con $A(t, T)$ dada por 2.24. En forma resumida

$$Y_t = d + F' X_t + v_t, \quad (3.7)$$

donde v_t es un vector aleatorio, normalmente distribuido con

$$E[v_t] = 0, \quad E[v_t v_s^T] = \begin{cases} \mathbf{V} & \text{para } s = t \\ 0 & \text{otro caso.} \end{cases}$$

En particular, la matriz de covarianza de v_t es del tipo

$$Cov[v_t] = \text{diag}\{s_1^2, s_2^2, \dots, s_n^2\},$$

donde s_n es un parámetro constante.

- Finalmente, se asume

$$E[\omega_t v_t] = 0, \quad (3.8)$$

y el vector estado inicial X_0 está normalmente distribuido, con $E[X_0] = m_0$ y $Cov[X_0] = C_0$. Asimismo, se asume

$$E[X_0 v_t] = 0 \quad y \quad E[X_0 \omega_t] = 0, \quad \forall t. \quad (3.9)$$

Observación 3.1.2. (Proceso ruido blanco). Cabe mencionar que en el espacio discreto se definen los procesos ruidos blancos v_t como aquellos procesos no correlacionados los cuales están idénticamente e independientemente distribuidos con media cero y varianza constante, en el sentido

$$Cor[v_t v_s^T] = \begin{cases} \mathbf{V} & \text{para } s = t \\ 0 & \text{otro caso.} \end{cases}$$

Consecuentemente,

$$\mathbf{V} = E[v_t v_t^T].$$

El marco de referencia para discutir los procesos ruidos blancos con media cero y varianza finita es el espacio de Hilbert con el producto interno $\langle \omega_t, v_t \rangle = E[\omega_t, v_t]$. En este caso, ω_t y v_t son ortogonales si y solamente si $\langle \omega_t, v_t \rangle = 0$. Este es la interpretación del supuesto en las ecuaciones 3.8 y 3.9. Además, notemos que un proceso Browniano discreto es un proceso ruido blanco con $\omega_t \sim \mathcal{N}(0, \Delta t)$ donde Δt es el tiempo discreto.

3.2. Simulación de precios del modelo

Para fines computacionales, es útil considerar la discretización de los procesos Brownianos, donde B_t es especificado para valores discretos de t en particiones $\Delta t = T/N$ para algún entero positivo N . Así, B_j denota a B_{τ_j} con $\tau_j = t_0 + j\Delta t$. Para $B_0 = 0$ tenemos que:

$$B_j - B_{j-1} = \sqrt{\Delta t} Z_j,$$

donde Z_j es una variable aleatoria independiente con distribución normal estándar, $Z_j \sim \mathcal{N}(0, 1)$. Una manera práctica de simular dos procesos Brownianos, consiste en generar variables aleatorias normal estándar

$$Z_j^1 \sim \mathcal{N}(0, 1)$$

$$Z_j^2 \sim \mathcal{N}(0, 1),$$

y si asumimos que están correlacionados, como en 3.5,

$$\text{Cor}(Z_j^1, Z_j^2) = \rho,$$

entonces construimos Z_j^2 como una forma lineal

$$Z_j^2 = \rho Z_j^1 + \epsilon_j, \tag{3.10}$$

donde $\epsilon_j \sim \mathcal{N}(0, 1 - \rho^2)$, ver Algoritmo 3.1.

De lo descrito en la sección anterior, tenemos que la discretización exacta del modelo del Lema 3.1.1 estaría dado por el siguiente sistema de ecuaciones:

$$\chi_j = \chi_{j-1} e^{-\kappa \Delta t} + \sigma_\chi \sqrt{\frac{1 - \exp(-2\kappa \Delta t)}{2\kappa}} Z_{j-1}^1 \tag{3.11a}$$

$$\xi_j = \xi_{j-1} + \mu_\xi \Delta t + \sigma_\xi \sqrt{\Delta t} Z_{j-1}^2, \tag{3.11b}$$

donde Z_t^1 y Z_t^2 definidos como en 3.10. La gráfica de la simulción del precio spot se encuentra en el anexo A.

Algoritmo 3.1 Simulación de dos Brownianos Correlacionados

```
(1) Generar n valores de la variable normal estándar  $Z_n^1$ 
(2) Generar n valores de la variable  $\epsilon_n$ , donde  $\epsilon_n \sim \mathcal{N}(0, 1 - \rho^2)$ ;
(3) Construir  $Z_n^2$  donde  $Z_n^2 = \rho Z_n^1 + \epsilon_n$ 
corGBM <- funcion(n, r, t=1/365, plot=TRUE)
{ #n es el número de muestra
#r es la correlacion
#t es el tamaño de paso
Z1 <- rnorm(n, mean=0, sd= 1)
se <- sqrt(1 - r^2) #deviacion estandar del error
e <- rnorm(n, mean=0, sd=se)
Z2 <- r*x + e
Z1 <- cumsum(Z1* sqrt(t))
Z2 <- cumsum(Z2* sqrt(t))
Max <- max(c(Z1,Z2))
Min <- min(c(Z1,Z2))
if(plot) {
plot(Z1, type="l", ylim=c(Min, Max))
lines(Z2, col="blue")
} return(cor(Z1,Z1))
}
#sample result
corGBM(10000, .85)
[1] 0.8523341
```

3.3. Modelo oculto de Markov

El modelo espacio estado descrito en la Definición 3.0.2 es un tipo particular de modelo oculto de Markov. A continuación, mostraremos la estructura probabilística que induce este sistema dinámico; para ello revisaremos brevemente la definición de Proceso de Markov en términos de un kernel de transición. Fijaremos dos espacios métricos \mathbb{S} y \mathbb{M} y denotaremos por \mathcal{S} y \mathcal{M} a sus respectivos σ -álgebra de Borelianos.

Definición 3.3.1. (Kernel de Transición). Dados $(\mathbb{S}, \mathcal{S}), (\mathbb{M}, \mathcal{M})$, y una función

$$K : \mathbb{S} \times \mathcal{M} \rightarrow [0, 1],$$

que cumple:

- (a) Para cada $x \in \mathbb{S}$, $K(x, \cdot)$ es una medida positiva sobre $(\mathbb{M}, \mathcal{M})$.
- (b) Para todo $A \in \mathcal{M}$, la función $x \rightarrow K(x, A)$ es medible, diremos que K es un kernel de transición cuando $K(x, \mathbb{M}) = 1$ para todo $x \in \mathbb{S}$; y K será un kernel de transición de Markov sobre $(\mathbb{S}, \mathcal{S})$ cuando $\mathbb{S} = \mathbb{M}$ y $K(x, \mathbb{S}) = 1, \forall x \in \mathbb{S}$.

Un kernel de transición K cuando admite una función densidad $K : \mathbb{S} \times \mathbb{M} \rightarrow [0, \infty]$ medible con respecto al sigma álgebra producto $\mathcal{S} \otimes \mathcal{M}$ lo denotaremos por,

$$K(x, A) = \int_A K(x, y) \mu(dy), \quad A \in \mathcal{M}.$$

Resumiremos algunas nociones importantes del kernel de transición que serán de utilidad para la formulación recursiva de la medida de filtraje.

- Las medidas positivas operan sobre un kernel de transición de dos maneras diferentes. Para la medida positiva μ definida sobre $(\mathbb{S}, \mathcal{S})$, definiremos la medida positiva μK sobre $(\mathbb{M}, \mathcal{M})$ del siguiente modo

$$\mu K(A) \triangleq \int \mu(dx) K(x, A), \quad A \in \mathcal{M}. \quad (3.12)$$

Además, definiremos la medida $\mu \otimes K$ sobre el espacio producto $(\mathbb{S} \times \mathbb{M}, \mathcal{S} \otimes \mathcal{M})$ por

$$\mu \otimes K(C) \triangleq \int \int 1_C(x, y) \mu(dx) K(x, dy), \quad C \in \mathcal{S} \otimes \mathcal{M}.$$

Si μ es una medida de probabilidad entonces μK y $\mu \otimes K$ también serán medidas de probabilidad.

- El kernel de transición opera sobre funciones. Para $f : \mathbb{M} \rightarrow \mathbb{R}$ medible y acotada, definiremos la función Kf sobre \mathbb{S} del siguiente modo

$$Kf(x) \triangleq \int_{\mathbb{S}} K(x, dy) f(y), \quad x \in \mathbb{S} \quad (3.13)$$

con Kf función medible y acotada en \mathbb{S} .

- Para cualquier medida positiva μ sobre $(\mathbb{S}, \mathcal{S})$ y $f : \mathbb{M} \rightarrow \mathbb{R}$ medible sobre $(\mathbb{M}, \mathcal{M})$, es fácil verificar que

$$(\mu K)(f) = \mu(Kf) = \int \int \mu(dx) K(x, dy) f(y). \quad (3.14)$$

Definiremos a continuación un proceso de Markov sobre el espacio de trayectorias en \mathbb{S} , asociados a $\{K(x, A) : x \in \mathbb{S}, A \in \mathcal{S}\}$.

Definición 3.3.2. (Proceso de Markov Homogéneo). Considere un espacio de probabilidad filtrado con medida de probabilidad \mathbb{P} y una colección de kernel de transición de Markov $K(x, A) \in [0, 1], (x, A) \in (\mathbb{S}, \mathcal{S})$. Llamaremos un proceso de Markov bajo \mathbb{P} , a un proceso estocástico $\{X_t\}_{t \geq 0} \in \mathbb{S}$ adaptado a la filtración $\{\mathcal{F}_t\}_{t \geq 0}$ tal que

$$\mathbb{P}(X_{t+1} \in A \mid \mathcal{F}_t) = K(X_t, A),$$

para todo $t \geq 0$ y $A \in \mathcal{S}$.

Llamaremos distribución inicial del proceso de Markov a la distribución de X_0 y a \mathbb{S} como el espacio de estados.

Además, del hecho que $\mathcal{F}_t^X \subseteq \mathcal{F}_t$ para todo $t \geq 0$ vale que un proceso de Markov con respecto a la filtración \mathcal{F}_t es también proceso de Markov con respecto a su filtración natural

$$\mathbb{P}(X_{t+1} \in A \mid X_0, \dots, X_t) = K(X_t, A),$$

para todo $t \geq 0$ y $A \in \mathcal{S}$.

Sea ν una medida de probabilidad sobre $(\mathbb{S}, \mathcal{S})$. La siguiente proposición, sin demostración, nos dice que una propiedad fundamental del proceso de Markov $\{X_t\}_{t \geq 0}$, es que su distribución finita dimensional, se caracteriza por su distribución inicial y producto de kernels de transición.

Proposición 3.3.3. *Sea $\{X_t\}_{t \geq 0}$ un proceso de Markov con distribución inicial ν y un kernel de transición K . Para todo $t \geq 0$ y una función medible y acotado $f : \mathbb{S}^{t+1} \rightarrow \mathbb{R}$ vale*

$$E[f(X_0, \dots, X_t)] = \int f(x_0, \dots, x_t) \nu(dx_0) K(x_0, dx_1) \dots K(x_{t-1}, dx_t).$$

El siguiente ejemplo muestra que el sistema lineal definido por la ecuación 3.1 es un proceso de Markov.

Ejemplo 3.3.4. Fijemos $X_0 = x_0 \in \mathbb{R}$ y $\{\omega_t\}_{t \in \mathbb{N}}$ un proceso ruido blanco con función densidad ρ tal que X_0 es independiente de ω_t . Mostraremos que $(X_t)_{t \geq 0}$

$$X_t = c + GX_{t-1} + \omega_t, \tag{3.15}$$

es un proceso de Markov homogéneo en el espacio estado $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Para este proceso, tenemos

$$\mathbb{P}(X_t \in A \mid X_{t-1}) = \int_A \rho(X_t - c - GX_{t-1}) dx.$$

por tanto

$$\mathbb{P}(X_t \in A \mid X_{t-1}) = \int_A \rho(X_t - c - GX_{t-1}) dx.$$

Podemos definir el kernel de transición

$$K(x, A) = \int_A \rho(u - c - Gx) du, \quad x \in \mathbb{R}, A \in \mathcal{B}(\mathbb{R}).$$

Bastará con probar que

$$\mathbb{P}(X_t \in A \mid X_{t-1}, \dots, X_0) = K(X_{t-1}, A).$$

Comenzaremos calculando

$$\mathbb{P}(X_t \in A \mid X_{t-1}, \dots, X_0) = E(1_A(X_t) \mid X_{t-1}, \dots, X_0),$$

y del supuesto de la independencia de ω_t respecto a X_{t-1}, \dots, X_0

$$\begin{aligned} \mathbb{P}(X_t \in A \mid X_{t-1}, \dots, X_0) &= E(1_A(c + GX_{t-1} + \omega_t) \mid X_{t-1}, \dots, X_0) \\ &= E(1_A(c + GX_{t-1} + \omega_t) \mid X_{t-1} = x \mid X_{t-1}, \dots, X_0) \\ &= E(1_A(c + GX_{t-1} + \omega_t) \mid X_{t-1} = x). \end{aligned}$$

Para $X_t = u$, se obtiene que $\omega_t = u - c - Gx$; entonces

$$\begin{aligned} \mathbb{P}(X_t \in A \mid X_{t-1}, \dots, X_0) &= \int_A \rho(u - c - Gx) du \\ &= K(x, A). \end{aligned}$$

Modelo Oculto de Markov

Definiremos a continuación cuándo un modelo se llama modelo oculto de Markov. Para ello fijaremos dos espacios métricos \mathbb{S} y \mathbb{M} con sus respectivos σ -álgebra de Borelianos, \mathcal{S} y \mathcal{M} . Llamaremos un modelo oculto de Markov a un proceso doblemente estocástico parcialmente observado asociado a un kernel de transición definido sobre el espacio producto $(\mathbb{S} \times \mathbb{M}, \mathcal{S} \otimes \mathcal{M})$ con una estructura particular. El kernel de transición debe verificar que tanto el proceso conjunto $\{X_t, Y_t\}_{t \geq 0}$ y el proceso (marginal) no observable $\{X_t\}_{t \geq 0}$ son de Markov.

- $\{X_t\}_{t \geq 0}$ es un proceso de Markov con transición kernel K y distribución inicial ν ;
- Condicionado sobre el proceso estado $\{X_t\}_{t \geq 0}$, las observaciones $\{Y_t\}_{t \geq 0}$ son independientes, y para cada t la distribución condicional de Y_t depende solamente de X_t .

Definición 3.3.5. (Modelo Oculto de Markov). Respecto a los espacios medibles $(\mathbb{S}, \mathcal{S})$ y $(\mathbb{M}, \mathcal{M})$ denotaremos por K y Q , respectivamente, a un kernel de transición de Markov en $(\mathbb{S}, \mathcal{S})$ y un kernel de transición de $(\mathbb{S}, \mathcal{S})$ a $(\mathbb{M}, \mathcal{M})$. Definiremos el kernel de transición de Markov T sobre el espacio producto $(\mathbb{S} \times \mathbb{M}, \mathcal{S} \otimes \mathcal{M})$ como

$$\begin{aligned} T[(x, y), C] &\triangleq K \otimes Q[(x, y), C] \\ &\triangleq \iint K(x, dx') Q(x', dy') 1_C(x', y'), \end{aligned} \quad (3.16)$$

para $(x, y) \in (\mathbb{S} \times \mathbb{M})$ y $C \in (\mathcal{S} \otimes \mathcal{M})$.

Llamaremos un modelo oculto de Markov al par $\{X_t, Y_t\}_{t \geq 0}$ como un proceso de Markov con espacio de estados $(\mathbb{S} \times \mathbb{M})$, con kernel de transición de Markov T y distribución inicial

$$v \otimes Q(C) \triangleq \iint v(dx) Q(x, dy) 1_C(x, y),$$

donde v es una medida de probabilidad sobre $(\mathbb{S}, \mathcal{S})$. Se le denomina oculto porque el proceso X_t es oculto en el sentido de que no se tiene data discreta. En contraste, el proceso Y_t es observable.

Cuando el espacio \mathbb{M} es un subconjunto de \mathbb{R}^d para algún entero d , asumiremos que lo siguiente cumple.

Definición 3.3.6. (Modelo Oculto de Markov parcialmente dominada). El modelo de la Definición 3.3.5 se dice que es parcialmente dominada si existe una medida de probabilidad μ sobre $(\mathbb{M}, \mathcal{M})$ tal que para todo $x \in \mathbb{S}$, $Q(x, \cdot)$ es absolutamente continua con respecto a μ ; $Q(x, \cdot) \ll \mu(\cdot)$, con función densidad de transición $g(x, \cdot)$. Entonces, para todo $A \in \mathcal{M}$,

$$Q(x, A) = \int 1_A(y) g(x, y) \mu(dy). \quad (3.17)$$

En ese caso el kernel de transición conjunta T se escribe

$$T[(x, y), C] \triangleq \iint K(x, dx') g(x', y') 1_C(x', y') \mu(dy'). \quad (3.18)$$

para $(x, y) \in (\mathbb{S} \times \mathbb{M})$ y $C \in (\mathcal{S} \otimes \mathcal{M})$.

Note que en tales modelos, generalmente reusaremos la notación v para denotar la función de densidad de probabilidad del estado inicial X_0 (con respecto a λ) en vez de la distribución en sí misma.

En el capítulo cinco, donde consideramos la estimación estadística para modelos oculto de Markov con parámetros desconocidos, requeriremos condiciones más fuertes y asumiremos que el modelo está completamente dominado en el siguiente sentido.

Definición 3.3.7. (Modelo Oculto de Markov totalmente dominado). Si en adición a los requerimientos de la Definición 3.3.6, existe una medida de probabilidad λ sobre $(\mathbb{S}, \mathcal{S})$ tal que $v \ll \lambda$ y, para todo $x \in \mathbb{S}$, $K(x, \cdot) \ll \lambda(\cdot)$ con función de densidad de transición $K(x, \cdot)$, entonces, para $A \in \mathcal{S}$,

$$K(x, A) = \int 1_A(x') K(x, x') \lambda(dx'),$$

el modelo se llama totalmente dominado. La transición kernel conjunta de Markov T es entonces dominada por la medida producto $\lambda \otimes \mu$ y admite la función de densidad de transición

$$t[(x, y), (x', y')] \triangleq K(x, x') g(x', y'). \quad (3.19)$$

La ecuación 3.18 nos permite expresar la distribución conjunta del proceso $\{X_t, Y_t\}_{t \geq 0}$ de la siguiente manera

$$E_v[f(X_0, Y_0, \dots, X_t, Y_t)] = \int \cdots \int f(x_0, y_0, \dots, x_t, y_t) v(dx_0) g(x_0, y_0) \times \prod_{k=1}^t \{K(x_{k-1}, dx_k) g(x_k, y_k)\} \times \mu_t(dy_0, \dots, dy_t), \quad (3.20)$$

donde $\mu_t(dy_0, \dots, dy_t) = \mu(dy_0) \mu(dy_1) \dots \mu(dy_t)$ y vale para cualquier función medible y acotada $f : (\mathbb{S} \times \mathbb{M})^{t+1} \rightarrow \mathbb{R}$.

A partir de la Definición 3.3.5 sobre modelo oculto de Markov $\{X_t, Y_t\}_{t \geq 0}$ podemos señalar dos propiedades de los procesos X_t e Y_t . El cual solo la enunciamos y la prueba se encuentra en Cappé (2005, pag 67).

Proposición 3.3.8. (Propiedad de Markov del vector estado). Los estados $\{X_t, t \geq 0\}$ forman una secuencia de Markov. Es decir, X_t dado X_{t-1} es independiente de los eventos ocurridos antes de $t - 1$

$$E[f(X_t) | X_0, \dots, X_{t-1}, Y_0, \dots, Y_{t-1}] = E[f(X_t) | X_{t-1}]$$

y los eventos pasados es independiente del futuro dado el presente para

$$E[f(X_{t-1}) | X_t, \dots, X_T, Y_t, \dots, Y_T] = E[f(X_{t-1}) | X_t], \quad t \leq T.$$

Considerando la distribución conjunta del proceso $\{X_t, Y_t\}_{t \geq 0}$ de la ecuación 3.20 y de la propiedad de Markov de X_t se tiene

$$\begin{aligned} E_v [f(X_{t+1}, Y_{t+1}) \mid X_0, Y_0, \dots, X_t, Y_t] &= E_v [f(X_{t+1}, Y_{t+1}) \mid X_t] \\ &= \int \int f(x, y) K(X_t, dx) g(x, y) \mu(dy). \end{aligned}$$

Otras de las implicancias de la Definición 3.3.5 es la característica de la independencia condicional del proceso Y_t , es decir

$$E_v [f(Y_t) \mid X_0, \dots, X_t, Y_0, \dots, Y_{t-1}] = E_v [f(Y_t) \mid X_t].$$

Este resultado se plantea en la siguiente proposición y nos dice que $\{Y_t\}_{t \geq 0}$ es condicionalmente independiente dado $\{X_t\}_{t \geq 0}$.

Proposición 3.3.9. *Sea $\{X_t, Y_t\}_{t \geq 0}$ un proceso de Markov sobre el espacio producto $\mathbb{S} \times \mathbb{M}$ con kernel de transición T dado por la ecuación 3.16. Entonces, para cualquier entero p y conjunto ordenado $\{k_1 < \dots < k_p\}$ de índices y funciones medibles y acotadas en \mathbb{M} ,*

$$E_v \left[\prod_{i=1}^p f_i(Y_{k_i}) \mid X_{k_1}, \dots, X_{k_p} \right] = \prod_{i=1}^p \int f_i(y) g(X_{k_i}, dy)$$

Demostración. Para cualquier función medible y acotada h se cumple

$$\begin{aligned} E_v \left[\prod_{i=1}^p f_i(Y_{k_i}) h(X_{k_1}, \dots, X_{k_p}) \right] &= \int \dots \int v(dx_0) g(x_0, y_0) \left[\prod_{i=1}^{k_p} K(x_{i-1}, dx_i) g(x_i, y_i) \right] \times \\ &\quad \left[\prod_{i=1}^p f_i(y_{k_i}) \right] h(x_{k_1}, \dots, x_{k_p}) \left[\prod_{i=1}^p f_i(y_{k_i}) \right] h(x_{k_1}, \dots, x_{k_p}). \end{aligned}$$

Simplificando, se obtiene

$$\begin{aligned} E_v \left[\prod_{i=1}^p f_i(Y_{k_i}) h(X_{k_1}, \dots, X_{k_p}) \right] &= \int \dots \int v(dx_0) \prod_{i=1}^{k_p} K(x_{i-1}, dx_i) h(x_{k_1}, \dots, x_{k_p}) \times \\ &\quad \int \dots \int \prod_{i \notin \{k_1, \dots, k_p\}} g(x_i, dy_i) \prod_{i \in \{k_1, \dots, k_p\}} \int f_i(y_i) g(x_i, dy_i). \end{aligned}$$

Como $\int g(x_i, dy_i) = 1$,

$$\begin{aligned}
E_v \left[\prod_{i=1}^p f_i(Y_{k_i}) h(X_{k_1}, \dots, X_{k_p}) \right] \\
= E_v \left[h(X_{k_1}, \dots, X_{k_p}) \prod_{i \in \{k_1, \dots, k_p\}} \int f_i(y_i) g(X_i, dy_i) \right].
\end{aligned}$$

Además, cualquier información adicional que obtenemos de X_{k_p} no tendrá influencia en la distribución de Y_k . Es decir, para cualquier entero p y p' , e índices $k_1 < \dots < k_p$ y $k'_1 < \dots < k'_p$ tal que $\{k_1 < \dots < k_p\} \cap \{k'_1 < \dots < k'_p\} = \emptyset$ y para cualquier función medible y acotada f se cumple

$$\begin{aligned}
E_v \left[f(Y_{k_1}, \dots, Y_{k_p}) \mid X_{k_1}, \dots, X_{k_p}, X_{k'_1}, \dots, X_{k'_p}, Y_{k'_1}, \dots, Y_{k'_p} \right] \\
= E_v \left[f(Y_{k_1}, \dots, Y_{k_p}) \mid X_{k_1}, \dots, X_{k_p} \right].
\end{aligned}$$

□

Esta propiedad de independencia condicional se encuentra formalizada en el siguiente Corolario.

Corolario 3.3.10. *Para k y p enteros y un conjunto ordenado $\{k_1 < \dots < k_p\}$ de índices tal que $k \notin \{k_1 < \dots < k_p\}$, las variables aleatorias Y_k y $(X_{k_1}, \dots, X_{k_p})$ son \mathbb{P}_v -condicionalmente independientes dado X_k .*

Demostración. Sea h una función medible acotada en (\mathbb{S}^p) . De la proposición 3.3.9, para cualquier función f medible y acotada en \mathbb{M} ,

$$\begin{aligned}
E_v \left[f(Y_k) h(X_{k_1}, \dots, X_{k_p}) \mid X_k \right] \\
= E_v \left[E_v \left[f(Y_k) \mid X_{k_1}, \dots, X_{k_p}, X_k \right] h(X_{k_1}, \dots, X_{k_p}) \mid X_k \right] \\
= E_v \left[f(Y_k) \mid X_k \right] E_v \left[h(X_{k_1}, \dots, X_{k_p}) \mid X_k \right].
\end{aligned}$$

En la sección 4.3 se muestra que el modelo de Schwartz y Smith es un modelo de Markov Oculto. □

Capítulo 4.

Filtro de Kalman discreto

En esta sección se describe el método del Filtro de Kalman para la estimación de las variables de estado del precio spot. La estructura probabilística del modelo oculto de Markov es adoptado para encontrar las ecuaciones recursivas de filtraje. Los resultados teóricos son ilustrados con el modelo de precios spot y la ecuación de precios futuros de Schwartz y Smith.

Para definir nuestro problema de filtraje estocástico, consideraremos un proceso de Markov $X = \{X_t, t \in \mathbb{N}\}$ definido sobre cierto espacio de probabilidad que toma valores en un espacio de estados general, \mathbb{S} . Adicionalmente a X_t , existe otro proceso Y_t que es una función ruidosa de X_t definido sobre algún otro espacio \mathbb{M} . Más precisamente, fijaremos dos espacios métricos \mathbb{S} y \mathbb{M} , y en todo este capítulo denotaremos por \mathcal{S} y \mathcal{M} a sus respectivos σ -álgebra de Borelianos. A continuación se plantea el problema de filtraje estocástico.

Sea un sistema estocástico en su forma espacio estado, según la Definición 3.0.2:

$$\begin{aligned} X_t &= c + GX_{t-1} + \omega_t, \quad t = 0, 1, 2, \dots, n \\ Y_t &= d + F'X_t + v_t. \end{aligned} \tag{4.1}$$

Fijemos una secuencia finita de variables aleatorias $Y = [Y_1, \dots, Y_t]$ y sea \mathcal{M}_t la σ -álgebra generada por $\{Y_s(\cdot), s \leq t\}$. Supongamos que observando Y podemos estimar X . Es decir, estamos suponiendo que existe una función $h : \mathbb{R}^n \rightarrow \mathbb{R}$ Borel medible tal que

$$X = h \circ Y. \tag{4.2}$$

es \mathcal{G}_t medible. Pero la igualdad en 4.2 resulta ser casi imposible por lo que en el problema de filtraje necesitamos definir el estimador de X , al cual lo denotaremos por \hat{X} .

Definición 4.0.1. Sean X e Y definidos como en la ecuación 4.1.

(a) Por estimador \hat{X} de X dado Y , nos referimos a la variable aleatoria

$$\hat{X} = h \circ Y, \tag{4.3}$$

donde h es una función Borel medible, llamado también función estimación.

(b) Si $y_1 = Y_1, \dots, y_n = Y_n$ son mediciones de Y , entonces

$$\hat{x} = h(y_1, y_2, \dots, y_n)$$

es el estimador de X dado y_1, \dots, y_n . Asimismo, cuando $Y = [y_1, \dots, y_n]$ se cumple

$$\hat{x} = \hat{X} = h(Y).$$

Pero de todos los posibles estimadores, buscaremos el mejor estimador es decir un estimador de mínima varianza.

Definición 4.0.2. Un estimador \hat{X} de X dado Y es de mínima varianza cuando existe una función g que cumple la ecuación 4.3 y

$$\|\hat{X} - X\| \leq \|g \circ Y - X\| \quad (4.4)$$

Equivalentemente podemos expresar la ecuación 4.4 como

$$\int_{\Omega} (\hat{X} - X)^2 dP \leq \int_{\Omega} (g \circ Y - X)^2 dP.$$

La existencia y unicidad de la solución al problema de filtraje es garantizado por el siguiente Corolario. Sea \mathcal{L}_2 el espacio de las funciones cuadrados integrables, definimos

$$\mathcal{M}(Y) = \{h \circ Y \mid h \text{ una función Borel, } h \circ Y \in \mathcal{L}_2\}.$$

y por el Teorema 2.3.4 en Catlin (1980) el conjunto $\mathcal{M}(Y)$ es un subespacio cerrado en \mathcal{L}_2 .

Corolario 4.0.3. *El mejor estimador de X dado Y coincide con la proyección de X sobre el subespacio $\mathcal{M}(Y)$. Por lo tanto, la solución \hat{X} existe, es único y cumple la condición $(\hat{X} - X) \perp \mathcal{M}(Y)$.*

De lo cual se desprende el siguiente resultado.

Teorema 4.0.4. *El estimador de mínima varianza de X dado Y está dado por*

$$\hat{X} = E(X \mid Y).$$

Demostración. Sea h una función medible donde $h \circ Y \in \mathcal{M}(Y)$. Del Teorema de Pitágoras y del hecho de que $X \in \mathcal{L}_2$ por el Lema 1.9.9 en Catlin (1980) es decir

$$[X - E(X | Y)] \perp [E(X | Y) - h \circ Y]$$

se sigue

$$\begin{aligned} \|X - h \circ Y\|^2 &= \|X - E(X | Y) + E(X | Y) - h \circ Y\|^2 \\ &= \|X - E(X | Y)\|^2 + \|E(X | Y) - h \circ Y\|^2. \end{aligned}$$

Se encuentra el mínimo cuando $h \circ Y = E(X | Y)$. □

A partir de este resultado y de la Definición 4.0.1, necesitamos conocer la función $h(\cdot)$ para encontrar el estimador \hat{x} de X dado las mediciones y_1, \dots, y_n y es lo que se muestra en la Proposición 4.1.7.

Antes de empezar con la exposición de la estimación recursiva de las distribuciones de filtraje, revisaremos brevemente la noción de distribución condicional.

4.1. Conceptos preliminares

Siguiendo con la notación introducida en la sección 3.3 sobre modelo oculto de Markov, a continuación se introducen los conceptos de verosimilitud y distribución de suavizamiento los cuales definen la estructura de las ecuaciones recursivas de filtraje.

Definición 4.1.1. (Distribución condicional). Sea X una $(\mathbb{S}, \mathcal{S})$ - variable aleatoria e Y una $(\mathbb{M}, \mathcal{M})$ - variable aleatoria definidos sobre el espacio probabilístico $(\Omega, \mathcal{F}, \mathbb{P})$. Un kernel de transición $\pi_{X|Y} : \mathbb{M} \times \mathcal{S} \rightarrow [0, 1]$ que verifica

$$\int f(x) \pi_{X|Y}(Y, dx) = \mathbb{E}(f(X) | Y), \mathbb{P} - \text{casi ciertamente}$$

para cada función medible y acotada $f : \mathbb{S} \rightarrow \mathbb{R}$, se le denomina distribución condicional de X dado Y .

En la definición anterior, la existencia y unicidad de las distribuciones condicionales está garantizada cuando \mathbb{S} es el espacio Borel como en nuestro caso. No obstante, este aspecto técnico se puede relajar cuando el proceso Y admite la representación dada por 3.18.

La construcción de $\pi_{X|Y}$ de la Definición 4.1.1 se obtiene a partir del siguiente teorema.

Teorema 4.1.2. (Fórmula de Bayes). Sea X una $(\mathbb{S}, \mathcal{S})$ -variable aleatoria e Y una $(\mathbb{M}, \mathcal{M})$ -variable aleatoria sobre el espacio probabilístico $(\Omega, \mathcal{F}, \mathbb{P})$. Suponga que exista una función medible $g : \mathbb{S} \times \mathbb{M} \rightarrow (0, \infty)$, la ley μ_X en \mathbb{S} , y la ley μ_Y en \mathbb{M} , tal que

$$\mathbb{E}(f(X, Y)) = \int f(x, y) g(x, y) \mu_X(dx) \mu_Y(dy),$$

para toda función medible acotada f . Entonces,

$$\pi_{X|Y}(y, A) = \frac{\int 1_A(x) g(x, y) \mu_X(dx)}{\int g(x, y) \mu_X(dx)}, \forall A \in \mathcal{S}, y \in \mathbb{M}, \quad (4.5)$$

es la distribución condicional de X dado Y .

Verosimilitud

La probabilidad conjunta de estados no observables y de las observaciones hasta el índice t es tal que para cualquier función medible y acotada f sobre $\{\mathbb{S} \times \mathbb{M}\}^{t+1}$

$$E_v[f(X_0, Y_0, \dots, X_t, Y_t)] = \int \cdots \int f(x_0, y_0, \dots, x_t, y_t) \times v(dx_0) g(x_0, y_0) \prod_{k=1}^t \{K(x_{k-1}, dx_k) g(x_k, y_k)\} \mu_t(dy_0, \dots, dy_t), \quad (4.6)$$

donde μ_t denota la distribución producto $\mu^{\otimes(t+1)}$ sobre $(\mathbb{S}^{t+1}, \mathcal{S}^{\otimes(t+1)})$. Tomando marginales con respecto a las variables no observables X_0, \dots, X_t , se obtiene la distribución marginal de la observación,

$$E_v[f(Y_0, \dots, Y_t)] = \int \cdots \int f(y_0, \dots, y_t) L_{v,t}(y_0, \dots, y_t) \mu_t(dy_0, \dots, dy_t), \quad (4.7)$$

donde $L_{v,t}$ es una medida de probabilidad importante que se define a continuación.

Definición 4.1.3. (Verosimilitud o Likelihood). La función de verosimilitud de las observaciones es la función de densidad de probabilidad de Y_0, Y_1, \dots, Y_t con respecto a μ_t definida para todo $(y_0, \dots, y_t) \in \mathbb{M}^{t+1}$ por

$$L_{v,t}(y_0, \dots, y_t) = \int \cdots \int v(dx_0) g(x_0, y_0) \times \prod_{k=1}^t \{K(x_{k-1}, dx_k) g(x_k, y_k)\}. \quad (4.8)$$

Ejemplo 4.1.4. Sea el proceso de observaciones $Y = \{Y_t, t \in \mathbb{N}\}$ un proceso estocástico con valores en \mathbb{R}^m definido como

$$Y_t \triangleq h(X_t) + v_t, \quad t > 0, \quad (4.9)$$

con $Y_0 = 0$, $h: \mathbb{R}^d \rightarrow \mathbb{R}^m$ es una función Borel medible y para todo $t \in \mathbb{N}$, v_t es un proceso ruido blanco. Se denota por $g(t, \cdot)$ a la función densidad de v_t con respecto a la medida de Lebesgue $\mu \in \mathbb{R}^m$ con $g(\cdot) \in B(\mathbb{R}^d)$ estrictamente positiva. La función densidad del vector aleatorio (Y_s, \dots, Y_t) para un arbitrario $s, t \in \mathbb{N}$, $s \leq t$

$$L_t(y_s, \dots, y_t) = \int_{(\mathbb{R})^{t-s+1}} \prod_{i=s}^t g_i(y_i - h(x_i)) \mathbb{P}_{X_s, \dots, X_t}(dx_s, \dots, dx_t),$$

donde $\mathbb{P}_{X_s, \dots, X_t}$ es la distribución de probabilidad del vector aleatorio (X_s, \dots, X_t) . Si asumimos que el proceso X_t es proceso de Markov se sigue

$$L_{v,t}(y_s, \dots, y_t) = \int_{(\mathbb{R})^{t-s+1}} v(dx_s) g(y_s - h(x_s)) \times \prod_{i=s+1}^t \{K(x_{i-1}, dx_i) g(y_i - h(x_i))\} \quad (4.10)$$

Un caso especial de 4.10 es cuando

$$\mathbb{P}(Y_t \in dy_t \mid X_t = x_t) = g(y_t - h(x_t)) dy_t.$$

Como g_i para $i = s, \dots, t$ son estrictamente positivas, la función densidad del vector (Y_s, \dots, Y_t) es también estrictamente positiva.

Si en el ejemplo anterior además asumimos que $v_t \sim \mathcal{N}(0, \mathbf{V})$, entonces

$$g(y_t - h(x_t)) = (2\pi)^{-\frac{m}{2}} \det(\mathbf{V})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_t - h(x_t))^\top \mathbf{V}^{-1}(y_t - h(x_t))\right\},$$

$$Q(x, B) = \int_{(\mathbb{R}^m)} 1_B(z) (2\pi)^{-\frac{m}{2}} \det(\mathbf{V})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(z - h(x))^\top \mathbf{V}^{-1}(z - h(x))\right\} dz.$$

En lo sucesivo usaremos la siguiente notación :

Observación 4.1.5. (Notación Concisa de Subsecuencias). Por simplicidad , usaremos en lo sucesivo la notación

$$X_{l:m} \triangleq (X_l, \dots, X_m) \quad e \quad Y_{l:m} \triangleq (Y_l, \dots, Y_m),$$

para denotar la colección de variables indexadas consecutivas X_l, \dots, X_m y de la misma forma para la secuencia $\{Y_t\}$. Por ejemplo, en 4.8 reescribiríamos $L_{v,t}(y_0, \dots, y_t)$ por $L_{v,t}(y_{0:t})$. Por convención, $Y_{t|t}$ se refiere a la variable Y_t .

Asimismo, denotaremos por $f \in \mathcal{F}_b(\mathbb{S})$ al conjunto de funciones Borel medibles y acotadas definido sobre el espacio \mathbb{S} .

Suavizamiento

A continuación, definimos de manera general lo que significa suavizamiento antes de derivar el resultado básico que forma el core de la obtención de las ecuaciones de filtraje. Considere el problema de estimación X_t usando información hasta el tiempo t , denotado por el conjunto de información $Y_{0:t} = \{Y_t, \dots, Y_0\}$, diferenciaremos entre tres casos de problemas: suavizamiento, filtraje y predicción.

Definición 4.1.6. (Suavizamiento, Filtraje y Predicción). Para índices positivos k, l , y t con $l \geq k$ denotaremos por $\pi_{v,k:l|t}$ a la distribución condicional de $X_{k:l}$ dado $Y_{0:t}$, es decir

- (a) $\pi_{v,k:l|t}$ es un kernel de transición de $(\mathbb{M})^{t+1}$ a $(\mathbb{S})^{t+1}$:
- Para cualquier conjunto $A \in \mathcal{S}^{\otimes(l-k+1)}$, $y_{0:t} \mapsto \pi_{v,k:l|t}(y_{0:t}, A)$ es una $\mathcal{M}^{\otimes(t+1)}$ función medible.
 - Para cualquier sub-secuencia $y_{0:t}$, $A \mapsto \pi_{v,k:l|t}(y_{0:t}, A)$ es una distribución de probabilidad sobre $(\mathbb{S}^{l-k+1}, \mathcal{S}^{\otimes(l-k+1)})$.
- (b) $\pi_{v,k:l|t}$ satisface para cualquier $f \in \mathcal{F}_b(\mathbb{S}^{l-k+1})$,

$$E_v[f(X_{k:l}) | Y_{0:t}] = \int \cdots \int f(x_{k:l}) \pi_{v,k:l|t}(Y_{0:t}, dx_{k:l}),$$

donde la igualdad cumple \mathbb{P}_v -casi siempre.

Para valores específicos de k y l se definen casos particulares de distribución:

- Distribución de Suavizamiento: $\pi_{v,0:t|t}$, para $t \geq 0$;
- Suavizamiento Marginal: $\pi_{v,k|t}$ para $t \geq k \geq 0$;
- Predicción : $\pi_{v,t+1|t}$ para $t \geq 0$. Par denotar la distribución inicial v usaremos la notación $\pi_{v,0|-1}$.

- p-pasos de Predicción adelante: $\pi_{v,t+p|t}$ para $t, p \geq 0$.
- Filtraje : $\pi_{v,t|t}$ para $t \geq 0$. En algunas ocasiones abreviaremos $\pi_{v,t|t}$ por $\pi_{v,t}$.

Podemos notar que $\pi_{v,k:l|t}$ es una versión de la distribución condicional de $X_{k:l}$ dado $Y_{0:t}$. Sin embargo, no es obvio que tal cantidad exista. La siguiente Proposición complementa la Definición en 4.1.6 .Mediante un enfoque constructivo se define las cantidades smoothing o suavizamiento para nuestro modelo oculto de Markov.

Proposición 4.1.7. *Considere un modelo oculto de Markov compatible con la Definición 3.3.5, sea t un entero positivo y supongamos $y_{0:t} \in \mathbb{M}^{t+1}$ es una subsecuencia tal que la función de verosimilitud $L_{v,t}(y_{0:t}) > 0$. La distribución conjunta $\pi_{v,0:t|t}$ entonces verifica*

$$\begin{aligned} \pi_{v,0:t|t}(y_{0:t}, f) &= L_{v,t}(y_{0:t})^{-1} \int \cdots \int f(x_{0:t}) \\ &\quad \times v(dx_0) g(x_0, y_0) \prod_{k=1}^t K(x_{k-1}, dx_k) g(x_k, y_k), \end{aligned} \quad (4.11)$$

para cualquier función medible y acotada f . Asimismo, para los índices $p \geq 0$,

$$\pi_{v,0:t+p|t}(y_{0:t}, f) = \int \cdots \int f(x_{0:t+p}) \times \pi_{0:t|t}(y_{0:t}, dx_{0:t}) \prod_{k=t+1}^{t+p} K(x_{k-1}, dx_k) \quad (4.12)$$

Demostración. En 4.11 se define $\pi_{v,0:t|t}$ en una manera que satisface la parte (a) de la Definición 4.1.6. Para probar la parte (b), recordemos la caracterización de esperanza condicional y consideremos una función h medible acotada definida sobre (\mathbb{M}^{t+1}) . De 1.8

$$\begin{aligned} \mathbb{E}_v[h(Y_{0:t}) f(X_{0:t})] &= \int \cdots \int h(y_{0:t}) f(x_{0:t}) v(dx_0) g(x_0, y_0) \times \\ &\quad \left[\prod_{k=1}^t K(x_{k-1}, dx_k) g(x_k, y_k) \right] \mu_t(dy_{0:t}). \end{aligned}$$

Aplicando la Definición 4.1.3 sobre $L_{v,n}$ y 4.11 para $\pi_{v,0:t|t}$ resulta

$$\begin{aligned} \mathbb{E}_v[h(Y_{0:t}) f(X_{0:t})] &= \int \cdots \int h(y_{0:t}) \pi_{v,0:t|t}(y_{0:t}, f) L_{v,t}(y_{0:t}) \mu_t(dy_{0:t}) \\ &= E_v[h(Y_{0:n}) \pi_{v,0:t|t}(Y_{0:t}, f)]. \end{aligned} \quad (4.13)$$

En la última línea usamos 4.7, por lo tanto, de la relación de igualdad en 4.13, concluimos que $\mathbb{E}_v [f(X_{0:t}) | Y_{0:t}] = \pi_{v,0:t|t}(Y_{0:t}, f)$ \mathbb{P} - casi ciertamente para cualquier función medible y acotada f definida sobre \mathbb{S}^{t-k+1} , es decir $f \in \mathcal{F}_b(\mathbb{S}^{t+1})$.

La prueba para 4.12 se realiza de manera similar, pero esta vez consideraremos la funciones $f \in \mathcal{F}_b(\mathbb{S}^{t+1+p})$ y $h \in \mathcal{F}_b(\mathbb{S}^{t+1})$. Primero aplicamos 1.8 para obtener

$$\begin{aligned} \mathbb{E}_v [h(Y_{0:t}) f(X_{0:t+p})] &= \int \cdots \int f(x_{0:t+p}) h(y_{0:t}) \times \\ &v(dx_0) g(x_0, y_0) \left[\prod_{k=1}^t K(x_{k-1}, dx_k) g(x_k, y_k) \right] \times \\ &\left[\prod_{l=t+1}^{t+p} K(x_{l-1}, dx_l) g(x_l, y_l) \right] \mu_{t+p}(dy_{0:t+p}). \end{aligned} \quad (4.14)$$

Al integrarse con respecto a la subsecuencia $y_{t+1:t+p}$, y si recordamos que $g(x, \cdot) = dQ(x, \cdot) / d\mu$ entonces se obtiene

$$\int_{\mathbb{M}} \cdots \int_{\mathbb{M}} \prod_{l=t+1}^{t+p} g(x_l, y_l) \mu_l(dy_l) = \prod_{l=t+1}^{t+p} Q(x_l, \mathbb{M}) = 1,$$

y por tanto, la tercera línea de 4.14 se reduce a $\prod_{l=t+1}^{t+p} K(x_{l-1}, dx_l) \mu_t(dy_{0:t})$. Finalmente, aplicando 4.8 y 4.11 se obtiene la siguiente relación para todo $f \in \mathcal{F}_b(\mathbb{S}^{t+1})$

$$\pi_{v,0:t|t}(y_{0:t}, dx_{0:t}) L_{v,t}(y_{0:t}) = v(dx_0) g(x_0, y_0) \prod_{k=1}^t K(x_{k-1}, dx_k) g(x_k, y_k),$$

y así 4.14 se reduce a la siguiente expresión

$$\begin{aligned} \mathbb{E}_v [h(Y_{0:t}) f(X_{0:t+p})] &= \int \cdots \int h(y_{0:t}) f(x_{0:t+p}) \times \pi_{v,0:t|t}(y_{0:t}, dx_{0:t}) L_{v,t}(y_{0:t}) \times \\ &\left[\prod_{l=t+1}^{t+p} K(x_{l-1}, dx_l) \right] \mu_t(dy_{0:t}). \end{aligned} \quad (4.15)$$

Si definimos el kernel

$$\gamma(y_{0:t}, f) \triangleq \int \cdots \int f(x_{0:t+p}) \pi_{v,0:t|t}(y_{0:t}, dx_{0:t}) \prod_{l=t+1}^{t+p} K(x_{l-1}, dx_l),$$

y usando 4.7, concluimos que 4.15 se puede describir como

$$\begin{aligned} \mathbb{E}_v [h(Y_{0:t}) f(X_{0:t+p})] &= \int \cdots \int h(y_{0:t}) \gamma(y_{0:t}, f) L_{v,t}(y_{0:t}) \mu_t(dy_{0:t}) \\ &= \mathbb{E}_v [h(Y_{0:t}) \gamma(Y_{0:t}, f)]. \end{aligned}$$

Eso prueba que para cualquier función $f \in \mathcal{F}_b(\mathbb{S}^{t+p+1})$,

$$\mathbb{E}_v [f(X_{0:t+p}) | Y_{0:t}] = \gamma(Y_{0:t}, f), \quad \mathbb{P} - \text{casi ciertamente.}$$

□

La Proposición 4.1.7 define también otros casos particulares de kernels de suavizamiento mencionados en la Definición 4.1.6, pues estos se obtienen de la marginalización.

Corolario 4.1.8. *El kernel de suavizamiento marginal $\pi_{v,k|t}$ para $0 \leq k \leq t$ es tal que para cualquier $y_{0:t} \in \mathbb{M}^{t+1}$ y $f \in \mathcal{F}_b(\mathbb{S})$,*

$$\pi_{v,k|t}(y_{0:t}, f) \triangleq \int \cdots \int f(x_k) \pi_{v,0:t|t}(y_{0:t}, dx_{0:k}), \quad (4.16)$$

donde $\pi_{v,0:t|t}$ está definida por 4.11. Asimismo, dado $y_{0:t} \in \mathbb{M}^{t+1}$, la distribución de p-pasos adelante de la predicción $\pi_{v,t+p|t}(y_{0:t}, \cdot)$ se obtiene por marginalización de la distribución conjunta $\pi_{v,0:t+p|t}(y_{0:t}, \cdot)$ con respecto a todas las variables x_k excepto el último ($k = t+p$).

Una examinación más cerca de la 4.12 junto con el uso de las ecuaciones de Chapman-Kolmogorov

$$K^{t+m}(x, A) = \int K^t(x, dy) K^m(y, A) \quad t, m \geq 0, \quad x \in \mathbb{S} \quad y \quad A \in \mathcal{S}$$

se obtiene

$$\pi_{v,t+p|t}(y_{0:t}, \cdot) = \pi_{v,t|t}(y_{0:t}, \cdot) K^p,$$

donde $\pi_{v,t|t}$ se refiere a la distribución de filtraje (distribución condicional de X_t dado $Y_{0:t}$).

Observación 4.1.9. Reemplazando la definición kernel $\pi_{v,0:t|t}$ según la ecuación 4.11 , obtenemos la siguiente expresión

$$\begin{aligned} \pi_{v,t|t} (y_{0:t}, f) &= L_{v,t} (y_{0:t})^{-1} \int \cdots \int f (x_t) v (dx_0) g (x_0, y_0) \times \\ &\quad \prod_{k=1}^t K (x_{k-1}, dx_k) g (x_k, y_k) . \end{aligned} \quad (4.17)$$

El siguiente ejemplo ilustra los conceptos descritos hasta el momento.

Ejemplo 4.1.10. Para $x_{0:2} \in \mathbb{S}^3$ e $y_{0:2} \in \mathbb{M}^3$ el kernel de transición $\pi_{v,0:2|2}$ se escribiría como

$$\begin{aligned} \pi_{v,0:2|2} (y_{0:2}, f) &= L_{v,2} (y_{0:2})^{-1} \int \int \int f (x_{0:2}) \\ &\quad \times v (dx_0) g (x_0, y_0) \prod_{k=1}^2 K (x_{k-1}, dx_k) g (x_k, y_k) , \end{aligned}$$

y su equivalente definido como medida

$$\pi_{v,0:2|2} (y_{0:2}, A) = L_{v,2} (y_{0:2})^{-1} \int 1_A (x_{0:2}) v (dx_0) g (x_0, y_0) \prod_{k=1}^2 K (x_{k-1}, dx_k) g (x_k, y_k) .$$

Por lo tanto, los marginales kernel

$$\begin{aligned} \pi_{v,2|2} (y_{0:2}, f) &= \int \int \int f (x_2) \pi_{v,0:2|2} (y_{0:2}, dx_{0:2}) , \\ \pi_{v,1|1} (y_{0:1}, f) &= \int \int f (x_1) \pi_{v,0:1|1} (y_{0:2}, dx_{0:1}) , \\ \pi_{v,0|0} (y_0, f) &= L_{v,0} (y_0)^{-1} \int f (x_0) v (dx_0) g (x_0, y_0) , \end{aligned}$$

donde por definición de $L_{v,t}$ se verifica

$$L_{v,0} (y_0) = \int v (dx_0) g (x_0, y_0) K (x_{-1}, dx_0) = \int v (dx_0) g (x_0, y_0) .$$

y por convención asumiremos

$$K (x_{-1}, dx_0) = 1.$$

4.2. Recursividad de la distribución de filtraje

En lo que sigue desarrollaremos una técnica para que un kernel de suavizamiento marginal $\pi_{v,k|t}$ pueda calcularse eficientemente para todos los valores de k en $\{0, \dots, t\}$ y para un valor específico de t . Nuestra principal herramienta será una simple representación de $\pi_{v,k|t}$ el cual se introduce a continuación. Reemplazando $\pi_{v,0:t|t}$ en 4.17 por la expresión dada en 4.11, se muestra que siempre es posible rescribir $\pi_{v,k|t}(y_{0:t}, f)$, para funciones $f \in \mathcal{F}_b(\mathbb{S})$, como

$$\begin{aligned} \pi_{v,k|t}(y_{0:t}, f) &= L_{v,t}(y_{0:t})^{-1} \int \cdots \int f(x_k) v(dx_0) g(x_0, y_0) \prod_{l=1}^k K(x_{l-1}, dx_l) g(x_l, y_l) \end{aligned}$$

donde el producto es vacío para $k = 0$ con la convención $K(x_{-1}, dx_0) g(x_0, y_0) = 1$.

Usaremos la siguiente definición para mostrar la recursividad.

Definición 4.2.1. Para $k \in \{0, \dots, t\}$, se define $h_{v,k}$ como un kernel no-negativo finito que va de $(\mathbb{M}^{k+1}, \mathcal{M}^{\otimes(k+1)})$ a $(\mathbb{S}, \mathcal{S})$ definido por

$$h_{v,k}(y_{0:k}, f) \triangleq \int \cdots \int f(x_k) v(dx_0) g(x_0, y_0) \prod_{l=1}^k K(x_{l-1}, dx_l) g(x_l, y_l). \quad (4.18)$$

Proposición 4.2.2. La medida definida por 4.18 puede obtenerse para todo $f \in \mathcal{F}_b(\mathbb{S})$ recursivamente para $k = 1, \dots, t$ según

$$h_{v,k}(y_{0:k}, f) = \int f(x') \int h_{v,k-1}(y_{0:k}, dx) K(x, dx') g(x', y_k)$$

con condición inicial

$$h_{v,0}(y_{0:k}, f) = \int f(x) g(x, y_0) v(dx).$$

Demostración. Este resultado es fácil de verificar por simple inspección

$$\begin{aligned} h_{v,k}(y_{0:k}, f) &= \int_{x_k \in \mathbb{S}} f(x_k) \int_{x_{k-1} \in \mathbb{S}} \left[\int \cdots \int_{x_0 \in \mathbb{S}, \dots, x_{k-2} \in \mathbb{S}} \right. \\ &\quad \left. v(dx_0) g(x_0, y_0) \prod_{l=1}^{k-1} K(x_{l-1}, dx_l) g(x_l, y_l) \right] K(x_{k-1}, dx_k) g(x_k, y_k), \end{aligned}$$

donde los términos entre corchetes es $h_{v,k-1}(y_{0:k-1}, dx_{k-1})$. □

Observación 4.2.3. (Notación Concisa). La relación en 4.18 puede ser rescrito como

$$\pi_{v,k|t}(y_{0:t}, f) = L_{v,t}(y_{0:t})^{-1} h_{v,k}(y_{0:k}, f). \quad (4.19)$$

Si definimos

$$\begin{aligned} w_{k|t}(y_{k+1:t}, x) \\ \triangleq \int \cdots \int K(x, dx_{k+1}) g(x_{k+1}, y_{k+1}) \prod_{l=k+2}^t K(x_{l-1}, dx_l) g(x_l, g_l), \end{aligned} \quad (4.20)$$

entonces obtenemos la siguiente caracterización.

Proposición 4.2.4. *Para todos los índices $k \in \{0, \dots, t\}$*

$$h_{v,k}(y_{0:k}, w_{k|t}) = L_{v,t}$$

y

$$h_{v,k}(1) = L_{v,k},$$

donde $L_{v,k}$ se refiere a la verosimilitud de las observaciones hasta el índice k .

Demostración. Este es resultado de 4.19 cuando $f = 1$ y la distribución marginal de suavizamiento $\pi_{v,k|t}(y_{0:t}, f)$ es una medida de probabilidad,

$$\pi_{v,k|t}(y_{0:t}, 1) \triangleq 1 = L_{v,t}(y_{0:t})^{-1} h_{v,k}(y_{0:k}, w_{k|t}).$$

Para el índice final $k = t$, $w_{k|k}$ es la función constante igual a uno y por lo tanto $h_{v,t}(1) = L_{v,t}$. \square

Recursión Normalizado

Las cantidades $h_{v,k}$ y $w_{k|t}$ así como están definidas están no normalizadas en el sentido de que se desconocen sus escalas. De otro lado, se conoce que $h_{v,k}(y_{0:k}, w_{k|t})$ es igual a $L_{v,t}$, la verosimilitud de las observaciones hasta el índice t bajo la medida de probabilidad \mathbb{P}_v .

El comportamiento a largo plazo de la verosimilitud $L_{v,t}$ tiene la propiedad asintótica. En los casos donde la recursión de la Proposición 4.2.2 es efectivamente computable, la dinámica de los valores numéricos de $h_{v,k}$ y $w_{k|t}$ es tan grande que rápidamente excede la representación computacional. La solución a este problema es el escalamiento y consiste en normalizar $h_{v,k}$ dentro de razonables cotas.

El esquema de escalamiento consistiría en reempalzar la medida $h_{v,k}$ y la función $w_{k|t}$ por versiones escaladas que verifique

$$\bar{h}_{v,k}(1) = 1 \quad y \quad \bar{h}_{v,k}(\bar{w}_{k|t}) = 1. \quad (4.21)$$

Para derivar la interpretación probabilística de $\bar{h}_{v,k}$, observemos de 4.19 y Proposición 4.2.4 para cuando $k = t$, tenemos que la distribución de filtraje $h_{v,t}$ puede describirse $[h_{v,t}(1)]^{-1} h_{v,t}$. Por lo tanto, la versión de normalización $\bar{h}_{v,k}$ de la medida $h_{v,k}$ coincide con la distribución de filtraje $\pi_{v,k}$ introducida en Definición 4.1.6. El único escalamiento que verifica 4.21 es considerar

$$\begin{aligned} \int f(x) \pi_{v,k|t}(dx) &= L_{v,t}^{-1} \int f(x) h_{v,k}(dx) w_{k|t}(x) \\ &= \int f(x) L_{v,k}^{-1} h_{v,k}(dx) L_{v,t}^{-1} L_{v,k} w_{k|t}(x) = \int f(x) \bar{h}_{v,k}(dx) \bar{w}_{k|t}(x), \end{aligned}$$

para todo $f \in \mathcal{F}_b(\mathbb{S})$. La siguiente definición resume esta conclusión, pero usando la notación $\pi_{v,k}$ en lugar de $\bar{h}_{v,k}$ pues ambos se refieren al mismo objeto - distribución de filtraje en el índice k .

Definición 4.2.5. (Variable Normalizada). Para $k \in \{0, \dots, t\}$, la medida normalizada $\bar{h}_{v,k}$ coincide con la distribución de filtraje $\pi_{v,k}$ y verifica

$$\pi_{v,k} = [h_{v,k}(1)]^{-1} h_{v,k} = L_{v,k}^{-1} h_{v,k}$$

Proposición 4.2.6. (Ecuación recursiva de filtraje). Las medidas de filtraje pueden obtenerse, para todo $f \in \mathcal{F}_b(\mathbb{S})$, de manera recursiva para $k = 1, \dots, t$ de acuerdo a las siguientes ecuaciones

$$\begin{aligned}
c_{v,k} &= \int \int \pi_{v,k-1|k-1}(dx) K(x, dx') g(x', y_k), \\
\pi_{v,k|k}(f) &= c_{v,k}^{-1} \int f(x) \int \pi_{v,k-1|k-1}(dx) K(x, dx') g(x', y_k),
\end{aligned} \tag{4.22}$$

con condición inicial

$$\begin{aligned}
c_{v,0} &= \int g_0(x) v(dx), \\
\pi_{v,0|0}(f) &= c_{v,0}^{-1} \int f(x) g(x, y_0) v(dx).
\end{aligned}$$

Demostración. Probaremos que 4.22 coincide con la medida normalizada según Definición 4.2.5. Por inducción de 4.22, es fácil de verificar que

$$\pi_{v,k|k} = \left(\prod_{l=0}^k c_{v,l} \right)^{-1} h_{v,k|k}. \tag{4.23}$$

Del hecho de que $\pi_{v,k|k}$ está normalizado,

$$\pi_{v,k|k}(1) \triangleq 1 = \left(\prod_{l=0}^k c_{v,l} \right)^{-1} h_{v,k|k}(1).$$

La Proposición 4.2.4 implica que para cualquier entero k ,

$$L_{v,k} = \prod_{l=0}^k c_{v,l}.$$

En otras palabras, $c_{v,0} = L_{v,0}$ y para índices subsecuentes $k \geq 1$, $c_{v,k} = L_{v,k}/L_{v,k-1}$. Por lo tanto, 4.23 coincide con la medida normalizada de la Definición 4.2.5. \square

Podemos concluir que el cálculo recursivo de las distribuciones de filtraje dadas por 4.22 incluye dos etapas:

- (a) Se calcula primero la distribución de predicción, $\pi_{k|k-1}$,
- (b) Luego se estima la distribución de actualización, $\pi_{k|k}$.

Por convención, denotaremos por $\pi_{v,0|_{-1}}$ a la distribución predictiva de X_0 cuando ninguna observación se encuentra disponible. Estas dos etapas se encuentran contenidas en el Lema 4.2.8.

Pero, previamente notemos de que como resultado de la Proposición 4.1.7 tenemos que la distribución de predicción está dado por

$$\pi_{v,k+1|k}(y_{0:k}, f) = \int \cdots \int f(x_{k+1}) \times \pi_{v,k|k}(y_{0:k}, dx_k) K(x_k, dx_{k+1}).$$

Sea $k \in \{0, 1, \dots, t\}$, se muestra a continuación la distribución de predicción para $k = t$.

Corolario 4.2.7. *La distribución $\pi_{v,t|t-1}$ se encuentra definido por*

$$\pi_{v,t|t-1}(y_{0:t-1}, dx_t) = (\pi_{v,t-1|t-1}K)(y_{0:t-1}, dx_t), \quad (4.24)$$

para $y_{0:t}$ tal que $L_{v,t}(y_{0:t}) > 0$.

Demostración. Aplicando 4.12 para $p = 1$ tenemos

$$\pi_{v,0:t|t-1}(y_{0:t-1}, dx_{0:t}) = \int 1_A(x_{0:t}) \pi_{v,0:t-1|t-1}(y_{0:t-1}, dx_{0:t-1}) K(x_{t-1}, dx_t).$$

Luego, para una función f medible y acotada, tendremos el kernel marginal

$$\begin{aligned} \pi_{v,t|t-1}(y_{0:t-1}, f) &= \int \cdots \int f(x_t) \pi_{v,0:t-1|t-1}(y_{0:t-1}, dx_{0:t-1}) K(x_{t-1}, dx_t). \\ &= \int \int f(x_t) \pi_{v,t-1|t-1}(y_{0:t-1}, dx_{t-1}) K(x_{t-1}, dx_t). \end{aligned} \quad (4.25)$$

La última igualdad se sigue de que para cualquier $A \in \mathcal{S}$ se tiene

$$\int_{\mathbb{S}^t} \pi_{v,0:t-1|t-1}(y_{0:t-1}, dx_{0:t-1}) K(x_{t-1}, A) = \int_{\mathbb{S}} \pi_{v,t-1|t-1}(y_{0:t-1}, dx_{t-1}) K(x_{t-1}, A),$$

gracias a 4.16.

Si rescribimos 4.25 integrando con respecto a x_{t-1} sobre $K(x_{t-1}, A)$ y considerando las noción de medidas positivas sobre un kernel de transición, ver 3.12, entonces 4.25 se rescribe de la siguiente forma

$$\pi_{v,t|t-1}(y_{0:t-1}, f) = \int f(x_t) (\pi_{v,t-1|t-1}K)(y_{0:t-1}, dx_t),$$

Eso prueba

$$\pi_{v,t|t-1}(y_{0:t-1}, dx_t) = (\pi_{v,t-1|t-1}K)(y_{0:t-1}, dx_t)$$

□

Gracias a este resultado obtenemos nuestro sistema ecuaciones de filtraje.

Lema 4.2.8. *La recursión en 4.22 puede describirse como la ecuación de actualización dado por 4.26 y la de predicción dado por 4.27*

$$\begin{aligned} c_{v,k} &= \int \int \pi_{v,k|k-1}(y_{0:k}, dx) g(x, y_k), \\ \pi_{v,k|k}(y_{0:k}, f) &= c_{v,k}^{-1} \int f(x) \int \pi_{v,k|k-1}(y_{0:k}, dx) g(x, y_k). \end{aligned} \quad (4.26)$$

$$\pi_{v,k+1|k} = \pi_{v,k|k} K, \quad (4.27)$$

para $k = 0, 1, \dots, t$ y $f \in \mathcal{F}_b(\mathbb{S})$.

Estimador óptimo

Necesitaremos de un concepto claro de lo que constituye para que un estimador sea estadísticamente óptimo en nuestro algoritmo de filtraje estocástico. El concepto estadístico que usaremos será el criterio de error cuadrático medio y está en línea con la Definición 4.0.1.

Definición 4.2.9. (Criterio Óptimo). Sea $\hat{X}_{k|k-1}$ el estimador de X_k basado en el conjunto de información $Y_{1:k-1}$, el cual induce el error de estimación $e_k = X_k - \hat{X}_{k|k-1}$. El estimador que minimiza el error cuadrático medio, **ECM**

$$\mathbf{ECM}(\hat{X}_{k|k-1}) \equiv E [X_k - \hat{X}_{k|k-1}]^2,$$

resulta ser el estimador óptimo con respecto a cualquier función cuadrática de e_k .

Del Teorema 4.0.4, se concluye el siguiente criterio de optimización del error.

Teorema 4.2.10. (*Estimador Óptimo*). Dado el conjunto de información $Y_{1:k-1}$, el estimador óptimo $X_{k|k-1}$ de X_k entre todos los estimadores $\hat{X}_{k|k-1}$ es el valor esperado de X_k condicionado sobre la información disponible hasta $k-1$. Entonces, el mínimo error cuadrático medio MECM está dado por $X_{k|k-1} = E[X_k | Y_{1:k-1}]$. El correspondiente ECM está dado por

$$\mathbf{ECM}(\hat{X}_{k|k-1}) = E [X_k - E[X_k | Y_{1:k-1}]]^2 = \mathit{Cov}[X_k | Y_{1:k-1}], \quad (4.28)$$

es decir, el error cuadrático medio coincide con la matriz covarianza.

Observación 4.2.11. (Notación). En el problema de estimación de X_k usando la información hasta el tiempo s , es decir el conjunto de información $Y_{1:s} = \{Y_s, \dots, Y_2, Y_1\}$, denotaremos por conveniencia $E[X_k | Y_{1:s}] \equiv m_{k|s}$, y para el segundo momento condicional usaremos $\mathit{Cov}[X_k | Y_{1:s}] \equiv C_{k|s}$.

En la siguiente sección se presenta el algoritmo de Filtro de Kalman bajo el supuesto Gaussiano de los errores y se interpreta al filtro de Kalman como el filtro óptimo en el sentido del estimador mínimo cuadrático medio. Cabe mencionar, que también es posible obtener el filtro de Kalman aún si no se verifica el supuesto de distribución Gaussiana. En este último caso hablaremos el filtro de Kalman como el filtro óptimo en el sentido de estimador lineal mínimo cuadrático medio.

4.3. Filtro de Kalman en el sistema lineal Markov Gaussiano

Para modelos espacio-estado lineales Gaussianos, la evaluación recursiva de $\pi_{v,k|t}$ para cierto valor fijo de k es factible (ver, Anderson y Moore, 1979, Capítulo 7). Los modelos espacio-estado lineales y Gaussianos constituyen un caso particular, pues $\pi_{v,k|t}$ resulta ser enteramente determinado por su media y varianza.

El principal propósito de esta sección es mostrar que el espacio estado del modelo de Schwartz y Smith es un modelo oculto de Markov. Fijemos los espacios $\mathbb{S} = \mathbb{R}^d$ y $\mathbb{M} = \mathbb{R}^p$ para algún entero d, p . Se definen los procesos X_t e Y_t sobre \mathbb{R}^d y \mathbb{R}^p , respectivamente y se asume que el kernel de transición T sigue la dinámica especificada por el Lema 3.1.1 ; es decir

$$X_t = c + GX_{t-1} + \omega_t, \quad (4.29)$$

$$Y_t = d + F'X_t + v_t \quad (4.30)$$

donde las matrices c, d, G y F son constantes con parámetros conocidos y dimensiones apropiadas, $\{\omega_t\}_{t \geq 0}$ y $\{v_t\}_{t \geq 0}$ son procesos ruidos blancos, ver Observación 3.1.2, con distribución Gaussiana

$$\omega_t \sim \mathcal{N}(0, \mathbf{W}) \quad \text{y} \quad v_t \sim \mathcal{N}(0, \mathbf{V}). \quad (4.31)$$

Del supuesto en la 4.31, las funciones densidades del kernel de transición de X_t e Y_t son Gaussianas¹ y por tanto se verifican las siguientes tres propiedades.

Observación 4.3.1. Propiedades del Modelo Lineal Markov-Gausiano.

- (a) Sea ρ la función densidad asociada a ω . Del ejemplo 3.3.4, el kernel de transición de Markov K está dado por

$$\begin{aligned} K(x, A) &= \int 1_A(x') \rho(x' - c - Gx) dx', \\ &= \int 1_A(x') (2\pi)^{-\frac{d}{2}} \det(\mathbf{W})^{-\frac{1}{2}} \times \\ &\quad \exp\left\{-\frac{1}{2}(x' - c - Gx)^\top \mathbf{W}^{-1}(x' - c - Gx)\right\} dx'. \end{aligned} \quad (4.32)$$

- (b) Sea g la función densidad de v , entonces g está dado por

$$\begin{aligned} g(x, y) &= (2\pi)^{-\frac{p}{2}} \det(\mathbf{V})^{-\frac{1}{2}} \\ &\quad \times \exp\left\{-\frac{1}{2}(y - d - Fx)^\top \mathbf{V}^{-1}(y - d - Fx)\right\}, \end{aligned}$$

con kernel de transición del proceso Y dado por

$$Q(x, B) = \int 1_B(y) g(x, y) dy. \quad (4.33)$$

¹Recordemos que un vector aleatorio $(Y_1, \dots, Y_p) \in \mathbb{R}^p$ es normal si y solamente si $c_1 Y_1 + \dots + c_p Y_p$ es normal para ciertos valores de $c_1, \dots, c_p \in \mathbb{R}$

(c) Se asume que $X_0 \sim \mathcal{N}(m_0, \mathbf{C}_0)$

$$\begin{aligned} \pi_{0|0} &= (2\pi)^{-\frac{d}{2}} \det(\mathbf{C}_0)^{-\frac{1}{2}} \times \\ &\quad \exp\left\{-\frac{1}{2}(x_0 - m_0)^\top \mathbf{C}_0^{-1}(x_0 - m_0)\right\} \end{aligned}$$

Estos tres supuestos son medulares para mostrar que la distribución de filtraje es Gaussiano en la Proposición 4.3.4.

En el siguiente Lema mostraremos que nuestro modelo lineal es un modelo oculto de Markov.

Lema 4.3.2. *La representación Espacio Estado dado en el Lema 3.1.1 es un modelo oculto de Markov.*

Demostración. Hemos mostrado que X_t es un proceso de Markov en el Ejemplo 3.3.4. Entonces basta probar

$$E[f(X_{t+1}, Y_{t+1}) | X_{0:t}, Y_{0:t}] = E[f(X_{t+1}, Y_{t+1}) | X_t]$$

Para ver esto comenzaremos con calcular

$$\begin{aligned} E[f(X_{t+1}, Y_{t+1}) | X_{0:t}, Y_{0:t}] \\ = E\left[f\left(X_{t+1}, d + F'X_t + v_t\right) | X_{0:t}, Y_{0:t}\right] \end{aligned} \quad (4.34)$$

y considerando la propiedad de la esperanza condicional, la expresión de 4.34 se puede describir como

$$E\left[E\left[f\left(X_{t+1}, d + F'X_t + v_t\right) | X_{0:t+1}, Y_{0:t}\right] | X_{0:t}, Y_{0:t}\right].$$

De la propiedad de Markov de X_t y del supuesto de independencia de V_{t+1} con respecto a (X_0, \dots, X_{t+1}) e (Y_0, \dots, Y_t) , 4.34 se reduce a la siguiente forma

$$\begin{aligned} E[f(X_{t+1}, Y_{t+1}) | X_{0:t}, Y_{0:t}] &= E\left[E\left[f\left(x, d_t + F'X_t + v_t\right) | X_{t+1} = x\right] | X_{0:t}, Y_{0:t}\right] \\ &= E\left[\int f(X_{t+1}, y) Q(X_{t+1}, dy) | X_{0:t}, Y_{0:t}\right] \\ &= \int f(x, y) Q(x, dy) K(X_t, dx) \\ &= E[f(X_{t+1}, Y_{t+1}) | X_t]. \end{aligned}$$

□

Una característica única del modelo espacio estado Gaussiano es que al ser los estados $X_{0:t}$ y las observaciones $Y_{0:t}$ Gaussianos conjuntamente multivariado para cualquier valor de t , las distribuciones de suavizamiento son también Gaussianas. Por tanto, las distribuciones de suavizamiento y sus marginales están determinadas por su vector media y matriz de covarianza; es decir el estimador predictivo $\pi_{k|k-1}$ y el estimador de filtraje π_k estarían dados por

$$\pi_{k|k-1} = \mathcal{N}(\hat{m}_{k|k-1}, \hat{C}_{k|k-1}), \quad (4.35)$$

$$\pi_{k|k} = \mathcal{N}(\hat{m}_{k|k}, \hat{C}_{k|k}), \quad (4.36)$$

sus respectivas medias y matrices de covarianza.

Observación 4.3.3. Hasta el momento, para $k \in \{0, 1, \dots, t\}$ hemos denotado por $\pi_{k|k}$ a la distribución de filtraje. Para evitar toda ambigüedad, usaremos la notación $\hat{m}_{k|k}$ y $\hat{C}_{k|k}$ para denotar los primeros dos momentos de la distribución de filtraje en nuestro modelo espacio estado lineal Gaussiano.

Proposición 4.3.4. *Sea el modelo espacio estado lineal Gaussiano, como la del Lema 3.1.1. Si la distribución de $\pi_{v,t-1|t-1}$ es Gaussiana, i.e. $\mathcal{N}(m_{t-1}, C_{t-1})$, entonces la distribución de filtraje $\pi_{v,t|t}$ es siempre Gaussiana.*

Demostración. Parte 1 : Ecuación de Predicción. Primero probaremos que la distribución predictiva $\pi_{v,t|t-1}$ es Gaussiana. De 4.27 se obtiene

$$\pi_{v,t|t-1}(y_{0:t-1}, x_t) = \int \underbrace{K(x_{t-1}, x_t)}_{\sim \mathcal{N}(c+Gx_{t-1}, \mathbf{W})} \times \underbrace{\pi_{v,t-1|t-1}(y_{0:t-1}, x_{t-1})}_{\sim \mathcal{N}(m_0, C_0)} dx_{t-1}, \quad (4.37)$$

Rescribiendo, en su forma Gaussiana, ver observación 4.3.1 obtenemos

$$\begin{aligned} \pi_{v,t|t-1}(y_{0:t-1}, x_t) &= \eta \int \exp \left\{ -\frac{1}{2} (x_t - c - Gx_{t-1})^\top \mathbf{W}^{-1} (x_t - c - Gx_{t-1}) \right\} \times \\ &\quad \exp \left\{ -\frac{1}{2} (x_{t-1} - m_{t-1})^\top C_{t-1}^{-1} (x_{t-1} - m_{t-1}) \right\} dx_{t-1}, \end{aligned} \quad (4.38)$$

donde η engloba los demás términos faltantes. En una forma más resumida, rescribimos 4.38

$$\pi_{v,t|t-1}(y_{0:t-1}, x_t) = \eta \int \exp \{-L_t\} dx_{t-1} \quad (4.39)$$

donde

$$L_t = \frac{1}{2} (x_t - c - Gx_{t-1})^\top \mathbf{W}^{-1} (x_t - c - Gx_{t-1}) + \frac{1}{2} (x_{t-1} - m_{t-1})^\top C_{t-1}^{-1} (x_{t-1} - m_{t-1}),$$

Para resolver la integral en 4.39 , reescribimos a L_t de la siguiente manera

$$L_t = L_t(x_{t-1}, x_t) + L_t(x_t), \quad (4.40)$$

tal que $L_t(x_{t-1}, x_t)$ contiene todos los términos cruzados que contiene a x_{t-1} mientras que $L_t(x_t)$ solo los componentes cuadráticos en x_t :

$$\begin{aligned} \pi_{v,t|t-1} &= \eta \int \exp \{-L_t(x_{t-1}, x_t) - L_t(x_t)\} dx_{t-1} \\ &= \eta \exp \{-L_t(x_t)\} \int \exp \{-L_t(x_{t-1}, x_t)\} dx_{t-1}. \end{aligned} \quad (4.41)$$

Además, elegiremos $L_t(x_{t-1}, x_t)$ tal que el valor de la integral en 4.41 no dependa de x_t . La integral será una constante relativo al problema de estimar la distribución sobre x_t ,

$$\pi_{v,t|t-1} = \eta \exp \{-L_t(x_t)\} \quad (4.42)$$

En lo que resta, el problema se centra en encontrar una función $L_t(x_{t-1}, x_t)$ cuadrática en x_{t-1} . Los coeficientes de esta función se calculan tomando las primeras dos derivadas de L_t , para mayor detalle de los cálculos ver Thrun (2006, pag.43).

$$L_t(x_{t-1}, x_t) = \frac{1}{2} \left(x_{t-1} - \Psi_t \left[G^\top \mathbf{W}^{-1} x_t + C_{t-1}^{-1} m_{t-1} \right] \right)^\top \Psi_t^{-1} \left(x_{t-1} - \Psi_t \left[G^\top \mathbf{W}^{-1} x_t + C_{t-1}^{-1} m_{t-1} \right] \right), \quad (4.43)$$

donde $\Psi_t^{-1} := G^\top \mathbf{W}^{-1} G + C_{t-1}^{-1}$ define la curvatura de $L_t(x_{t-1}, x_t)$. Cabe mencionar que $L_t(x_{t-1}, x_t)$ es una forma cuadrática que comunmente aparece en el exponente de una distribución normal; es decir

$$\int \det(2\pi\Psi_t)^{-\frac{1}{2}} \exp \{-L_t(x_{t-1}, x_t)\} dx_{t-1} = 1, \quad (4.44)$$

de esto se sigue que

$$\int \exp\{-L_t(x_{t-1}, x_t)\} dx_{t-1} = \det(2\pi\Psi_t)^{\frac{1}{2}}. \quad (4.45)$$

Lo importante de esto es notar que el valor de esta integral es independiente de x_t , el cual es nuestra principal variable de interés. Por lo tanto, para calcular la distribución sobre x_t , esta integral es constante; por lo tanto, se le considera dentro del factor η y así se justifica 4.42.

Luego, $L_t(x_t)$ se obtiene de la ecuación 4.40,

$$\begin{aligned} L_t(x_t) &= L_t - L_t(x_{t-1}, x_t) \\ &= \frac{1}{2}(x_t)^\top \mathbf{W}^{-1}x_t + \frac{1}{2}m_{t-1}^\top C_{t-1}^{-1}m_{t-1} - \frac{1}{2}[G^\top \mathbf{W}^{-1}x_t + C_{t-1}^{-1}m_{t-1}]^\top \times \\ &\quad (G^\top \mathbf{W}^{-1}G + C_{t-1}^{-1})^{-1} [G^\top \mathbf{W}^{-1}x_t + C_{t-1}^{-1}m_{t-1}] \end{aligned}$$

$L_t(x_t)$ es cuadrática en x_t , lo cual muestra que $\pi_{v,t|t-1}$ es Gaussiana : donde la media y varianza de $\pi_{v,t|t-1}$ coincide con la primera y curvatura de $L_t(x_t)$.

$$\begin{aligned} \hat{m}_{t|t-1} &= c + G \hat{m}_{t-1|t-1} \\ \hat{C}_{t|t-1} &= G\hat{C}_{t-1|t-1}G^\top + \mathbf{W}. \end{aligned}$$

Parte 2 : Ecuación de Actualización. Analizaremos la distribución de filtraje. Del Lema 4.2.8

$$\begin{aligned} \pi_{v,t|t}(y_{0:t}, x_t) &= \underbrace{c_t^{-1} g(x_t, y_t)}_{\sim \mathcal{N}(d+Fx_t, \mathbf{V})} \times \underbrace{\pi_{v,t-1|t-1}(y_{0:t-1}, x_{t-1})}_{\sim \mathcal{N}(\hat{m}_{t|t-1}, \hat{C}_{t|t-1})} dx_{t-1}, \\ &= \eta \exp\{-J_t\}, \end{aligned}$$

donde de la ecuación 4.33 J_t está determinada por

$$J_t := \frac{1}{2}(y_t - d - Fx_t)^\top \mathbf{V}^{-1}(y_t - d - Fx_t) + \frac{1}{2}(x_t - \hat{m}_{t|t-1})^\top \hat{C}_{t|t-1}(x_t - \hat{m}_{t|t-1})$$

De la observación de que J_t es cuadrática en x_t es suficiente para probar que $\pi_{v,t|t}(y_{0:t}, x_t)$ es Gaussiana. La media y covarianza de $\pi_{v,t|t}(y_{0:t}, x_t)$ se obtienen de la primera y segunda derivada de J_t :

$$\begin{aligned}\hat{m}_{t|t} &= \hat{m}_{t|t-1} + \hat{C}_{t|t-1} F^\top (F \hat{C}_{t|t-1} F^\top + \mathbf{V})^{-1} (Y_t - d - F \hat{m}_{t|t-1}) \\ \hat{C}_{t|t} &= \hat{C}_{t|t-1} - \hat{C}_{t|t-1} F^\top (F \hat{C}_{t|t-1} F^\top + \mathbf{V})^{-1} F \hat{C}_{t|t-1}.\end{aligned}$$

□

Observación 4.3.5. Recordemos que en el Lema 3.1.1 habíamos definido la dinámica de X_k para $k \in \{0, \dots, t\}$:

$$X_k = c + G X_{k-1} + \omega_k. \quad (4.46)$$

El siguiente Lema elemental es instrumental para calcular el estimador del estado predictivo y de filtraje de las ecuaciones 4.35 y 4.36.

Proposición 4.3.6. (*Modelo Lineal Gaussiano Condicionado*). Sean X y v dos vectores aleatorios Gaussianos independientes con $E[X] = m_X$, $Cov(X) = C_X$, y asuma que $E[v] = 0$. Considere el modelo

$$Y_k = d + F' X_k + v_k \quad (4.47)$$

donde F es una matriz determinística de apropiadas dimensiones. Sea $Cov(v) = \mathbf{V}$, asumimos que $FC_X F^\top + \mathbf{V}$ es una matriz de rango completo. Entonces, la distribución de X condicionado sobre Y es Gaussiana con media y covarianza

$$\begin{aligned}E[X | Y] &= E[X] + Cov(X, Y) \{Cov(Y)\}^{-1} (Y - E[Y]) \\ &= m_X + C_X F^\top \{FC_X F^\top + \mathbf{V}\}^{-1} (Y - d - F m_X) \quad y\end{aligned} \quad (4.48)$$

$$\begin{aligned}Cov(X | Y) &= Cov(X - E[X | Y]) = E[(X - E[X | Y]) X^\top] \\ &= Cov(X) - Cov(X, Y) \{Cov(Y)\}^{-1} Cov(X, Y)^\top \\ &= C_X - C_X F^\top \{FC_X F^\top + \mathbf{V}\}^{-1} F C_X.\end{aligned} \quad (4.49)$$

Demostración. Denotemos por \hat{m} el lado derecho de la ecuación 4.48. Entonces

$$X - \hat{m} = X - E[X] + Cov(X, Y) \{Cov(Y)\}^{-1} (Y - E[Y])$$

el cual implica que

$$Cov(X - \hat{m}, Y) = Cov(X, Y) - Cov(X, Y) \{Cov(Y)\}^{-1} Cov(Y) = 0. \quad (4.50)$$

Los vectores aleatorios Y e $X - \hat{m}$ por tanto son conjuntamente Gaussianos (como transformaciones lineales de vector aleatorio multivariado Gaussiano) y no correlacionado. Por tanto, Y e $X - \hat{m}$ son también independientes. Escribiendo

$$X = \hat{m} + (X - \hat{m}),$$

donde \hat{m} es $\sigma(Y)$ medible (como combinación lineal de los componentes de Y) y $X - \hat{m}$ es independiente de \hat{m} , es fácil de verificar que $\hat{m} = E[X | Y]$ y que,

$$Cov(X | Y) \triangleq Cov[(X - \hat{m})(X - \hat{m})^T | Y] = Cov(X - \hat{m}).$$

Finalmente, la ecuación 4.49 se obtiene de notar

$$Cov(X - \hat{m}) = E[(X - \hat{m})(X - \hat{m})^T] = E[(X - \hat{m})X^T],$$

usando la ecuación 4.50 y el hecho de que \hat{m} es una transformación lineal de Y . Las segundas líneas de las ecuaciones 4.48 y 4.49 se siguen de la estructura lineal de la ecuación 4.47. \square

Según 4.28, la Proposición 4.3.6 señala que el estimador óptimo de X condicionado sobre Y está dado por $E[X | Y]$ con $Cov(X | Y)$ caracterizando el error medio cuadrático, ECM, de esta estimación:

$$E[(X - E[X | Y])(X - E[X | Y])^T] = Cov(X | Y). \quad (4.51)$$

La Proposición 4.3.6 también nos dice de que mientras $\hat{m}_{k|k-1}$ o $\hat{m}_{k|k}$ dependen de las observaciones, las matrices de covarianzas, $\hat{C}_{k|k-1}$ y $\hat{C}_{k|k}$ están completamente determinadas por los parámetros del modelo. Nuestro resultado principal simplemente consiste en aplicar la fórmula de la Proposición 4.3.6 en modelos lineales Gaussianos para obtener ecuaciones recursivas equivalentes a la ecuación 4.27 en términos de los parámetros de nuestro modelo.

Veremos que al deducir las ecuaciones del filtro de Kalman, se comienza con el objetivo de encontrar una ecuación que calcule una estimación del estado $\hat{m}_{k|k}$ como una combinación lineal de la estimación $\hat{m}_{k|k-1}$ y un promedio ponderado de la diferencia entre la actual medida Y_k y una predicción de la medida $Y_{k|k-1}$ como se muestra a continuación.

Proposición 4.3.7. (*Filtraje en Modelos Lineales Gaussianos*). Sea el modelo del Lema 3.1.1. La media y matrices de covarianzas predictivas y de filtraje del sistema dado en 3.1.1 pueden actualizarse recursivamente como sigue, para $k \geq 0$.

Predicción:

$$\hat{m}_{k|k-1} = c + G \hat{m}_{k-1|k-1} \quad (4.52)$$

$$\hat{C}_{k|k-1} = G \hat{C}_{k-1|k-1} G^T + \mathbf{W}. \quad (4.53)$$

Filtraje:

$$\hat{m}_{k|k} = \hat{m}_{k|k-1} + \hat{C}_{k|k-1} F^T (F \hat{C}_{k|k-1} F^T + \mathbf{V})^{-1} (Y_k - d - F \hat{m}_{k|k-1}) \quad (4.54)$$

$$\hat{C}_{k|k} = \hat{C}_{k|k-1} - \hat{C}_{k|k-1} F^T (F \hat{C}_{k|k-1} F^T + \mathbf{V})^{-1} F \hat{C}_{k|k-1} \quad (4.55)$$

con la convenciones $\hat{m}_{0|-1} = 0$ y $\hat{C}_{0|-1} = C_v$.

Demostración. Predicción: Para encontrar la predicción un paso adelante, tomamos la esperanza condicional de las variables de estado de la ecuación 4.46 dada la información hasta el tiempo $t - 1$, $k = 0, 1, \dots, t$.

$$\begin{aligned} \hat{m}_{k|k-1} &\equiv E[X_k | Y_{1:k-1}] = E[E[X_k | X_{k-1}, Y_{1:k-1}] | Y_{1:k-1}] \\ &= c + G \hat{m}_{k-1|k-1} \end{aligned} \quad (4.56)$$

La correspondiente matriz varianza-covarianza para las variables de estado están dados por

$$\begin{aligned} \hat{C}_{k|k-1} &\equiv E[(X_k - E[X_k | Y_{1:k-1}]) (X_k - E[X_k | Y_{1:k-1}])^T | Y_{1:k-1}] \\ &= E[(X_k - \hat{m}_{k|k-1}) (X_k - \hat{m}_{k|k-1})^T | Y_{1:k-1}] \\ &= E\left[(c + G X_{k-1} + \omega_k - c - G \hat{m}_{k-1|k-1}) \right. \\ &\quad \left. (c + X_{k-1}^T G^T + \omega_k - c - \hat{m}_{k-1|k-1}^T G^T) | Y_{1:k-1}\right] \end{aligned} \quad (4.57)$$

$$\begin{aligned} &= GE \left[(X_{k-1} - \hat{m}_{k-1|k-1}) (X_{k-1} - \hat{m}_{k-1|k-1})^T | Y_{1:k-1} \right] G^T + E[\omega_k \omega_k^T | Y_{1:k-1}] \\ &= G \hat{C}_{k-1|k-1} G^T + \mathbf{W}. \end{aligned} \quad (4.58)$$

Las expresiones en 4.56 y 4.58 se conocen como ecuaciones de predicción del Filtro de Kalman.

Como un resultado intermedio, definimos el error de predicción denotado por ϵ_k , usando la data de las observaciones Y_k , como

$$\epsilon_k \equiv Y_k - Y_{k|k-1} = Y_k - E[Y_k | Y_{1:k-1}] = Y_k - d - F\hat{m}_{k|k-1}, \quad (4.59)$$

y calculemos la matriz covarianza de ϵ_k como

$$\begin{aligned} Z_{k|k-1} &\equiv Cov[\epsilon_k | Y_{1:k-1}] \\ &= E[(Y_k - E[Y_k | Y_{1:k-1}])(Y_k - E[Y_k | Y_{1:k-1}])^\top | Y_{1:k-1}] \\ &= E\left[\left(d + F'X_k + v_k - d - F\hat{m}_{k|k-1}\right)\left(d + X_k^\top F^\top + v_k - d - \hat{m}_{k|k-1}^\top F^\top\right) | Y_{1:k-1}\right] \\ &= FE\left[\left(X_{k-1} - \hat{m}_{k-1|k-1}\right)\left(X_{k-1} - \hat{m}_{k-1|k-1}\right)^\top | Y_{1:k-1}\right]F^\top + E[v_k v_k^\top | Y_{1:k-1}] \\ &= F\hat{C}_{k|k-1}F^\top + \mathbf{V}. \end{aligned} \quad (4.60)$$

Actualización de la medición: Para encontrar $\hat{m}_{k|k}$ y $\hat{C}_{k|k}$ en términos de $\hat{m}_{k|k-1}$ y $\hat{C}_{k|k-1}$, condicionamos la ecuación 4.47, $Y_k = d + F'X_k + v_k$ con respecto a $Y_{1:k-1}$:

$$\begin{aligned} E[Y_k | Y_{1:k-1}] &= E\left[E\left[d + F'X_k + v_k | X_k, Y_{1:k-1}\right] | Y_{1:k-1}\right] \\ &= d + E\left[F'X_k | Y_{1:k-1}\right] \\ &= d + F\hat{m}_{k|k-1}. \end{aligned}$$

$$Var[Y_k | Y_{1:k-1}] = Cov[\epsilon_k | Y_{1:k-1}] = Z_{k|k-1}.$$

Del hecho de que v_k e Y_{k-1} son independientes tenemos que $E[X_k | Y_{1:k-1}]$ y $E[Y_k | Y_{1:k-1}]$ son conjuntamente Gaussianos con media y matriz varianza-covarianza

$$\begin{bmatrix} \hat{m}_{k|k-1} \\ d + F\hat{m}_{k|k-1} \end{bmatrix}, \quad \begin{bmatrix} \hat{C}_{k|k-1} & \hat{C}_{k|k-1}F^\top \\ F\hat{C}_{k|k-1} & Z_{k|k-1} \end{bmatrix}, \quad (4.61)$$

donde para $Y_{k|k-1} \equiv E[Y_k | Y_{1:k-1}]$ la covarianza se obtiene de

$$\begin{aligned} E\left[\left(X_k - \hat{m}_{k|k-1}\right)\left(Y_k - Y_{k|k-1}\right) | Y_{1:k-1}\right] &= E\left[\left(X_k - \hat{m}_{k|k-1}\right)\left(X_k - \hat{m}_{k|k-1}\right)^\top | Y_{1:k-1}\right]F^\top \\ &= \hat{C}_{k|k-1}F^\top \end{aligned}$$

Para la derivación de las ecuaciones de filtraje es resultado directo de aplicar la Proposición 4.3.6. Solo basta observar la propiedad de independencia de Markov: la media y covarianza de $E[(X_k | Y_{1:k-1}) | (Y_k | Y_{1:k-1})]$ es equivalente a $E[(X_k | Y_{1:k})]$:

$$X_k \sim \mathcal{N} \left(\hat{m}_{k|k-1} + \hat{C}_{k|k-1} F^\top \left(F \hat{C}_{k|k-1} F^\top + \mathbf{V} \right)^{-1} \left(Y_k - d - F \hat{m}_{k|k-1} \right), \right. \\ \left. \hat{C}_{k|k-1} - \hat{C}_{k|k-1} F^\top \left(F \hat{C}_{k|k-1} F^\top + \mathbf{V} \right)^{-1} F \hat{C}_{k|k-1} \right).$$

Actualización del tiempo: Si incrementamos el tiempo, usando $X_{k+1} = c + GX_k + \omega_k$, y condicionamos sobre $Y_{1:k}$ conseguimos

$$\begin{aligned} \hat{m}_{k+1|k} &= c + G \hat{m}_{k|k} \quad \text{y} \\ \hat{C}_{k+1|k} &= G \hat{C}_{k|k} G^\top + \mathbf{W}. \end{aligned} \tag{4.62}$$

La ecuación de actualización de la medición y del tiempo nos da la solución recursiva que comienza con la media $\hat{m}_{0|-1} = m_0$ y covarianza $\hat{C}_{0|-1} = C_0$. Aplicamos la ecuación de actualización

$$\begin{aligned} \hat{m}_{k|k} &= \hat{m}_{k|k-1} + \hat{C}_{k|k-1} F^\top \left(F \hat{C}_{k|k-1} F^\top + \mathbf{V} \right)^{-1} \left(Y_k - d - F \hat{m}_{k|k-1} \right) \\ \hat{C}_{k|k} &= \hat{C}_{k|k-1} - \hat{C}_{k|k-1} F^\top \left(F \hat{C}_{k|k-1} F^\top + \mathbf{V} \right)^{-1} F \hat{C}_{k|k-1}, \end{aligned}$$

y conseguimos $\hat{m}_{0|0}$ y $\hat{C}_{0|0}$, luego incrementamos el tiempo, ver la ecuación 4.62 y así conseguimos $\hat{m}_{1|0}$ y $\hat{C}_{1|0}$, este proceso se repite recursivamente. \square

Debemos notar de la Proposición 4.3.7 que si combinamos la ecuación de predicción con la de filtraje obtenemos

$$\hat{C}_{k+1|k} = G \hat{C}_{k|k-1} G^\top + \mathbf{W} - G \hat{C}_{k|k-1} F^\top \left(F \hat{C}_{k|k-1} F^\top + \mathbf{V} \right)^{-1} F \hat{C}_{k|k-1} G^\top, \tag{4.63}$$

el cual se le conoce como la **recursión de Riccati**, con condición inicial $\hat{C}_{0|-1} = C_0$. La matriz de covarianza $\hat{C}_{k|k-1}$ está completamente determinadas por los parámetros del modelo, y por lo tanto, podemos calcular la covarianza del error de estimación $\hat{C}_{k|k-1}$ sin considerar ninguna observación de la data Y_k . Este hecho es importante porque, como se muestra en Anderson y Moore (1979) y Harvey (1994), la recursión de Riccati cuando los parámetros con constantes converge a un valor de estado de equilibrio (constante), \hat{C}_x donde de 4.63 \hat{C}_x es solución de esta ecuación:

$$\hat{C}_x = G \hat{C}_x G^\top + \mathbf{W} - G \hat{C}_x F^\top \left(F \hat{C}_x F^\top + \mathbf{V} \right)^{-1} F \hat{C}_x G^\top.$$

Gracias a este hecho, el algoritmo de Filtro de Kalman es más eficiente cuando se inicia desde \hat{C}_x .

De otro lado, las ecuaciones recursivas del Filtro de Kalman son representadas por una matriz K_k la cual es escogida para ser la ganancia o factor de mezcla que minimiza la covarianza del error $\hat{C}_{k|k}$. Una de las formas de llegar a la minimización de K_k es :

$$K_k = \underbrace{\hat{C}_{k|k-1}F^T}_{\text{Error Estimación}} \times \underbrace{\left(F\hat{C}_{k|k-1}F^T + \mathbf{V}\right)^{-1}}_{\text{Error Estimación+Error Medición}}. \quad (4.64)$$

Podemos describir las ecuaciones de actualización de la Proposición 4.3.7 :

$$\hat{m}_{k|k} = \hat{m}_{k|k-1} + K_k \left(Y_k - d - F\hat{m}_{k|k-1} \right) \quad (4.65)$$

$$\begin{aligned} \hat{C}_{k|k} &= E \left[\left(X_k - \hat{m}_{k|k} \right) \left(X_k - \hat{m}_{k|k} \right)^T \mid Y_{1:k} \right] \\ &= \hat{C}_{k|k-1} - K_k F \hat{C}_{k|k-1} \\ &= \left(\mathbf{I} - K_k F \right) \hat{C}_{k|k-1}. \end{aligned} \quad (4.66)$$

Observando las ecuaciones 4.64 y 4.65, notamos las siguientes relaciones:

- Si V tiende a cero, entonces K_k tenderá a uno (se ajusta principalmente con la actualización de la medición Y_k y le dará menos peso a la medición estimada $Y_{k|k-1}$, es decir $d - F\hat{m}_{k|k-1}$).
- Si V es lo suficientemente grande, entonces K_k tenderá a cero (se ajusta principalmente con la predicción de $\hat{m}_{k|k-1}$).
- Si $\hat{C}_{k|k-1}$ tiende a cero, entonces la actualización de las mediciones Y_k son casi ignorados y se ajustará principalmente por la medición estimada $Y_{k|k-1}$.

Se puede describir la ecuación 4.65 como

$$\begin{aligned} \hat{m}_{k+1|k} &= c + G\hat{m}_{k|k-1} + \underbrace{G\hat{C}_{k|k-1}F^T}_{K_k} \underbrace{\left(F\hat{C}_{k|k-1}F^T + \mathbf{V}\right)^{-1} \left(Y_k - d - F\hat{m}_{k|k-1}\right)}_{\epsilon_k}, \\ &= c + G\hat{m}_{k|k-1} + GK_k\epsilon_k, \end{aligned}$$

donde $\hat{Y}_{k|k-1} = d + F\hat{m}_{k|k-1}$. Ahora, si definimos el error de estimación de la variable estado $\tilde{X}_{k|k-1}$ por $\tilde{X}_{k|k-1} := X_k - \hat{m}_{k|k-1}$, entonces

$$\begin{aligned} Y_k - \hat{Y}_{k|k-1} &= d + Fx_k + v_k - d - F\hat{m}_{k|k-1} \\ &= F\tilde{X}_{k|k-1} + v_k \quad y \end{aligned}$$

$$\begin{aligned}
 \tilde{X}_{k+1|k} &= X_{k+1} - \hat{m}_{k+1|k} \\
 &= c + GX_k + \omega_k - c - G\hat{m}_{k|k-1} - GK_k (F \tilde{X}_{k|k-1} + v_k) \\
 &= (G - GK_k F) \tilde{X}_{k|k-1} + \omega_k - GK_k v_k,
 \end{aligned}$$

y tenemos que la estimación del error $\tilde{X}_{k|k-1}$ de la variable estado X_k se propaga según un sistema lineal con la dinámica de loop-cerrado $G - GK_k F$ generada por el proceso $\omega_k - GK_k v_k$, el cual es un ruido blanco con media cero y covarianza , donde $(G - GK_k F)$ será estable bajo ciertas condiciones, ver Anderson y Moore (1979).

Finalmente, a continuación se muestra que para encontrar las ecuaciones de filtraje o ecuación de actualización dado por 4.54 y 4.55 no es necesario asumir la distribución Gaussiana de los errores, ver 4.31 , basta que sean procesos ruidos blancos; es decir, que los errores ω_t y v_t sean independientes.

Comenzaremos por asumir de que $\hat{m}_{k|k}$ es una combinación lineal de la información presente (Y_k, c, d) y de la información pasada $Y_{1:k-1}$. Asumiremos, la forma lineal

$$\hat{m}_{k|k} = K_k Y_k + Ld + Mc + N\hat{m}_{k-1|k-1}, \quad (4.67)$$

con matrices arbitrarias K, L, M y N . Estas matrices son elegidos tal que $\hat{m}_{k|k}$ es el mínimo error cuadrático medio ECM; por tanto, definimos el error de estimación

$$\begin{aligned}
 e_k &= X_k - \hat{m}_{k|k} \\
 &= c + GX_{k-1} + \omega_k - K_k (d + F' (c + GX_{k-1} + \omega_k) + v_k) - Ld - Mc - N (X_{k-1} - e_{k-1}) \\
 &= (G - N - K_k F G) X_{k-1} - (L + K_k) d + (\mathbf{I} - M - K_k F) c + \\
 &\quad N e_{k-1} + (\mathbf{I} - K_k F) \omega_k - K_k v_k.
 \end{aligned} \quad (4.68)$$

Para que $\hat{m}_{k|k}$ sea el mínimo error cuadrático medio, el error de estimación e_k debe cumplir las siguientes dos condiciones : (i) incondicionalmente insesgado, i.e. el error de estimación debe tener media cero, lo cual resulta en las siguientes restricciones,

$$\begin{aligned}
 G - N - K_k F G &= 0, \\
 L + K_k &= 0, \\
 \mathbf{I} - M - K_k F &= 0
 \end{aligned}$$

para 4.68. Equivalentemente, si reemplazamos $N = (\mathbf{I} - K_k F) G$, el error de estimación

$$e_k = (\mathbf{I} - K_k F) (G e_{k-1} + \omega_k) - K_k v_k.$$

Además, se requiere que el error de estimación e_k debe tener (ii) mínima varianza, i.e. necesitamos minimizar la matriz de covarianza

$$\begin{aligned}
 \hat{C}_{k|k} &= E [((\mathbf{I} - K_k F) (G e_{k-1} - \omega_k) - K_k v_k) \\
 &\quad ((\mathbf{I} - K_k F) (G e_{k-1} - \omega_k) - K_k v_k)^\top] \\
 &= (\mathbf{I} - K_k F) (G \hat{C}_{k-1|k-1} G^\top + \mathbf{W}) (\mathbf{I} - K_k F)^\top + K_k \mathbf{V} K_k^\top \\
 &= (\mathbf{I} - K_k F) \hat{C}_{k|k-1} (\mathbf{I} - K_k F)^\top + K_k \mathbf{V} K_k^\top
 \end{aligned} \tag{4.69}$$

con respecto a K_k ; tomando la primera derivada parcial

$$\begin{aligned}
 \frac{\partial \hat{C}_{k|k}}{\partial K_k} &= \frac{\partial}{\partial K_k} \left(\hat{C}_{k|k-1} - K_k F \hat{C}_{k|k-1} - \hat{C}_{k|k-1} F^\top K_k^\top \right. \\
 &\quad \left. + K_k F \hat{C}_{k|k-1} F^\top K_k^\top + K_k \mathbf{V} K_k^\top \right) \\
 &= - \left(F \hat{C}_{k|k-1} \right)^\top - \hat{C}_{k|k-1} F^\top + 2 F \hat{C}_{k|k-1} F^\top K_k + 2 \mathbf{V} K_k \\
 &= -2 \hat{C}_{k|k-1} F^\top + 2 F \hat{C}_{k|k-1} F^\top K_k + 2 \mathbf{V} K_k.
 \end{aligned} \tag{4.70}$$

Igualando el resultado de 4.70 a cero según la condición necesaria de minimización, se obtiene la matriz de ganancia de Kalman

$$K_k = \hat{C}_{k|k-1} F_t^\top \left(F \hat{C}_{k|k-1} F^\top + \mathbf{V} \right)^{-1} \tag{4.71}$$

el cual es equivalente a la expresión dada por 4.64. Si reemplazamos 4.71 en 4.67 y 4.69 obtenemos el mismo resultado de la Proposición 4.3.6. A continuación se muestra el Algoritmo 4.1 sobre las recursiones de Filtro de Kalman.

Algoritmo 4.1 Filtro de Kalman con Parámetros Conocidos

```

1:Set the parameters c,d,G,F,W,V
2:Initialize  $m_0, C_0$ 
3:while (need to estimate the system state) do
4: for i = 1:N do
5: Prediction Equations
    $\hat{m}_{t_i} = c + Gm_{t_{i-1}}$ ;
    $\hat{C}_{t_i} = GC_{t_{i-1}}G' + W$ ;
    $\hat{Y}_{t_i} = d + F\hat{m}_{t_i}$ ;
6: Get the measurement  $Y_t$ 
7: Innovation
    $\epsilon_{t_i} = Y_{t_i} - \hat{Y}_{t_i}$ ;
8: Updating Equations
    $Z_{t_i} = F\hat{C}_{t_i}F' + V$ ;
    $K_{t_i} = \hat{C}_{t_i}F'Z_{t_i}^{-1}$ ;
    $m_{t_i} = \hat{m}_{t_i} + K_{t_i}\epsilon_{t_i}$ ;
    $C_{t_i} = (I - K_{t_i}F)\hat{C}_{t_i}$ ;
    $dZ_{t_i} = \det Z_{t_i}$ ;
   if  $dZ_{t_i} \leq 0$  then
      $dZ_{t_i} = 10^{-10}$ 
   end if
   end for
9:end while

```

Capítulo 5.

Estimación paramétrica con Máxima Verosimilitud

Los parámetros de nuestro modelo necesitan ser estimados a partir de una secuencia de observaciones históricas de precios spot y futuros $Y = [y_1 \ y_2 \ \cdots \ y_N]$. Schwartz y Smith sugieren usar el método de estimación paramétrica por máxima verosimilitud. La función de verosimilitud, ver la Definición 4.1.3, puede calcularse como la función densidad de probabilidad conjunta de las observaciones:

$$L(\theta, \hat{Y}) = \text{Prob}(y_t = \hat{y}_t, t = 1, 2, \dots, N), \quad (5.1)$$

donde $\theta = \{\kappa, \mu_\xi, \sigma_\chi, \sigma_\xi, \rho_{\chi\xi}, \lambda_\chi, \lambda_\xi, s_1, s_2, \dots, s_n, m_0, C_0\}$ denota al conjunto de parámetros del modelo, $\hat{Y} = [\hat{y}_1 \ \hat{y}_2 \ \cdots \ \hat{y}_N]$ es la estimación $\hat{y}_{t|t-1}$ de la predicción un paso adelante de Y_t basado sobre las observaciones y_1, y_2, \dots, y_{t-1} así como también de m_0 y C_0 donde $X_0 \sim \mathcal{N}(m_0, C_0)$.

El problema de la estimación por máxima verosimilitud se divide en dos partes.

- Primero, la predicción óptima donde se asume que los parámetros del modelo θ es conocido. Este es el problema de filtro de Kalman y cuya solución se abordó en el capítulo anterior.
- Una vez determinado la forma de la función de verosimilitud $L(\theta, \hat{Y})$, la segunda parte es la estimación paramétrica.

La segunda parte es lo que mostraremos en este capítulo. Previamente, presentaremos algunas notaciones del problema de estimación paramétrica por máxima verosimilitud adicional a lo presentado en la sección 4.1 del capítulo precedente.

Planteamiento del problema : Dado λ una medida σ -finita sobre $(\mathbb{S}, \mathcal{S})$, consideremos una familia $\{f(\cdot, \theta)\}_{\theta \in \Theta}$ de funciones no negativas λ -integrables sobre \mathbb{S} . Esta familia está indexada por un parámetro $\theta \in \Theta$, donde Θ es un subconjunto de \mathbb{R}^{d_θ} (para algún entero d_θ). El problema bajo consideración es la maximización de la integral

$$L(\theta) \triangleq \int f(x; \theta) \lambda(dx) \quad (5.2)$$

con respecto al parámetro θ .

En lo que sigue, asumiremos que $L(\theta)$ es positivo, y por lo tanto maximizar $L(\theta)$ es equivalente a maximizar

$$l(\theta) \triangleq \log L(\theta). \quad (5.3)$$

A l se le conoce como log-likelihood o log-verosimilitud. Asociaremos para cada función $f(\cdot, \theta)$ una función de densidad de probabilidad $p(\cdot; \theta)$ (con respecto a la medida λ) definida por

$$p(x; \theta) \triangleq f(x; \theta) / L(\theta). \quad (5.4)$$

Se interpreta a $p(x; \theta)$ como la densidad condicional de X dado Y .

5.1. Solución por el método del Gradiente

El método del gradiente en el contexto de modelo oculto de Markov consiste en evaluar las derivadas de la función objetivo $l(\theta)$ con respecto al parámetro θ y para ello, hace uso de herramientas de métodos de optimización basados en la gradiente.

La condición para aplicar este método es que $l(\theta)$ esté bien definida. A continuación se impone un conjunto de condiciones mínimas que garantiza que la primera y segunda derivada de $l(\theta)$ se encuentren bien definidas.

(A.1) El conjunto de parámetros Θ es un subconjunto abierto de \mathbb{R}^{d_θ} (para algún entero d_θ).

(A.2) Para toda $\theta \in \Theta$, $L(\theta)$ es positiva y finita.

(A.3) Para todo $(\theta, \theta') \in \Theta \times \Theta$, $\int |\nabla_\theta \log p(x, \theta)| p(x; \theta') \lambda(dx)$ es finito.

El tercer supuesto implica que en particular las distribuciones de probabilidad en la familia $\{p(\cdot; \theta) d\lambda\}_{\theta \in \Theta}$ son absolutamente continuas una con respecto a otro.

Antes de enunciar la proposición de Identidades de Fisher y Loui definiremos $\mathcal{H}(\theta, \theta')$, como

$$\mathcal{H}(\theta, \theta') \triangleq - \int \log p(x; \theta) p(x; \theta') \lambda(dx).$$

Se le llama a $\mathcal{H}(\theta, \theta')$ la entropía de la función densidad de probabilidad $p(\cdot; \theta')$, para mayor detalle ver Cappé (2005, pag. 350).

Proposición 5.1.1. (*Identidades de Fisher y Louis*). Asuma los tres supuestos (A.1)-(A.3) y las siguientes condiciones.

- (a) $\theta \mapsto L(\theta)$ es dos veces continuamente diferenciable sobre Θ .
- (b) Para cualquier $\theta' \in \Theta$, $\theta \mapsto \mathcal{H}(\theta; \theta')$ es dos veces continuamente diferenciable sobre Θ . Además, $\int |\nabla_{\theta}^k \log p(x; \theta)| p(x; \theta') \lambda(dx)$ es finito para $k = 1, 2$ y cualquier $(\theta; \theta') \in \Theta \times \Theta$, y

$$\nabla_{\theta}^k \int \log p(x; \theta) p(x; \theta') \lambda(dx) = \int \nabla_{\theta}^k \log p(x; \theta) p(x; \theta') \lambda(dx).$$

Entonces las siguientes identidades cumplen:

$$\nabla_{\theta} l(\theta') = \int \nabla_{\theta} \log f(x; \theta) |_{\theta=\theta'} p(x; \theta') \lambda(dx), \quad (5.5)$$

$$\begin{aligned} \nabla_{\theta}^2 l(\theta') &= - \int \nabla_{\theta}^2 \log f(x; \theta) |_{\theta=\theta'} p(x; \theta') \lambda(dx) \\ &\quad + \int \nabla_{\theta}^2 \log p(x; \theta) |_{\theta=\theta'} p(x; \theta') \lambda(dx). \end{aligned} \quad (5.6)$$

La segunda igualdad puede describirse en la forma equivalente

$$\begin{aligned} \nabla_{\theta}^2 l(\theta') + \{ \nabla_{\theta} l(\theta') \} \{ \nabla_{\theta} l(\theta') \}^t &= \int [\nabla_{\theta}^2 \log f(x; \theta) |_{\theta=\theta'} + \{ \nabla_{\theta} \log f(x; \theta) |_{\theta=\theta'} \} \\ &\quad \times \{ \nabla_{\theta} \log f(x; \theta) |_{\theta=\theta'} \}^t] p(x; \theta') \lambda(dx). \end{aligned} \quad (5.7)$$

La 5.5 comúnmente se le denomina la identidad de Fisher. Cuando la función de verosimilitud está asociada a un modelo estadístico en particular, se le conoce también como la función score. La 5.6 muestra que la función score puede evaluarse calculando la esperanza bajo $p(\cdot; \theta')$ de la función $\nabla_{\theta} \log f(X; \theta) |_{\theta=\theta'}$. Esta función se refiere a su vez a la función score completa pues $\log f(x; \theta)$ es el log-likelihood conjunta de la data completa (X, Y) .

Algoritmo de Newton-Raphson

Uno de los métodos para implementar la optimización basado en el enfoque gradiente es el algoritmo de Newton-Raphson:

$$\theta^{i+1} = \theta^i - H^{-1}(\theta^i) \nabla_{\theta} l(\theta^i), \quad (5.8)$$

donde $H(\theta^i) = \nabla_{\theta}^2 l(\theta^i)$ es la Hessian de la función objetivo. La iteración de Newton está basada en la aproximación de segundo orden

$$l(\theta) \approx l(\theta') + \nabla l(\theta') (\theta - \theta') + \frac{1}{2} (\theta - \theta')^t H(\theta') (\theta - \theta').$$

Si la secuencia $\{\theta^i\}_{i \geq 0}$ producida por el algoritmo converge al punto θ_* en el cual el Hessiano es definido positivo, la convergencia es cuadrática en el sentido de que para suficientemente largo valor de i existe una constante positiva β tal que

$$\|\theta^{i+1} - \theta_*\| \leq \beta \|\theta^i - \theta_*\|^2.$$

5.2. Aplicación al modelo oculto de Markov

El modelo oculto de Markov, descrito en la sección 3.1, corresponde a la subcategoría de modelos de data incompleta conocido como modelos de data perdida. En tales modelos, la data observada Y es un subconjunto de cierta data completa no completamente observable (X, Y) . Asumiremos que la distribución conjunta de X e Y , dado cierto valor de parámetro θ , admite una función de densidad de probabilidad conjunta $f(x, y; \theta)$ con respecto a la medida producto $\lambda \otimes \mu$. Es importante entender que f es una función de densidad de probabilidad solamente cuando es función de x e y . Para cierto valor fijo de y , es decir cuando $f(\cdot, \theta)$ es función solamente de x , diremos que f es una función positiva integrable. En efecto, el likelihood de la observación, el cual es definido como la función densidad de probabilidad de Y con respecto a μ , se obtiene como marginalización

$$L(y; \theta) \triangleq \int f(x; y; \theta) \lambda(dx). \quad (5.9)$$

Para cierto valor de y , este es un caso particular de 5.2. En nuestro modelo, la familia de funciones de densidad de probabilidad $\{p(\cdot; \theta)\}_{\theta \in \Theta}$ definido en 5.4 puede por tanto interpretarse como

$$p(x | y; \theta) = \frac{f(x; y; \theta)}{\int f(x; y; \theta) \lambda(dx)}, \quad (5.10)$$

la función densidad de probabilidad condicional de X dado Y .

Usando la notación del capítulo anterior, asumiremos que las observaciones de Y_0 a Y_t (o en representación mas concisa $Y_{0:t}$) están disponibles. Como solamente consideraremos el modelo oculto de Markov totalmente dominada según la Definición 3.3.7, usaremos las notaciones v y $\pi_{k|t}$ para referirnos a las funciones de densidad de probabilidad de estas distribuciones (de X_0 y de X_k dado $Y_{0:t}$) con respecto a la medida dominante λ . La función de densidad de la probabilidad conjunta de los estados $X_{0:t}$ y las observaciones asociadas $Y_{0:t}$, con respecto a la medida producto $\lambda^{\otimes(t+1)} \otimes \mu^{\otimes(t+1)}$, está dado por 3.20 :

$$f_t(x_{0:t}, y_{0:t}; \theta) = v(x_0; \theta) g(x_0, y_0; \theta) K(x_0, x_1; \theta) g(x_1, y_1; \theta) \\ \times \cdots K(x_{t-1}, x_t; \theta) g(x_t, y_t; \theta), \quad (5.11)$$

donde hemos usado la misma convención de 5.9 para indicar la dependencia con respecto al parámetro θ .

Como principalmente consideramos la estimación paramétrica del vector θ de secuencias de observaciones, no tiene mucho sentido considerar a v como un parámetro independiente. No se puede estimar v consistentemente pues solo se tiene una variable aleatoria X_0 , que además no es observable. En lo que sigue asumiremos que v es conocido o toma un valor fijo. Por ejemplo, v puede ser la distribución estacionaria asociada con la función de transición $K(\cdot, \cdot; \theta)$. Indiferentemente de si v es fijo o determinado por θ , se omite la dependencia con respecto a v en nuestra notación.

La función de verosimilitud de las observaciones $L(y_{0:t}; \theta)$ se obtiene integrando 5.11 con respecto a x , las variables de estado, bajo la medida $\lambda^{\otimes(t+1)}$. Tomando logaritmos en 5.11,

$$\log f_t(x_{0:t}, y_{0:t}; \theta) = \log v(x_0; \theta) + \sum_{k=0}^{t-1} \log K(x_k, x_{k+1}; \theta) \\ + \sum_{k=0}^t \log g(x_k, y_k; \theta). \quad (5.12)$$

En el caso del modelo oculto de Markov, la identidad de Fisher - 5.5 - se simplifica a

$$\nabla_{\theta} l_t(\theta) = E_{\theta} [\nabla_{\theta} \log v(X_0; \theta) | Y_{0:t}] + \sum_{k=0}^t E_{\theta} [\nabla_{\theta} \log g_k(X_k; \theta) | Y_{0:t}] \\ + \sum_{k=0}^{t-1} E_{\theta} [\nabla_{\theta} \log K(X_k, X_{k+1}; \theta) | Y_{0:t}]. \quad (5.13)$$

En el caso de 5.7 se puede reescribir como

$$\begin{aligned} \nabla_{\theta} l_t(\theta) + \{\nabla_{\theta} l_t(\theta)\} \{\nabla_{\theta} l_t(\theta)\}^t &= \sum_{k=0}^t E_{\theta} \left[\nabla_{\theta}^2 \log g_k(X_k; \theta) \mid Y_{0:t} \right] \\ &+ \sum_{k=0}^t \sum_{j=0}^t E_{\theta} \left[\{\nabla_{\theta} \log g_k(X_k; \theta)\} \{\nabla_{\theta} \log g_j(X_j; \theta)\}^t \mid Y_{0:t} \right]. \end{aligned}$$

Observación 5.2.1. Relación con el Enfoque Recursivo

Existe un enfoque alternativo para estimar la gradiente de la función de verosimilitud sin hacer uso de la identidad de Fisher. Este método se conoce como 'sensitivity equations' para mayor detalle ver Cappé (2005, pag 364).

Recordemos del capítulo anterior de que

$$l_t(\theta) = \sum_{k=0}^t \log \underbrace{\int \pi_{t|t-1}(x_k; \theta) g_k(x_k; \theta) \lambda(dx_k)}_{c_k(\theta)}, \quad (5.14)$$

La ecuación de filtraje dado por 4.27 implica que

$$\pi_{k+1}(x_{k+1}; \theta) = c_{k+1}^{-1}(\theta) \int \pi_k(x_k; \theta) K(x_k, x_{k+1}; \theta) g_{k+1}(x_{k+1}; \theta) \lambda(dx_k). \quad (5.15)$$

Para la primera derivada de 5.15 con respecto a θ , se asume que $c_{k+1}(\theta)$ no es nulo y usamos la identidad

$$\nabla_{\theta} \frac{u(\theta)}{v(\theta)} = v^{-1}(\theta) \nabla_{\theta} u(\theta) - \frac{u(\theta)}{v(\theta)} \nabla_{\theta} \log v(\theta)$$

para obtener

$$\nabla_{\theta} \pi_{k+1}(x_{k+1}; \theta) = \rho_{k+1}(x_{k+1}; \theta) - \pi_{k+1}(x_{k+1}; \theta) \nabla_{\theta} \log c_{k+1}(\theta), \quad (5.16)$$

donde

$$\rho_{k+1}(x_{k+1}; \theta) \triangleq c_{k+1}^{-1}(\theta) \nabla_{\theta} \int \pi_k(x_k; \theta) K(x_k, x_{k+1}; \theta) g_{k+1}(x_{k+1}; \theta) \lambda(dx_k) \quad (5.17)$$

Podemos intercambiar la integración con respecto a λ y diferenciar con respecto a λ y diferenciar con respecto a θ . Debido a que $\pi_{k+1}(\cdot; \theta)$ es una función de probabilidad de densidad, $\int \pi_{k+1}(x_{k+1}; \theta) \lambda(dx_{k+1}) = 1$ y

$$\nabla_{\theta} \int \pi_{k+1}(x_{k+1}; \theta) \lambda(dx_{k+1}) = \int \nabla_{\theta} \pi_{k+1}(x_{k+1}; \theta) \lambda(dx_{k+1}) = 0.$$

. Por lo tanto la integración de ambos lados de 5.16 con respecto a $\lambda(dx_{k+1})$ resulta

$$0 = \int \rho_{k+1}(x_{k+1}; \theta) \lambda(dx_{k+1}) - \nabla_{\theta} \log c_{k+1}(\theta).$$

Por lo tanto, se puede evaluar la gradiente del incremental log-likelihood en términos de ρ_{k+1} según

$$\nabla_{\theta} \log c_{k+1}(\theta) \triangleq \nabla_{\theta} (l_{k+1}(\theta) - l_k(\theta)) = \int \rho_{k+1}(x_{k+1}; \theta) \lambda(dx_{k+1}). \quad (5.18)$$

Evaluaremos ahora la derivada en 5.17 asumiendo que K y g_k son diferentes de cero para obtener

$$\begin{aligned} \rho_{k+1}(x_{k+1}; \theta) &= c_{k+1}^{-1}(\theta) \int \{[\nabla \log K(x_k, x_{k+1}; \theta) + \nabla_{\theta} \log g_{k+1}(x_{k+1}; \theta)] \\ &\quad \times \pi_{k+1}(x_k; \theta) + \nabla_{\theta} \pi_k(x_k; \theta)\} K(x_k, x_{k+1}; \theta) g_{k+1}(x_{k+1}; \theta) \lambda(dx_k). \end{aligned}$$

Reemplazando 5.16 en la expresión de arriba resulta una fórmula de actualización de ρ_{k+1} ,

$$\begin{aligned} \rho_{k+1}(x_{k+1}; \theta) &= c_{k+1}^{-1}(\theta) \int \{[\nabla \log K(x_k, x_{k+1}; \theta) + \nabla_{\theta} \log g_{k+1}(x_{k+1}; \theta)] \\ &\quad \times \pi_{k+1}(x_k; \theta) + \rho_k(x_k; \theta)\} K(x_k, x_{k+1}; \theta) g_{k+1}(x_{k+1}; \theta) \lambda(dx_k) \\ &\quad - \pi_{k+1}(x_k; \theta) \nabla_{\theta} \log c_k(\theta), \end{aligned} \quad (5.19)$$

5.3. Aplicación al modelo de Schwartz y Smith

Para nuestro propósito de estimación del vector de parámetros θ dada la data Y , se usa el enfoque de Schweppe (1965) conocido como el enfoque de descomposición del error de predicción de la función de verosimilitud a ser maximizado con respecto θ . Es decir, expresamos la función de verosimilitud en términos del error de predicción. Como la varianza del error de predicción

$$\epsilon_t = y_t - E[y_t | y_{1:t-1}] \quad (5.20)$$

es igual a la varianza condicional de y_t ,

$$\text{Cov}[\epsilon_t | y_{1:t-1}] = \text{Cov}[y_t | y_{1:t-1}],$$

podemos determinar la función densidad de las observaciones. La función general de densidad de $E[f(y_t) | y_{t-1}, \theta]$ está dada por la distribución Gaussiana con media condicional $E[y_t | \mathcal{F}_{t-1}] = d - F\hat{\eta}_{k|k-1}$ y matriz varianza-covarianza $\text{Cov}[y_t | \mathcal{F}_{t-1}] = F\hat{C}_{k|k-1}F^T + \mathbf{V}$ al cual lo hemos denotado por $Z_{t|t-1}$ - ver 4.60

$$g(x_t, y_t; \theta) = \frac{1}{\sqrt{2\pi |Z_{t|t-1}|}} \exp\left[-\frac{1}{2}\epsilon_t^T Z_{t|t-1} \epsilon_t\right], \quad (5.21)$$

con ϵ_t definido en 5.20. Luego, tomando logaritmos a 5.21 resulta

$$\log g(x_t, y_t; \theta) = -\frac{k}{2} \log 2\pi - \frac{1}{2} \log |Z_{t|t-1}| - \frac{1}{2} \epsilon_t^T Z_{t|t-1} \epsilon_t,$$

donde k representa la dimensión del vector y_t . La función de verosimilitud en logaritmos resulta

$$\begin{aligned} l(\theta) &= \sum_{t=1}^n \log g(x_t, y_t; \theta) \\ &= -\frac{1}{2} \sum_{t=1}^n \left(-\frac{k}{2} \log 2\pi + \log |Z_{t|t-1}| + \epsilon_t^T Z_{t|t-1} \epsilon_t \right). \end{aligned}$$

en lugar de 5.1. Este resultado se resume en el siguiente Lema.

Lema 5.3.1. *La recursión para la función de verosimilitud está dado por*

$$l_t(\theta) = l_{t-1}(\theta) + \frac{1}{2} \log |2\pi Z_t(\theta)| + \frac{1}{2} \epsilon_t^\top(\theta) Z_t^{-1}(\theta) \epsilon_t(\theta), \quad (5.22)$$

donde los términos $\epsilon_t(\theta)$, y $Z_t(\theta)$ se obtienen del filtro de Kalman con parámetros fijos para θ .

- Predicción

$$\begin{aligned} \hat{m}_{t|t-1}(\theta) &= c + G_\theta \hat{m}_{t-1}(\theta) \\ \hat{C}_{t|t-1}(\theta) &= G_\theta \hat{C}_{t-1}(\theta) G_\theta^\top + \mathbf{W}_\theta \end{aligned} \quad (5.23)$$

- Actualización

$$\begin{aligned} \epsilon_t(\theta) &= y_t - d - F_\theta \hat{m}_{t|t-1}(\theta) \\ Z_t(\theta) &= F_\theta \hat{C}_{t|t-1}(\theta) F_\theta^\top + \mathbf{V}_\theta \\ K_t(\theta) &= \hat{C}_{t|t-1}(\theta) F_\theta^\top Z_t^{-1}(\theta) \\ m_t(\theta) &= \hat{m}_{t|t-1}(\theta) + K_t(\theta) \epsilon_t(\theta) \\ C_t(\theta) &= \hat{C}_{t|t-1}(\theta) - K_t(\theta) F_\theta \hat{C}_{t|t-1}(\theta) \end{aligned} \quad (5.24)$$

El lema nos dice que podemos encontrar los valores para la esperanza condicional $E[y_{t|t-1}]$ y la matriz de varianza condicional $Cov[y_{t|t-1}]$ para cierto parámetro θ usando el algoritmo de Filtro de Kalman. Por lo tanto, podemos calcular el valor de la función condicional de log-verosimilitud numéricamente. Dada la función verosimilitud, se puede usar un método computacional de optimización para encontrar los parámetros, en particular el método de la gradiente.

$$\hat{\theta} = \arg \max_{\theta \in \Theta} l_t(\theta).$$

En Tai y Fu (2009), se encuentra un abordaje computacional eficiente y explotan el estado de equilibrio de Filtro de Kalman en el contexto del modelo de Schwartz y Smith (2000).

El algoritmo y diagrama de Flujo del Lema 5.3.1 se presenta a continuación :

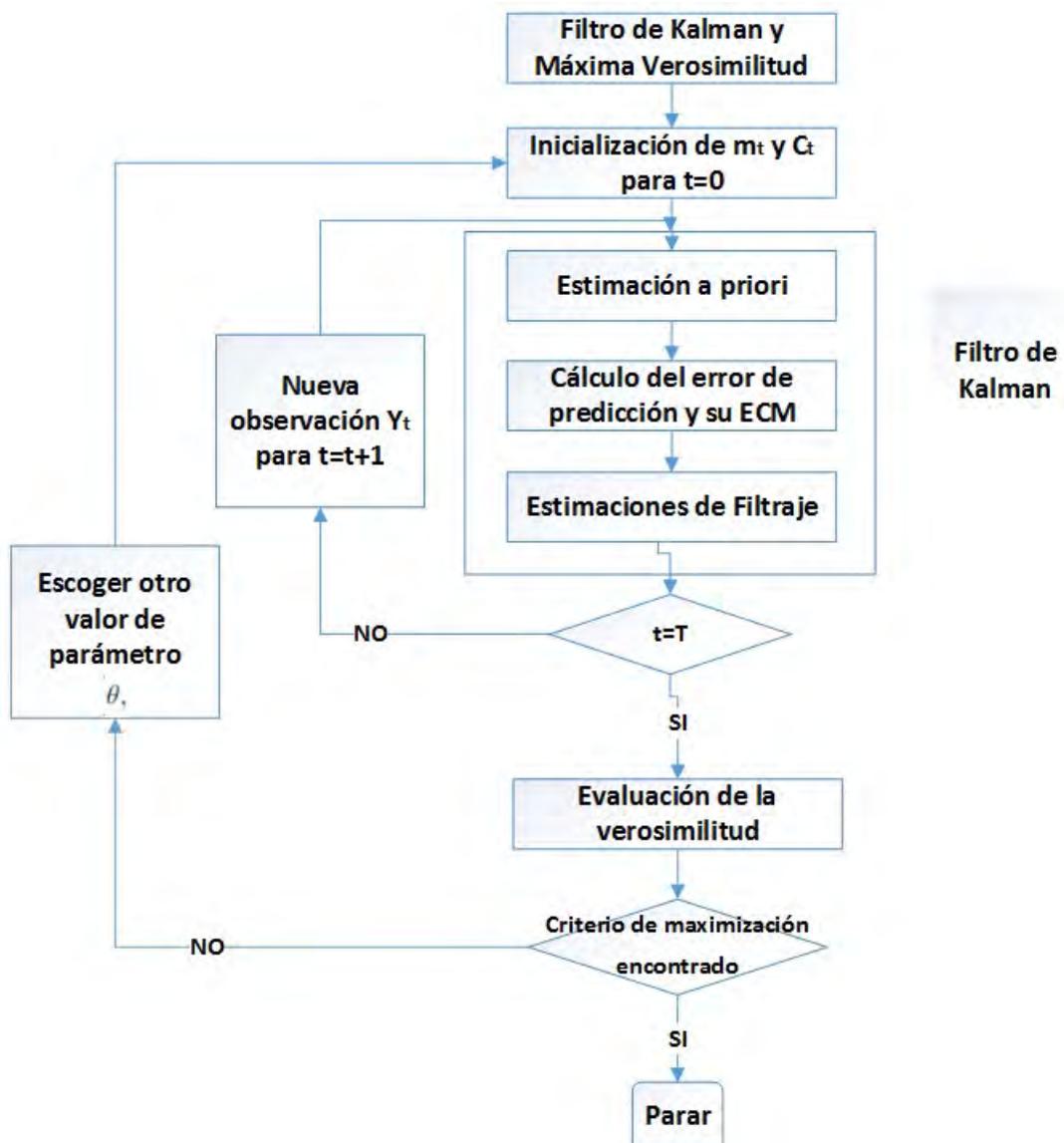
Algoritmo 5.1 Filtro de Kalman con Máxima Verosimilitud

```

1:Set initial guess c,d,G,F,W,V
2:Initialize  $m_0, C_0$ 
3:while (need to estimate the system state, optimal  $\theta$  has not been found) do
  ( Kalman Filter)
4: for i = 1:N do
5: Prediction Equations
    $\hat{m}_{t_i} = c + Gm_{t_{i-1}}$ ;
    $\hat{C}_{t_i} = GC_{t_{i-1}}G' + W$ ;
    $\hat{Y}_{t_i} = d + F\hat{m}_{t_i}$ ;
6: Get the measurement  $Y_t$ 
7: Innovation
    $e_{t_i} = Y_{t_i} - \hat{Y}_{t_i}$ ;
8: Updating Equations
    $Z_{t_i} = F\hat{C}_{t_i}F' + V'$ ;
    $K_{t_i} = \hat{C}_{t_i}F'Z_{t_i}^{-1}$ ;
    $m_{t_i} = \hat{m}_{t_i} + K_{t_i}e_{t_i}$ ;
    $C_{t_i} = (I - K_{t_i}F)\hat{C}_{t_i}$ ;
    $dZ_{t_i} = \det Z_{t_i}$ ;
   if  $dZ_{t_i} \leq 0$  then
      $dZ_{t_i} = 10^{-10}$ 
   end if
9: (Log-Likelihood function per iteration)
    $\log l(i) = -(n/2) * \ln(2 * \pi) - 0.5 * \ln(dZ_{t_i}) - 0.5 * e_{t_i}' * Z_{t_i}^{-1} * e_{t_i}$ ;
   end for
LogL =  $\sum_i \log l(i)$ 
(Adjustment for  $\varphi = \log l(i)$  via a Matlab optimization routine)
10:end while

```

Figura 5.1.: Diagrama de Flujo de Filtro de Kalman y Máxima Verosimilitud



Capítulo 6.

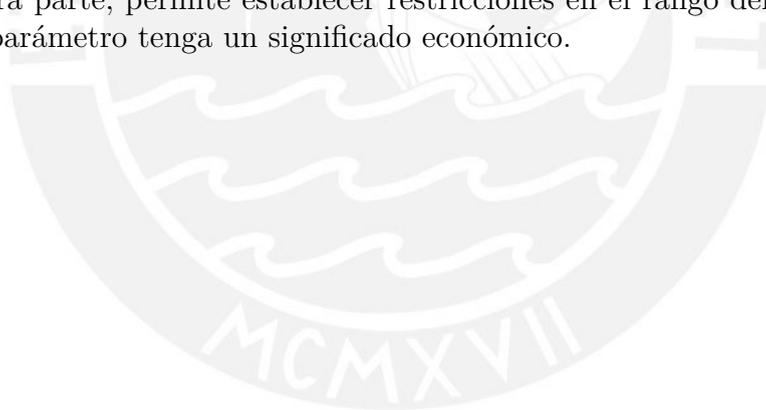
Conclusión

- El modelo spot de Schwartz y Smith es un modelo estocástico lineal homogéneo y bivariado. Los procesos que la definen son del tipo de procesos de difusión: ξ_t sigue la dinámica de un proceso Browniano aritmético y χ_t , la de un proceso Ornstein-Uhlenbeck. Aplicando cálculo estocástico, se encontró la solución del modelo de manera explícita, la cual se distribuye conjuntamente como una Gaussiana bivariada y por lo tanto, el precio spot sigue una distribución log-normal.
- La curva de futuros del modelo de Schwartz y Smith se obtuvo aplicando la ecuación de valoración de riesgo neutral para contratos futuros, el cual por el Teorema 2.2.3 resulta ser una martingala bajo la medida de riesgo neutral. El modelo spot de Schwartz y Smith con un activo libre de riesgo induce un mercado incompleto libre de arbitraje, como resultado de aplicar el Meta-Teorema 2.1.1 y de observar que los procesos que describen el modelo spot son factores no observables que no tienen precios; por el cual se le denomina también modelo de factores. Del “Segundo Teorema Fundamental de Valoración de Activos”, a partir del modelo no es posible obtener directamente la “correcta y única” medida de riesgo neutral tal que verifica la condición del Teorema 2.3.1. sobre el Teorema de Girsanov.
- La medida de riesgo neutral se define asumiendo que los precios de mercado de riesgo $\lambda := (\lambda_1, \lambda_2)$ son constantes, aplicando el Teorema 2.3.2. Al comparar la dinámica de los procesos bajo la medida \mathbb{P} con los obtenidos bajo \mathbb{Q} , se observa que el proceso χ_t bajo \mathbb{Q} sigue siendo un proceso Ornstein-Uhlenbeck pero ahora revierte a $-\lambda_\chi/\kappa$, mientras que bajo \mathbb{P} este revertía a 0. El proceso ξ_t bajo \mathbb{Q} sigue siendo un movimiento Browniano aritmético, pero con drift $\mu_\xi^* \triangleq \mu_\xi - \lambda_\xi$; por lo tanto el vector $(\chi_T, \xi_T)^\top$ se sigue distribuyendo como un proceso Gaussiano.
- La curva de precios futuros se muestra en la ecuación 2.23, la cual resulta ser lineal con distribución Gaussiana, $Y_t \equiv \log(F_{t,T}) = e^{-\kappa(T-t)}\chi_t + \xi_t + A(t, T)$. Para contratos próximos a vencer (i.e. $T = 0$), la varianza es igual a $Cov(\chi_t + \xi_t)$. Cuando T aumenta, la contribución de la varianza de χ_t disminuye, y para $T \rightarrow \infty$, la varianza de $\ln(F_t)$ se aproxima a σ_ξ .
- La curva de precios futuros y el proceso (χ_T, ξ_T) definen la representación espacio estado dado en el Lema 3.1.1, donde la estructura probabilística es la de un modelo oculto de Markov según la Proposición 3.3.9 y el cual se demuestra en el Lema 4.3.2.
- Las ecuaciones recursivas de filtraje de un modelo oculto de Markov se obtienen de la distribución marginal de la distribución conjunta dada en la Proposición 4.1.7, siendo

una aplicación directa de la Proposición 4.2.6. Para modelos lineales Gaussianos como el sistema inducido por el modelo spot y la curva de futuros se muestran las ecuaciones de filtraje y de predicción en la Proposición 4.3.7, el cual es aplicación directa de las propiedades de un modelo lineal Gaussiano condicionado. En el Lema 5.3.1 se presenta la estimación recursiva de la función de verosimilitud de nuestro modelo.

Trabajos Futuros

Una de las líneas de trabajos futuros se presenta en el área de abordaje computacional. Por ejemplo, Sauvageau y Kumral (2016) recientemente realizan un estudio de simulación de precios futuros para un proyecto minero de depósitos de cobre. Lo novedoso es que alternativamente al método tradicional de optimización de Expectation-Maximisation emplean el algoritmo Genético con Filtros de Kalman. Lo que sucede es que los valores iniciales del modelo tendrán un efecto significativo sobre la convergencia de la función objetivo de maximización. Para garantizar que la solución no se corresponda a un máximo local, el algoritmo de Expectation-Maximisation se realiza varias veces con diferentes valores iniciales. Frente a este limitante, se buscan métodos más eficientes, entre ellos los algoritmos genéticos para optimizar los parámetros del modelo. Aunque son más lentos que los algoritmos tradicionales porque usan generadores de números aleatorios para buscar la solución óptima, el algoritmo genético optimiza una población de soluciones en lugar de trabajar en una sola solución a la vez. Por otra parte, permite establecer restricciones en el rango del parámetro, lo cual asegura que el parámetro tenga un significado económico.

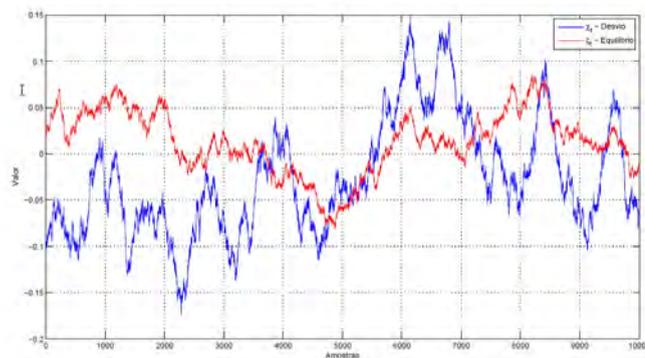


Apéndice A.

Precios spot simulado

Este anexo hace referencia a de la simulación de las ecuaciones de la sección 3.3.

Figura A.1.: Realización de los Procesos χ_t y ξ_t .



En el Cuadro 3.1 se muestra los parámetros asumidos para la simulación de los precios spot.¹

Cuadro A.1.: Parámetros del Modelo

Variable	Símbolo	Valor
Variación del Precio de corto plazo	χ_0	0.461
Precio de largo plazo	ξ_0	-0.1985
Drift en el precio de largo plazo	μ_ξ	0.00 %
Volatilidad del precio de largo plazo	σ_ξ	5.00 %
Tasa de reversión a la media	κ	0.2700
Volatilidad de la variación del Precio de corto plazo	σ_χ	25.00 %
Correlación	$\rho_{\chi\xi}$	0.1000

¹Figura tomado de Da Silva.M (2011) en “Calibracao do Modelo de Schwartz-Smith com Filtro de Kalman” y de Graham (2012)

Bibliografía

- [1] Arnold, L.1992. *Stochastic Differential Equations: Theory and Applications*. Krieger Publishing Company, Malabar.
- [2] Anderson, B y Moore, J. 1979. Optimal Filtering. Information and System Sciences Series.
- [3] Björk, Tomas.Arbitrage theory in continuous time .Oxford ; New York : Oxford University Press, 2009.
- [4] Black,F y Scholes,M. The pricing of options and corporate liabilities. Journal of Political Economy, 81(3):637–654, 1973.
- [5] Burger,M , Graeber, B y Schindlmayr, G. 2007. Managing Energy Risk: An Integrated View on Power and Other Energy Markets. The Wiley Finance Series.
- [6] Catlin, D. 1989. Estimation, Control, and the Discrete Kalman Filter. Springer-Verlag.
- [7] Cappé, O, Moulines, E y Ryden, T.2005.Inference in Hidden Markov Models. Springer.
- [8] Cappé, O, Moulines, E y Ryden, T.2009.Inference in Hidden Markov Models.Springer.
- [9] Carmona, R. 2014. Quantitative Energy Finance:Modeling, Pricing, and Hedging in Energy and Commodity Markets.Editado por Benth, F.
- [10] Durrett, R. 1984. Brownian Motion and Martingales in Analysis. Wadsworth.
- [11] Elliott, R. y Ekkehard, P. 2005. Mathematics of Financial Markets. Second edition.Springer.
- [12] Espend, F. 2011. The Risk Premium in Commodity Markets.Unpublished working pape pero recuperado el 17/07/2018 en https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=finanzas2009&paper_id=111.
- [13] Gibson, R.,E. Schwartz. 1990. Stochastic convenience yield and the pricing of oil contingent claims. J. Finance 45 959 - 976.
- [14] Graham A. Davis.Technical Note: Simulating the Two-Factor Schwartz and Smith Model of Commodity Prices.Journal The Engineering Economist A Journal Devoted to the Problems of Capital Investment Volume 57, 2012 - Issue 2.
- [15] Harvey,A. 1994. Forecasting, Structural Time Series Models and the Kalman Filter.
- [16] Iacus, S.2008 .Simulation and Inference for Stochastic Differential Equations: With R Examples. Springer Series in Statistics.
- [17] Jarrow, R. 1987. The Pricing of Commodity Options with Stochastic Interest Rates. Advances in Futures and Options Research, Vol. 2, pages 19-45.

- [18] Karatzas y Shreve. 1998 . Methods of mathematical finance. Springer.
- [19] Kalman, R.E. A new approach to linear filtering and prediction, Journal of Basic Engineering (ASME), 82D, March 1960, 35-45.
- [20] Kloeden, P, y Platen,E . 1992. Numerical Solution of Stochastic Differential Equations. Springer.
- [21] Mastro, M. 2013. Financial Derivative and Energy Market Valuation: Theory and Implementation in MATLAB. A Jhon Wiley & Sons , Inc. Publication.
- [22] Muirhead, S. 2011. Financial Geometry: Pricing multi-asset barrier options using the generalised reflection principle. Honours Thesis, The University of Melbourne.
- [23] Mendiola, A ; Aguirre, C y otros. 2014. Valoración de una empresa con opciones reales: el caso de Minera Aurífera Peruana.– Lima : Universidad ESAN, 2014. – 102 p. – (Serie Gerencia para el Desarrollo ; 34)
- [24] Øksendal, B. Stochastic Differential Equations. An Introduction with Applications.Springer. Fifth Edition, 1998.
- [25] Särkkä, S. 2012. Applied Stochastic Differential Equations. Lecture notes of the course Becs-114.4202 Special Course in Computational Engineering II held in November 2012. Version 1.0 (November 21, 2012).
- [26] Sauvageau, M. y Kumral, M. 2016. Genetic algorithms for the optimisation of the Schwartz–Smith two-factor model: a case study on a copper deposit. International Journal of Mining, Reclamation and Environment, DOI: 10.1080/17480930.2016.1260858
- [27] Schwartz, E. S. and Smith, (2000). Short-Term variations and long-term dynamics in commodity prices. . Management Science, 46(7):893–911, July 2000.
- [28] Schwartz, S.(1997) The stochastic behavior of commodity prices: Implications for valuation and hedging. Journal of Finance, 52(3):923–73, July 1997.
- [29] Schweppe, F . (1965), Evaluation of likelihood functions for gaussian signals. IEEE Trans. Information Theory IT-11, 61-70.
- [30] Shreve, S. 2004. Stochastic Calculus for Finance II Continuous - Time Models. Springer
- [31] Tai, X y Fu,M. (2009). On Parameter Estimation of the Schwartz-Smith Short-Term/Long-Term Model.
- [32] Tresierra, A y Carrasco, C. 2016. Valorización de opciones reales: modelo Ornstein-Uhlenbeck. Journal of Economics, Finance and Administrative Science 21 (2016) 56-62.
- [33] Thrun, S., Burgard, W. 2006. Probabilistic Robotics. Intelligent Robotics and Autonomous Agents series Edition. Massachusetts Institute of Technology
- [34] Wets, R y Rios, I. 2015.Modeling and estimating commodity prices: copper. Springer.