



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Modelos Dinámicos Lineales Aplicados a
Series de Tiempo

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

PRESENTA:

ALEXANDRO MAYORAL TERÁN

DIRECTOR DE TESIS:

M. en C. JOSÉ SALVADOR ZAMORA MUÑOZ



2013



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Índice general

Introducción	1
I Introducción a los Modelos Dinámicos Lineales	5
1 Modelos Dinámicos	6
1.1 Modelos de Espacio de Estados	6
1.2 Modelos Dinámicos Lineales	9
1.2.1 DLM en R	10
II Estimación en los Modelos Dinámicos Lineales	14
2 Estimación y Predicción de los Estados	15
2.1 Filtración	15
2.2 Suavizamiento (Análisis Retrospectivo)	21
2.3 Pronóstico	24
2.4 Procesos de Innovación	28
III Análisis de Series de Tiempo Univariadas mediante DLM	32
3 Representación alternativa de los ARIMA	33
3.1 Representación de un ARIMA mediante un DLM	35
3.2 Modelos para la Tendencia	38
3.3 Modelos para la Estacionalidad	41
3.4 Modelos de Regresión	42

IV	Implementación de los DLM	46
4	DLM en la modelación de rendimientos	47
4.1	Conceptos básicos	47
4.1.1	Línea del Mercado de Capitales y Línea del Mercado de Títulos	50
4.2	Implementación	52
4.2.1	Modelo de Regresión Simple	52
4.2.2	Modelo de Regresión Simple Dinámico	55
4.2.3	Estimación por Máxima Verosimilitud	56
4.2.4	Estimación por MCMC	60
V	Conclusiones	77
	Apéndice A	80
	Apéndice B	83

Introducción

En este trabajo se presenta de manera introductoria la modelación de series de tiempo, usando los Modelos Dinámicos Lineales (DLM, por sus siglas en inglés), ya que estos modelos poseen una mayor flexibilidad en el tratamiento de series de tiempo no estacionarias, tienen una interpretación más sencilla que los reconocidos Modelos Autorregresivos Integrados de Medias Móviles (ARIMA, por sus siglas en inglés) y una mayor “libertad” en su implementación.

Este trabajo está determinado en esencia por dos bloques. En el primero, se desarrollan los métodos estadísticos necesarios para el análisis de series de tiempo mediante los DLM, para ello, se presentan los conceptos básicos para la comprensión del modelo. En este apartado se explicará el modelo de manera teórica, empezando con su representación, así como las herramientas que se necesitan para ello. Posteriormente se presentan las especificaciones del modelo, como caso particular, también observamos que los ARIMA se pueden considerar como un caso particular de los DLM. También se incluye la metodología necesaria para la estimación de los parámetros del modelo. Cabe mencionar que en la modelación de algún evento de interés, es común encontrarse con parámetros desconocidos, al respecto existen distintos enfoques para estimarlos, ya sea por el método de mínimos cuadrados generalizados, por máxima verosimilitud, o mediante un enfoque Bayesiano, en nuestro caso utilizaremos los dos últimos.

El segundo consta de la implementación de la metodología presentada en el primer bloque, para ello se utiliza el software estadístico R. En este apartado nos enfocaremos únicamente en una de las aplicaciones del modelo, en particular, en los modelos dinámicos de regresión. Para darle un contexto, utilizaremos lo que en finanzas conocen como los Modelos de Valoración de Activos Financieros (CAPM, por sus siglas en inglés) por lo que también se incluyeron conceptos básicos para el entendimiento de este modelo, así como una breve explicación del mismo. A partir de lo cual relajamos los supuestos de este modelo (CAPM) mediante los mencionados modelos dinámicos lineales. Los datos utilizados en este apartado se obtuvieron de internet¹.

¹<http://finance.yahoo.com>.

Una vez implementado el modelo se finaliza el trabajo con las conclusiones que se obtuvieron del modelo y de su implementación.

Antes de proceder con el apartado teórico de este trabajo comencemos con comentar una de las importancias de comprender las series de tiempo y su relevancia en nuestro entorno. Para ello es clave darnos cuenta que nuestra vida está conformada por una serie de eventos y situaciones, en las cuales tenemos que tomar decisiones que afectan nuestro futuro. Esto no es algo que se limite únicamente a nuestra vida personal, sino que trasciende a otros ámbitos en los que nos desenvolvemos, como en el familiar, en el que un padre tiene que decidir a dónde llevará a sus hijos, si a un parque de diversiones o a un museo. Igualmente en el empresarial, consideremos una tienda, que tiene que elegir qué medidas tomar para el crecimiento de la misma. Así como en estos casos, existen muchos otros donde el objetivo es buscar la mejor opción entre diversas alternativas.

Uno de los aspectos importantes a considerar en la toma de decisiones, es la visión de posibles escenarios futuros. Aquí se encuentran dos aspectos importantes, uno es la *predicción*, con la que se intenta “*calcular*” algún hecho futuro a través del análisis de los datos disponibles, lo cual no es una tarea sencilla. El otro punto importante es el *tiempo*. Mucha de la información de la que disponemos varía a través del tiempo. Una gran cantidad de información que se usa para la planeación y la toma de decisiones, se registra a intervalos de tiempo regulares, a este conjunto de datos se les conoce como *series de tiempo*.

El problema de las predicciones siempre ha estado presente en la teoría de procesos estocásticos y series de tiempo. Kolmogorov (1941) estudió el problema en los procesos estocásticos discretos estacionarios, Wiener (1949) estudió los procesos estocásticos continuos, reduciendo el problema de la predicción a la solución de la también llamada ecuación integral de Wiener-Hopf, pero aún estaba sujeta a importantes limitaciones teóricas y prácticas. Una nueva vista del problema la dio Kalman (1960), usando la representación Bode-Shannon de procesos aleatorios y “estados de transición” un método de análisis de sistemas dinámicos. Su solución, conocida también como filtro de Kalman, se puede aplicar tanto a procesos aleatorios estacionarios como no estacionarios, aunque fue reconocida en la literatura estadística hasta tiempo después.

Los Modelos de Espacio de Estados aparecieron en la literatura de series de tiempo en los 70's, y se establecieron durante los 80's, han tenido una expansión rápida debido, por una parte, a los diversos campos de estudio en que se pueden aplicar, y por otra, al desarrollo de herramientas computacionales, ya que uno de los problemas que se presentaba en el modelo anteriormente se debía, en mayor medida, a este tipo de problemas, que en su momento eran difíciles de

resolver. Hoy en día esas dificultades se están enfrentando por medio de los métodos de Monte Carlo con el enfoque Bayesiano, permitiendo el manejo de situaciones más complejas.

Un caso particular de los Modelos de Espacio de Estados son los Modelos Dinámicos Lineales, también conocidos como modelos de espacio de estados lineales. Éstos se empezaron a desarrollar a principios de los años 60's para el monitoreo y control de sistemas dinámicos. El análisis de series de tiempo usando modelos dinámicos lineales fue ampliamente desarrollado en los 1970-80's.

Los Modelos Dinámicos Lineales están basados en la idea de interpretar la salida de un sistema dinámico. En ese sentido una serie de tiempo se considera la salida de un sistema dinámico que es perturbado por errores aleatorios. He aquí el punto clave, ya que esta interpretación nos permite considerar a una serie de tiempo como la integración de varios componentes, tales como la tendencia y estacionalidad, entre otros. Además de plantear una estructura probabilística, esto debido a los errores aleatorios. Así, los problemas de estimación y predicción se resuelven por medio del cálculo recursivo de la distribución condicional de las cantidades de interés, dada la información disponible. En este sentido, es natural pensar su tratamiento mediante el enfoque Bayesiano.

Cabe recordar que el análisis de series de tiempo no estacionarias mediante modelos ARIMA, requiere al menos una transformación preliminar de los datos para volverla estacionaria. En ese sentido es donde se encuentra una de las principales ventajas de los Modelos de Espacio de Estados, que se pueden aplicar a series de tiempo no estacionarias sin requerir alguna transformación preliminar de los datos, esto debido a la forma en que se plantean los modelos dinámicos lineales.

Parte I

Introducción a los Modelos Dinámicos Lineales

Capítulo 1

Modelos Dinámicos

En este capítulo se abordan los conceptos sobre los que se sustenta la teoría de los Modelos Dinámicos Lineales. Antes de iniciar con el planteamiento del modelo haremos unas especificaciones en la notación que se utilizará. Debido a que haremos frecuentemente referencia a sucesiones finitas tanto de observaciones, como de sus posibles valores, se considerará la siguiente notación; $Y_{s:t}$, denotará la sucesión de observaciones entre los periodos de tiempo s a t , incluyendo tanto al momento s como a t , análogamente, $y_{s:t}$, representará los posibles valores para estas observaciones. Las funciones de densidad de probabilidad, se denotarán por $\pi(\cdot)$, donde el argumento nos dirá a qué distribución nos estamos refiriendo, por ejemplo, $\pi(y)$ denota la densidad de Y .

1.1 Modelos de Espacio de Estados

Un *proceso estocástico* es una familia de variables aleatorias asociadas a un conjunto índice de números reales, tal que a cada elemento del conjunto índice le corresponda una y sólo una variable, es decir, $\{Y_t; t \in T\}$, donde T es el conjunto índice y Y_t es la variable aleatoria correspondiente al elemento t de T . Con base en lo anterior, una *serie de tiempo* es un conjunto de observaciones (y_1, \dots, y_t) , generadas por un proceso estocástico, cuyo conjunto índice se toma con relación al tiempo. En esta tesis se trabaja con series de tiempo donde el conjunto índice T es un conjunto finito o infinito numerable, también conocidas como series de tiempo discretas, además de considerar que las observaciones se realizan a intervalos de tiempo iguales.

Sabemos que el comportamiento de un proceso $\{Y_t\}_{t \geq 1}$ queda completamente determinado por la distribución conjunta de las variables (Y_1, \dots, Y_t) , pero obtener dicha distribución es una tarea complicada sin la suposición de independencia o intercambiabilidad, y estas son difíciles de sostener debido a que harían esencialmente al tiempo irrelevante, lo que no es lógico en una serie de tiempo.

Una forma de abordar el problema con los modelos de espacio de estados, es hacer uso de la propiedad de Markov para facilitar la tarea de especificar la distribución conjunta de las variables Y'_t 's, pero suponer que nuestra serie cumple la propiedad tampoco tiene mucho sentido, por lo que la solución que se ocupa en los Modelos de Espacio de Estados, es considerar otro proceso auxiliar $\{\theta_t\}_{t \geq 1}$ que nos ayude a obtener la distribución que buscamos, para ello este proceso debe cumplir la propiedad de Markov, además de estar relacionado con la serie, esta relación queda especificada cuando definamos el modelo, a este proceso auxiliar también se le llama *proceso de estados*, este nombre tiene que ver con el hecho de que en los Modelos de Espacio de Estados se liga la idea de un *sistema dinámico*, que a grandes rasgos es un conjunto de objetos relacionados entre sí que cambian con el tiempo, donde se dice que un sistema dinámico queda determinado por su *estado*, con el análisis de series de tiempo. Ya que se considera a una serie de tiempo como el resultado de la interacción de estos objetos, es decir, como el resultado del sistema dinámico, de ahí el motivo por el que al proceso auxiliar que estamos considerando, se le conozca como proceso de estados, ya que este nos ayudará a determinar nuestro sistema.

Propiedad de Markov: Se dice que una sucesión de variables aleatorias $\{Y_t\}_{t \geq 1}$ es una *Cadena de Markov* si $\forall t > 1$:

$$\pi(y_t \mid y_{1:t}) = \pi(y_t \mid y_{t-1}).$$

donde $\pi(y)$, es la distribución marginal de Y . Esto significa que una sucesión $\{Y_t\}_{t \geq 1}$ es una Cadena de Markov, si la información que se tiene acerca de la variable Y_t de las observaciones al tiempo $t - 1$, depende de que se posee solamente de la observación y_{t-1} , bajo esta definición se puede escribir la distribución conjunta de las variables (Y_1, \dots, Y_t) como sigue:

$$\pi(y_{1:t}) = \pi(y_1) \prod_{j=2}^t \pi(y_j \mid y_{j-1}).$$

Apartir de estas ideas surge la definición de los modelos de espacio de estados: Un *modelo de espacio de estados* consiste de los procesos real valuados, $\{\theta_t : t = 0, 1, \dots\} \in \mathbb{R}^p$ y $\{y_t : t = 0, 1, \dots\} \in \mathbb{R}^m$ que satisfacen:

(A.1) $\{\theta_t\}$ es una cadena de Markov.

(A.2) Condicionando sobre θ_t las Y'_t 's son independientes, Y_t depende de θ_t

solamente.

De lo anterior se nota que el modelo queda completamente especificado por la distribución inicial o distribución a priori, $\pi(\theta_0)$, y las densidades condicionales $\pi(\theta_t | \theta_{t-1})$ y $\pi(y_t | \theta_t)$, esto es $\forall t > 0$:

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \prod_{j=1}^t \pi(\theta_j | \theta_{j-1}) \pi(y_j | \theta_j).$$

Con lo que observamos que se pueden, ya sea condicionando o calculando las marginales, obtener las distribuciones que sean de interés.

El “flujo de la información” o estructura de dependencia en el Modelo de Espacio de Estados se puede representar mediante la Figura 1.1 que es un caso particular de una gráfica acíclica, la cual puede usarse para deducir las propiedades de independencia condicional de las variables aleatorias del modelo. De hecho, se puede probar que dos conjuntos de variables, A y B, son condicionalmente independientes dado un tercer conjunto de variables C, si y sólo si C separa A y B, es decir, que no existe una trayectoria que conecta una variable en A con otra variable en B sin que pase a través de C¹. Por ejemplo, usando esta propiedad podemos demostrar que Y_t y $(\theta_{0:t-1}, Y_{1:t-1})$ son condicionalmente independientes dado θ_t , por lo antes mencionado. La demostración consistiría en mostrar que cualquier trayectoria que une a Y_t con algún elemento de $(\theta_{0:t-1}, Y_{1:t-1})$, tiene que pasar a través de θ_t , por lo que se tendría que:

$$\pi(y_t | \theta_{0:t-1}, Y_{1:t-1}) = \pi(y_t | \theta_t)$$

De igual manera, podemos demostrar que θ_t y $(\theta_{0:t-2}, Y_{1:t-1})$ son condicionalmente independientes dado θ_{t-1} , es decir:

$$\pi(\theta_t | \theta_{0:t-2}, Y_{1:t-1}) = \pi(\theta_t | \theta_{t-1})$$

¹Para una introducción a las gráficas acíclicas y la demostración consultar: Cowell (1999), *Probabilistic networks and expert systems*, Springer.

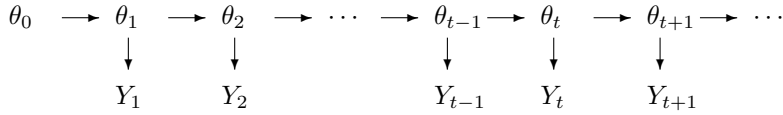


Figura 1.1: Gráfica Acíclica: Estructura de Dependencia de un modelo de espacio de estados

1.2 Modelos Dinámicos Lineales

Uno de los casos particulares y más importantes de los Modelos de Espacio de Estados son los Modelos de Espacio de Estados Gaussianos Lineales, también conocidos como Modelos Dinámicos Lineales. Un Modelo Dinámico Lineal, queda especificado por una distribución a priori Normal p -dimensional, para el estado del vector al tiempo $t = 0$, es decir;

$$\theta_0 \sim \mathcal{N}_p(m_0, C_0) \tag{1.1a}$$

junto con el par de ecuaciones para cada tiempo $t \geq 1$,

$$Y_t = F_t \theta_t + v_t, \quad v_t \sim \mathcal{N}_m(0, V_t) \tag{1.1b}$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}_p(0, W_t) \tag{1.1c}$$

donde Y_t es un vector de orden m , θ_t es un vector de orden p , G_t y F_t son matrices conocidas de orden $p \times p$ y $m \times p$, respectivamente, y $\{v_t\}_{t \geq 1}$ y $\{w_t\}_{t \geq 1}$ son dos sucesiones de variables aleatorias independientes con distribución Normal, con media cero y varianzas $\{V_t\}_{t \geq 1}$ y $\{W_t\}_{t \geq 1}$, respectivamente. La ecuación (1.1b) es conocida como la *ecuación de las observaciones*, mientras que la ecuación (1.1c) se le conoce como la *ecuación de los estados* o *sistema de ecuaciones*. Se supone además, que θ_0 es independiente de $\{v_t\}$ y $\{w_t\}$. Con lo anterior podemos mostrar por qué los DLM son un caso particular de los Modelos de Espacios de Estados, esto es debido a que en los Modelos de Espacio de Estados, se puede especificar cualquier otra distribución a priori distinta a la Normal, junto con las ecuaciones:

$$\begin{aligned} Y_t &= h_t(\theta_t, v_t), \\ \theta_t &= g_t(\theta_{t-1}, w_t), \end{aligned}$$

para cualquier función g_t y h_t , y en los DLM, la distribución a priori es Normal, además de que g_t y h_t son funciones lineales, en los modelos de espacio de estados no existe esta restricción.

Cuando las matrices G_t , F_t y las matrices de covarianza V_t y W_t son constantes, se dice que el modelo es *invariante con respecto al tiempo*. Los DLM pueden ser considerados como una generalización de los modelos de regresión lineal, porque permiten la variación de los coeficientes de regresión a través del tiempo, recordando que el modelo de regresión lineal simple describe la relación de la variable aleatoria Y y la variable no aleatoria explicativa x , mediante la siguiente relación:

$$Y_t = \theta_1 + \theta_2 x_t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

Donde se considera a (Y_t, x_t) , $t = 1, 2, \dots$ como observaciones a través del tiempo. Al permitir la variación de los parámetros de regresión, se puede modelar la relación no lineal entre x y y , además de permitir el estudio de los cambios estructurales del modelo. Con esta idea se puede formular un modelo dinámico de regresión lineal simple de la siguiente forma:

$$Y_t = \theta_{t,1} + \theta_{t,2} x_t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_t^2)$$

junto con la ecuación de estados, que describe la evolución del sistema:

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \stackrel{iid}{\sim} \mathcal{N}(0, W_t)$$

que es un DLM con $F_t = [1, x_t]$ y estados $\theta_t = (\theta_{t,1}, \theta_{t,2})'$. Note que si $G_t = I$, la matriz identidad, $\sigma_t^2 = \sigma^2$ y $w_t = 0$ para cada t , se tiene el modelo de regresión lineal simple.

1.2.1. DLM en R

Como mencionamos inicialmente, el software que utilizaremos para realizar el análisis de la serie de tiempo es *R*; para la representación de los DLM, ocuparemos el paquete **dlm**, el cual permite ésto por medio de listas con una serie

de atributos, los cuales forman un objeto llamado “*d1m*”. Estos objetos pueden representar un DLM constante o que varíe en el tiempo. Ahora bien, un DLM constante queda especificado una vez que se determinan los valores de las matrices **F**, **V**, **G**, **W**, C_0 y el vector m_0 . En R estos valores se almacenan, en un objeto *d1m*, mediante los elementos **FF**, **V**, **GG**, **W**, **CO** y **m0**, respectivamente. Como ejemplo, consideremos un modelo que se utiliza en series de tiempo univariadas $\{Y_t : t = 1, 2, \dots\}$ conocido como *random walk plus noise* también llamado *modelo polinomial de primer orden*, definido por las ecuaciones:

$$Y_t = \mu_t + v_t, \quad v_t \sim \mathcal{N}(0, V) \quad (1.2a)$$

$$\mu_t = \mu_{t-1} + w_t, \quad W_t \sim \mathcal{N}(0, W) \quad (1.2b)$$

Donde podemos observar que $F_t = G_t = [1]$, y los únicos parámetros del modelo son la observación y la evolución de las varianzas V y W , que usualmente son estimadas usando la información disponible y aplicando, ya sea el método de máxima verosimilitud o técnicas bayesianas. Ahora veamos un pequeño fragmento de código para definir un modelo polinomial de primer orden, con $V = 0.5$ y $W = 0.05$:

```
----- Código en R -----
> # Definimos un modelo polinomial de primer orden, con V = 0.5
> # y W = 0.05
> mod.d1m <- d1m(FF = 1, V = .5, GG = 1, W = .05,
                 m0 = 0, CO = 100)
> # Para obtener los componentes individuales del modelo,
> # es de la siguiente forma:
> FF(mod.d1m)
      [,1]
[1,]    1
>
> W(mod.d1m)
      [,1]
[1,] 0.05
>
> # 0 también se puede asignar un valor a una componente:
> V(mod.d1m) <- 0.8
-----
```


Observemos que la media y la varianza de la distribución a priori de θ_0 tienen que especificarse. El paquete *dlm* proporciona otras funciones para la creación de DLMs estándar, más adelante veremos algunas de estas funciones. Aunque para modelos más complicados siempre se puede usar simplemente la función general *dlm*. Consideremos ahora un DLM variable en el tiempo, entonces al menos una de las entradas de F_t, V_t, G_t o W_t cambia con el tiempo. Se puede pensar a las entradas de estas matrices como series de tiempo, es decir las matrices se almacenan como series de tiempo multivariadas. Por tanto, en lugar de almacenar completamente las matrices F_t, V_t, G_t y W_t , se tiene que almacenar en una especie de plantilla o contenedor para cada una las entradas invariantes, y las entradas que varíen en el tiempo, en una matriz separada, esta matriz es la componente X, una matriz de dimensión $n \times m$, donde n es el número de observaciones, que también se encuentra definida en un objeto *dlm*. Para lograr un manejo adecuado del modelo pensando de esta forma, se tiene entonces que saber qué entrada de qué matriz corresponde a qué columna de X, por lo que se utilizan las componentes llamadas JFF, JV, JGG y JW, también definidas en el objeto *dlm*. Estas son matrices de enteros con dimensión igual a las matrices FF, V, GG y W, respectivamente. Entonces, por ejemplo la entrada (i,j) de JFF es cero si la correspondiente entrada de FF es invariante, ó k si el valor de $F_s[i,j]$ es $X[s,k]$ $s \geq 1$, es decir, k si los valores para los distintos tiempos de $F[i,j]$, son almacenados en $X[,k]$, de manera similar se usan JV, JGG y JW.

Cabe hacer hincapié que se debe tener cuidado en la declaración de las matrices y considerar los valores predeterminados que tienen las funciones, así como con qué tipo de modelo se trabajará.

Parte II

Estimación en los Modelos Dinámicos Lineales

Capítulo 2

Estimación y Predicción de los Estados

Antes de comenzar con la estimación de los estados, es importante mencionar que una de las principales dificultades en muchos problemas estadísticos es la construcción de los modelos, sin ser la excepción los modelos de espacio de estados. Se desea crear un modelo que sea claro y no tan complejo, donde los estados tengan un significado intuitivo, y que el conocimiento que se tiene sobre el evento de interés pueda reflejarse en probabilidades. Por el momento consideremos que se tiene construido el modelo, partiendo de eso, uno de los objetivos es la inferencia de los estados del sistema, para las predicciones. Usando la información disponible, como ya se ha mencionado anteriormente, la estimación y predicción se resuelve con métodos computacionales, para obtener las distribuciones condicionales de interés.

Supongamos que disponemos de información hasta el tiempo t (i.e. $y_{1:t}$), y estamos interesados en hacer inferencias del vector de estados, entonces deseamos obtener la densidad condicional $\pi(\theta_s | y_{1:t})$, para ello tendríamos que distinguir tres diferentes casos en los que podríamos estar interesados: *filtración*; cuando $s = t$, *pronóstico*; cuando $s > t$ y *suavizamiento* o análisis retrospectivo; cuando $s < t$.

2.1 Filtración

En el caso de filtración, se considera que la información se puede ir actualizando continuamente, apartir de ello el problema se resuelve obteniendo la densidad condicional de θ_t dado $y_{1:t}$ (i.e. $\pi(\theta_t | y_{1:t})$), la fórmula que permite resolver esta cuestión, que es actualizar la inferencia del vector de estados con los datos disponibles, o conforme nueva información es obtenida, se le conoce como: *Filtro de Kalman*, y su nombre se debe a que, con los datos disponibles, se pasa de la densidad $\pi(\theta_t | y_{1:t})$ a $\pi(\theta_{t+1} | y_{1:t+1})$, a través de la densidad $\pi(\theta_{t-1} | y_{1:t-1})$,

una especie de filtro, la fórmula tiene además la ventaja de poder aplicarse de manera recursiva, de ahí la idea de su implementación computacional. Los pasos a seguir para resolver el problema de filtrado de una manera general, son los siguientes:

(i) Se obtiene la distribución predictiva del estado θ_t , dadas las observaciones $y_{1:t-1}$, (i.e. $\pi(\theta_t | y_{1:t-1})$), usando la densidad “filtrada” $\pi(\theta_{t-1} | y_{1:t-1})$ y la distribución condicional $\pi(\theta_t | \theta_{t-1})$;

(ii) Se obtiene la distribución predictiva de la siguiente observación y_t dadas las observaciones anteriores $y_{1:t-1}$, (i.e. $\pi(y_t | y_{1:t-1})$);

(iii) Se obtiene la distribución filtrada $\pi(\theta_t | y_{1:t})$ usando la regla de Bayes con $\pi(\theta_t | y_{1:t-1})$ como la distribución a priori y $\pi(y_t | \theta_t)$ como la de verosimilitud.

Cabe mencionar que partiendo de la distribución inicial $\theta_0 \sim \pi(\theta_0)$, los pasos anteriores se pueden realizar de manera recursiva, para $t = 1, 2, \dots$. Esta es una de las ventajas de los modelos de espacio de estados, que se debe a la estructura Markoviana de los estados y el supuesto de independencia condicional de las observaciones.

Filtro de Kalman para los DLM

En general, aunque el filtro de Kalman es de gran utilidad, la obtención de las distribuciones condicionales relevantes necesarias en el proceso no es una tarea sencilla, principalmente en el caso general, aunque en el caso de los DLM, esta labor se reduce considerablemente debido a las suposiciones de las distribuciones Normales. Se puede probar que el vector aleatorio $(\theta_0, \theta_1, \dots, \theta_t, Y_1, \dots, Y_t)$ tiene una distribución Normal para $t \geq 1$, usando el hecho de la estructura Markoviana del modelo, de lo que se sigue que, las distribuciones marginales y condicionales tienen también distribución Normal. La forma en que se resuelve el problema de filtración en el caso de los DLM, como se mencionó anteriormente, es a través del Filtro de Kalman, que enunciaremos a continuación:

Proposición 2.1 (Filtro de Kalman.) *Consideremos el DLM especificado en 1.1, y supongamos:*

$$\theta_{t-1} | y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$$

Entonces los siguientes enunciados se cumplen:

(i) La distribución predictiva (del primer paso) de θ_t dado $y_{1:t-1}$ es Gaussiana con parámetros:

$$\begin{aligned} a_t &= E(\theta_t \mid y_{1:t-1}) = G_t m_{t-1}, \\ R_t &= \text{Var}(\theta_t \mid y_{1:t-1}) = G_t C_{t-1} G_t' + W_t. \end{aligned} \quad (2.1a)$$

(ii) La distribución predictiva (del primer paso) de Y_t dado $y_{1:t-1}$ es Gaussiana con parámetros:

$$\begin{aligned} f_t &= E(Y_t \mid y_{1:t-1}) = F_t a_t, \\ Q_t &= \text{Var}(Y_t \mid y_{1:t-1}) = F_t R_t F_t' + V_t. \end{aligned} \quad (2.1b)$$

(iii) La distribución filtrada de θ_t dado $y_{1:t}$ es Gaussiana con parámetros:

$$\begin{aligned} m_t &= E(\theta_t \mid y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t, \\ C_t &= \text{Var}(\theta_t \mid y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t, \end{aligned} \quad (2.1c)$$

donde $e_t = Y_t - f_t$, es el error de predicción.

DEMOSTRACIÓN. Sabemos que el vector $(\theta_0, \theta_1, \dots, \theta_t, Y_1, \dots, Y_t)$ tiene una distribución conjunta Normal, al igual que las distribuciones condicionales y marginales. De manera más general, cualquier subconjunto de este vector tiene distribución Normal, así, como cualquier subconjunto del vector condicionado a otro subconjunto, entonces, la distribución predictiva y la de filtración son Normales, por lo que, para que estas queden determinadas, basta con obtener sus medias y varianzas, que es lo que se hará a continuación:

(i) Sabemos que $\theta_{t-1} \mid y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$, entonces: sea $\theta_t \mid y_{1:t-1} \sim \mathcal{N}(a_t, R_t)$, se obtendrán a_t y R_t :

$$\begin{aligned} a_t &= E(\theta_t \mid y_{1:t-1}) = E(E(\theta_t \mid \theta_{t-1}, y_{1:t-1}) \mid y_{1:t-1}) \\ &= E(E(G_t \theta_{t-1} + w_t \mid \theta_{t-1}, y_{1:t-1}) \mid y_{1:t-1}) \\ &= E(G_t \theta_{t-1} \mid y_{1:t-1}) = G_t m_{t-1}. \end{aligned}$$

y

$$\begin{aligned}
R_t &= \text{Var}(\theta_t \mid y_{1:t-1}) = \text{Var}(E(\theta_t \mid \theta_{t-1}, y_{1:t-1}) \mid y_{1:t-1}) \\
&\quad + E(\text{Var}(\theta_t \mid \theta_{t-1}, y_{1:t-1}) \mid y_{1:t-1}) \\
&= \text{Var}(E(G_t \theta_{t-1} + w_t \mid \theta_{t-1}, y_{1:t-1}) \mid y_{1:t-1}) \\
&\quad + E(\text{Var}(G_t \theta_{t-1} + w_t \mid \theta_{t-1}, y_{1:t-1}) \mid y_{1:t-1}) \\
&= \text{Var}(G_t \theta_{t-1} \mid y_{1:t-1}) + E(w_t) \\
&= G_t C_{t-1} G_t' + W_t.
\end{aligned}$$

(ii) Sea $Y_t \mid y_{1:t-1} \sim \mathcal{N}(f_t, Q_t)$, se obtendrán f_t y Q_t :

$$\begin{aligned}
f_t &= E(Y_t \mid y_{1:t-1}) = E(E(Y \mid \theta_t, y_{1:t-1}) \mid y_{1:t-1}) \\
&= E(E(F_t \theta_t + V_t \mid \theta_t, y_{1:t-1}) \mid y_{1:t-1}) \\
&= E(F_t \theta_t \mid y_{1:t-1}) = F_t a_t.
\end{aligned}$$

y

$$\begin{aligned}
Q_t &= \text{Var}(Y_t \mid y_{1:t-1}) = \text{Var}(E(Y_t \mid \theta_t, y_{1:t-1}) \mid y_{1:t-1}) \\
&\quad + E(\text{Var}(Y_t \mid \theta_t, y_{1:t-1}) \mid y_{1:t-1}) \\
&= \text{Var}(E(F_t \theta_t + v_t \mid \theta_t, y_{1:t-1}) \mid y_{1:t-1}) \\
&\quad + E(\text{Var}(F_t \theta_t + v_t \mid \theta_t, y_{1:t-1}) \mid y_{1:t-1}) \\
&= \text{Var}(F_t \theta_t \mid y_{1:t-1}) + E(v_t) \\
&= F_t R_t F_t' + V_t.
\end{aligned}$$

(iii) En este caso se aplica el teorema de Bayes, utilizando como distribución a priori a $\pi(\theta_t \mid y_{1:t-1})$, y como verosimilitud a $\pi(y_t \mid \theta_t)$, lo cual lleva a un problema de inferencia bayesiana. Cabe resaltar que en el caso de los DLM, el problema se puede plantear como un modelo de regresión lineal dinámico:

$$Y_t = F_t \theta_t + v_t \quad v_t \sim \mathcal{N}(0, V_t),$$

con parámetro de regresión θ_t , con distribución a priori $\mathcal{N}(0, V_t)$, además se supondrá que V_t es conocida, lo que conduce al resultado¹:

$$\theta_t \mid y_{1:t} \sim \mathcal{N}(m_t, C_t),$$

¹Para ver esta parte de la demostración completa, véase el Apéndice A.

donde:

$$m_t = a_t + R_t F_t' Q_t^{-1} (Y_t - F_t a_t)$$

$$C_t = R_t - R_t F_t' Q_t^{-1} F_t R_t$$

■

Filtrado con información faltante

Cuando se realiza el análisis de alguna serie de tiempo, o cualquier tipo de información, es común que los datos no estén completos. En lo que se refiere a series de tiempo, se pueden identificar dos tipos de casos, el primero se refiere a una pérdida total del vector de observaciones en algún punto del tiempo t , mientras que el segundo implica una pérdida parcial, es decir, sólo se cuenta con algunos componentes del vector de observaciones. En el caso de series de tiempo univariadas sólo se puede hablar de una falta de información o no, es decir, la pérdida total de información. Ahora, la forma de enfretar este problema en los modelos de espacio de estados, debido a su forma recursiva, es la siguiente, si la observación al tiempo t falta, entonces y_t no aporta ninguna información, así que

$$\pi(\theta_t | y_{1:t}) = \pi(\theta_t | y_{1:t-1})$$

Esto implica, en el caso de los DLM, que debido a que $\theta_t | y_{1:t} \sim \mathcal{N}(a_t, R_t)$ lo único que se necesita hacer es actualizar $m_t = a_t$ y $C_t = R_t$, y en $t + 1$ el algoritmo se sigue realizando de manera usual, siempre que y_{t+1} no falte. Formalmente, en un DLM, el que no se disponga de y_t es lo mismo que, $F_t = 0$ ó $V_t = \infty$, en el primer caso y_t no está vinculado a θ_t de ninguna forma, y en el segundo, la observación es tal que su información es poco confiable. Viéndolo de manera más exacta en el algoritmo, consideremos un DLM, y suponga que faltan algunas de las componentes de y_t , entonces consideramos a \tilde{y}_t como el vector que contiene sólo las componentes en las que se posee información, para esto, denotemos a \tilde{m}_t como la dimensión del vector \tilde{y}_t y consideremos a la matriz M_t de orden \tilde{m}_t por m , que se obtiene al remover de la matriz identidad las filas correspondientes a los componentes faltantes de y_t , así que $\tilde{y}_t = M_t y_t$, por lo que el hecho de tener el vector de observaciones \tilde{y}_t implica que en la actualización de la distribución a priori $\mathcal{N}(a_t, R_t)$ a la posteriori $\mathcal{N}(m_t, C_t)$, se tiene que considerar la siguiente ecuación de observaciones:

$$\tilde{y}_t = \tilde{F}_t' \theta_t + \tilde{v}_t \quad \tilde{v}_t \sim \mathcal{N}(0, \tilde{V}_t),$$

con $\tilde{F}_t = M_t F_t$ y $\tilde{V}_t = M_t V_t m'_t$, por lo que simplemente lo que hacemos al calcular el filtro de Kalman, es remplazar a F_t con \tilde{F}_t y a V_t con \tilde{V}_t .

Filtración en R

El filtro de Kalman es implementado mediante la función *dlmFilter*, cuyos argumentos son la información, *y*, en forma de un vector numérico, matriz o serie de tiempo, y el modelo, *mod*, un objeto de clase *dlm*. Los datos de salida producidos por la función incluyen, tanto los datos originales y el modelo, así como las medias de la distribución predictiva y filtrada (componentes *a* y *m*), y la descomposición en valores singulares (SVD, por sus siglas en inglés) de las varianzas R_t y C_t de la distribución predictiva y filtrada (componente *U.R*, *D.R*, *U.C* y *D.C*), además de la componente *f* que provee las predicciones del primer paso. Es útil mencionar que se pueden reconstruir las varianzas de su descomposición en valores singulares mediante la función *dlmSvd2var* implantada en el paquete que estamos describiendo. Por último, la función *dlmFilter* acepta información que contenga *NA*'s, sin necesidad de hacer ajustes. A continuación vemos un pequeño fragmento de código para ver como se usa la función:

```
----- Código en R -----
> # Definimos el filtro de Kalman de la siguiente forma:
> dlm.Filt <- dlmFilter(datos, mod.dlm)
>
> # Podemos observar que "mod.dlm" es un objeto de clase dlm:
> class(mod.dlm)
[1] "dlm"
>
> # Además un tipo de dato que se aceptan son de la clase "xst",
> # que son para series de tiempo:
> class(datos)
[1] "xts" "zoo"
>
> # Como mencionamos, podemos ver la salida que produce el Filtro
> # de Kalman
> ls(dlm.Filt)
[1] "a" "D.C" "D.R" "f" "m" "mod" "U.C" "U.R" "y"
>
> # La función head(), arroja las primeras partes de un vector,
> # matriz, tabla, data frame o función.
```

```

>
> # Podemos acceder a estos campos de la siguiente forma:
> head(dlm.Filt$m)
           [,1]      [,2]
2012-01-03 0.00000 0.0000000
2012-01-04 -0.01513 0.0000778
2012-01-05 -0.01531 -0.0343681
2012-01-06 -0.01537 -0.1807432
2012-01-09 -0.01609 -0.2605561
2012-01-09 -0.01706 -0.2010143

```

2.2 Suavizamiento (Análisis Retrospectivo)

En series de tiempo usualmente se tienen observaciones $\{Y_t\}$ en un periodo de tiempo T , con $t = 1, \dots, T$. El problema de suavizamiento, o análisis retrospectivo, consiste en la estimación de algún estado θ_t o “bloques” de estados, por ejemplo $(\theta_1, \dots, \theta_t)$, dada la información disponible, y_1, \dots, y_T , de ahí el nombre de análisis retrospectivo. Este problema es frecuente en el análisis de series de tiempo, donde se busca reconstruir el comportamiento del sistema a partir de los datos disponibles. Para ello se utiliza un algoritmo recursivo regresivo, con el que se obtiene la distribución condicional $\pi(\theta_t | y_{1:T})$, para cualquier $t < T$, empezando con la distribución filtrada $\pi(\theta_T | y_{1:T})$ y estimando de forma retrospectiva todos los estados. A continuación presentamos el algoritmo para el caso de los DLM:

Proposición 2.2 (Kalman smoother.) *Consideremos un DLM especificado en 1.1, si:*

$\theta_{t+1} | y_{1:T} \sim \mathcal{N}(s_{t+1}, S_{t+1})$, entonces $\theta_t | y_{1:T} \sim \mathcal{N}(s_t, S_t)$, donde

$$\begin{aligned}
 s_t &= m_t + C_t G'_{t+1} R_{t+1}^{-1} (s_{t+1} - a_{t+1}) \\
 S_t &= C_t - C_t G'_{t+1} R_{t+1}^{-1} (R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t
 \end{aligned}$$

DEMOSTRACIÓN. Por las propiedades de la distribución Normal se tiene que $\pi(\theta_t | y_{1:T})$ es normal, por lo que es suficiente con obtener su media y varianza, esto es:

$$s_t = E(\theta_t | y_{1:T}) = E(E(\theta_t | \theta_{t+1}, y_{1:T}) | y_{1:T})$$

y

$$S_t = \text{Var}(\theta_t \mid y_{1:T}) = \text{Var}(E(\theta_t \mid \theta_{t+1}, y_{1:T}) \mid y_{1:T}) \\ + E(\text{Var}(\theta_t \mid \theta_{t+1}, y_{1:T}) \mid y_{1:T})$$

Como ya se mencionó anteriormente, usando la gráfica acíclica dirigida (DAG, la figura 1.1), se puede probar que θ_t y $Y_{t+1:T}$ son condicionalmente independientes dado θ_{t+1} , por lo que $\pi(\theta_t \mid \theta_{t+1}, y_{1:T}) = \pi(\theta_t \mid \theta_{t+1}, y_{1:t})$, además, se puede probar que θ_{t+1} y $Y_{1:t}$ son condicionalmente independientes dado θ_t , lo que permite expresar a la función de verosimilitud como, $\pi(\theta_{t+1} \mid \theta_t, y_{1:t}) = \pi(\theta_{t+1} \mid \theta_t)$. Con lo anterior se puede proceder de manera similar al último inciso de la demostración anterior para el filtro de Kalman, es decir, se nota un modelo de regresión lineal dinámico ($\theta_{t+1} = G_{t+1}\theta_t + w_{t+1}$ $w_t \sim \mathcal{N}(0, W_t)$), con parámetro de regresión θ_t , con distribución a priori $\pi(\theta_t \mid y_{1:t}) \sim \mathcal{N}(m_t, C_t)$, y distribución de verosimilitud,² $\pi(\theta_{t+1} \mid \theta_t) \sim \mathcal{N}((G'_{t+1}W_t^{-1}G_{t+1})^{-1}G'_{t+1}W_t^{-1}\theta_{t+1}, (G'_{t+1}W_t^{-1}))$, lo que conduce a:

$$E(\theta_t \mid \theta_{t+1}, y_{1:t}) = m_t + C_t G'_{t+1} (G_{t+1} C_t G'_{t+1} + W_{t+1})^{-1} (\theta_{t+1} - G_{t+1} m_t) \\ = m_t + C_t G'_{t+1} R_{t+1}^{-1} (\theta_{t+1} - a_{t+1})$$

$$\text{Var}(\theta_t \mid \theta_{t+1}, y_{1:t}) = C_t - C_t G'_{t+1} R_{t+1}^{-1} G_{t+1} C_t,$$

de lo que sigue que:

$$s_t = E(E(\theta_t \mid \theta_{t+1}, y_{1:t}) \mid y_{1:T}) = m_t + C_t G'_{t+1} R_{t+1}^{-1} (s_{t+1} - a_{t+1}) \\ S_t = \text{Var}(E(\theta_t \mid \theta_{t+1}, y_{1:t}) \mid y_{1:T}) + E(\text{Var}(\theta_t \mid \theta_{t+1}, y_{1:t}) \mid y_{1:T}) \\ = C_t - C_t G'_{t+1} R_{t+1}^{-1} G_{t+1} C_t + C_t G'_{t+1} R_{t+1}^{-1} S_{t+1} R_{t+1}^{-1} G_{t+1} C_t \\ = C_t - C_t G'_{t+1} R_{t+1}^{-1} (R_{t+1} - S_{t+1}) G_{t+1} C_t$$

■

Observe que la recursividad depende de la información usando el filtro de Kalman.

Análisis Retrospectivo en R

El caso de suavizamiento, en R es implementado mediante la función *dlmSmooth*, cuyos argumentos son la información, y , en forma de un vector numérico, matriz o serie de tiempo, y el modelo, *mod*, un objeto de clase *dlm*. Los datos de salida

²Para observar el desarrollo, véase el Apéndice A.

producidos por la función son una lista con los componentes, s , las medias de la distribución suavizada y $U.S$, $D.S$, sus varianzas, dadas en términos de su SVD, al igual que en el caso de filtración, se pueden reconstruir las varianzas de su descomposición en valores singulares mediante la función `dlmSvd2var` implantada en el paquete `dlm`, y también la función `dlmSmooth` acepta información que contenga NA 's, sin necesidad de hacer ajustes. A continuación vemos un pequeño fragmento de código para ver como se usa la función:

```
----- Código en R -----
> # Definimos un Suavizamiento de la siguiente forma:
> dlm.Smooth <- dlmSmooth(datos, mod.dlm)
>
> # Podemos observar que "mod.dlm" es un objeto de clase dlm:
> class(mod.dlm)
[1] "dlm"
>
> # Además un tipo de dato que se aceptan son de la clase "xst",
> # que son para series de tiempo:
> class(datos)
[1] "xts" "zoo"
>
> # Como mencionamos, podemos ver la salida que produce el
> # Suavizamiento
> ls(dlm.Smooth)
[1] "D.S" "s" "U.S"
>
> # La función head(), arroja las primeras partes de un vector,
> # matriz, tabla, data frame o función.
>
> # Podemos acceder a estos campos de la siguiente forma:
> head(dlm.Smooth$s)
      [,1] [,2]
      0.004055 1.013
2012-01-03 0.004055 1.013
2012-01-04 0.004055 1.012
2012-01-05 0.004055 1.012
2012-01-06 0.004055 1.012
2012-01-09 0.004055 1.013
-----
```

2.3 Pronóstico

En series de tiempo uno de los objetivos principales es realizar predicciones adecuadas, así que la estimación de los estados es sólo un paso para poder predecir observaciones futuras. El “primer paso” sería la predicción de la siguiente observación, es decir Y_{t+1} , esto, basado en los datos disponibles $y_{1:t}$, es decir, $\pi(Y_{t+1} | y_{1:t})$, esta distribución también recibe el nombre de la predicción del primer paso. La idea es la siguiente, primero se estima el valor del estado θ_{t+1} , luego, basados en esta estimación, se puede obtener la predicción de Y_{t+1} , así sucesivamente. Ahora, la densidad predictiva del estado del primer paso es $\pi(\theta_{t+1} | y_{1:t})$, observemos que está basada en la densidad “filtrada” de θ_t , a partir de la cual se obtiene la densidad predictiva de la observación del primer paso $\pi(y_{t+1} | y_{1:t})$, eso suena bien, aunque se podría estar interesado no solamente en la predicción del siguiente dato, sino en una predicción al tiempo $t+k$, por lo que se tendría que hacer una estimación del sistema al tiempo $t+k$ (θ_{t+k}), para alguna $k \geq 1$. Una forma de hacerlo es realizando “k-pasos”, es decir, realizar el primer paso k-veces, y así, obtener la densidad predictiva $\pi(\theta_{t+k} | y_{1:t})$ para la observación futura al tiempo $t+k$. Cabe mencionar que algo que sucede al hacer que nuestro horizonte de tiempo sea mayor es que nuestra predicción se vuelva más incierta, pero también es cierto que podemos cuantificar esta incertidumbre a través de la función de densidad de probabilidad, para ello, un estimador óptimo del valor de Y_{t+1} es la esperanza condicional $E(Y_{t+1} | y_{1:t})$, que se obtiene minimizando la esperanza condicional del error de predicción al cuadrado. Como función de k , $E(Y_{t+k} | y_{1:t})$ es conocida como la *función de predicción*. En los modelos de espacio de estados, una ventaja es la existencia de un algoritmo que se puede aplicar de forma recursiva, permitiendo la obtención de la predicción de primer paso y conforme se actualiza la información repetir el proceso. Esto supone que es posible realizar dicha actualización, lo cual no siempre sucede.

Dada la naturaleza Markoviana del modelo, la distribución filtrada al tiempo t , sirve como la distribución inicial para la futura evolución del modelo, es decir, la distribución conjunta del presente y futuros estados $(\theta_{t+k})_{k \geq 0}$ y futuras observaciones $(Y_{t+k})_{k \geq 1}$ es la de un modelo de espacio de estados con distribuciones condicionales $\pi(\theta_{t+k} | \theta_{t+k-1})$ y $\pi(y_{t+k} | \theta_{t+k})$, y distribución inicial $\pi(\theta_t | y_{1:t})$, es decir, la información acerca del “futuro” está contenida en esta distribución. En el caso de los DLM, para darnos una idea del funcionamiento del algoritmo que resuelve este problema, recordemos cómo es el flujo de información del modelo. En el siguiente diagrama podemos observar el camino de $Y_{1:t}$ hacia Y_{t+k} y que $Y_{1:t}$ proporciona información sobre θ_t , que a su vez

le proporciona información al estado futuro θ_{t+k} y consecuentemente a Y_{t+k} , además nos damos cuenta de lo que ya se ha mencionado, de que conforme k es mayor, la información que se le proporciona a los pronósticos es menor, haciéndolas más inciertas:

$$\begin{array}{ccccccc}
 \theta_t & \longrightarrow & \theta_{t+1} & \longrightarrow & \dots & \longrightarrow & \theta_{t+k} \\
 \downarrow & & & & & & \downarrow \\
 Y_{1:t} & & & & & & Y_{t+k}
 \end{array}$$

A continuación se enuncia el algoritmo que sirve para obtener la distribución de los pronósticos para los estados y observaciones:

Proposición 2.3 *Considera el DLM especificado en 1.1, y supongamos que: $a_t(0) = m_t$ y $R_t(0) = C_t$. Entonces los siguientes enunciados se cumplen:*

(i) *La distribución predictiva de θ_{t+k} dado $y_{1:t-1}$ es Gaussiana con parámetros:*

$$\begin{aligned}
 a_t(k) &= G_{t+k} a_t(k-1), \\
 R_t(k) &= G_{t+k} R_t(k-1) G'_{t+k} + W_{t+k}.
 \end{aligned}$$

(ii) *La distribución predictiva de Y_{t+k} dado $y_{1:t}$ es Gaussiana con parámetros:*

$$\begin{aligned}
 f_t(k) &= F_{t+k} a_t(k), \\
 Q_t(k) &= F_{t+k} R_t(k) F'_{t+k} + V_{t+k}.
 \end{aligned}$$

Antes de proceder a la demostración definamos los siguientes valores, para $k \geq 1$, definimos:

$$a_{t+k}(k) = E(\theta_{t+k} \mid y_{1:t}), \quad (2.3a)$$

$$R_{t+k}(k) = \text{Var}(\theta_{t+k} \mid y_{1:t}), \quad (2.3b)$$

$$f_{t+k}(k) = E(Y_{t+k} \mid y_{1:t}), \quad (2.3c)$$

$$Q_{t+k}(k) = \text{Var}(Y_{t+k} \mid y_{1:t}). \quad (2.3d)$$

DEMOSTRACIÓN. Dada la formulación de un DLM, sabemos que todas las distribuciones condicionales son Gaussianas, por lo que basta proporcionar las

medias y varianzas. La demostración se hace por inducción, entonces para $k = 1$ se cumple como vimos en la preposición 2.1, ahora para $k > 1$:

$$\begin{aligned}
a_{t+k}(k) &= E(\theta_{t+k} \mid y_{1:t}) = E(E(\theta_{t+k} \mid \theta_{t+k-1}, y_{1:t}) \mid y_{1:t}) \\
&= E(E(G_{t+k}\theta_{t+k-1} + w_{t+k} \mid \theta_{t+k-1}, y_{1:t}) \mid y_{1:t}) \\
&= E(G_{t+k}\theta_{t+k-1} \mid y_{1:t}) = G_{t+k}E(\theta_{t+k-1} \mid y_{1:t}) \\
&= G_{t+k}a_t(k-1),
\end{aligned}$$

$$\begin{aligned}
R_{t+k} &= \text{Var}(\theta_{t+k} \mid y_{1:t}) = \text{Var}(E(\theta_{t+k} \mid \theta_{t+k-1}, y_{1:t}) \mid y_{1:t}) \\
&\quad + E(\text{Var}(\theta_{t+k} \mid \theta_{t+k-1}, y_{1:t}) \mid y_{1:t}) \\
&= \text{Var}(E(G_{t+k}\theta_{t+k-1} + w_{t+k} \mid \theta_{t+k-1}, y_{1:t}) \mid y_{1:t}) \\
&\quad + E(\text{Var}(G_{t+k}\theta_{t+k-1} + w_{t+k} \mid \theta_{t+k-1}, y_{1:t}) \mid y_{1:t}) \\
&= \text{Var}(G_{t+k}\theta_{t+k-1} \mid y_{1:t}) + E(w_{t+k} \mid y_{1:t}) \\
&= G_{t+k}\text{Var}(\theta_{t+k-1} \mid y_{1:t})G'_{t+k} + W_{t+k} \\
&= G_{t+k}R_t(k-1)G'_{t+k} + W_{t+k},
\end{aligned}$$

$$\begin{aligned}
f_t(k) &= E(Y_{t+k} \mid y_{1:t}) = E(E(Y_{t+k} \mid \theta_{t+k}, y_{1:t}) \mid y_{1:t}) \\
&= E(E(F_{t+k}\theta_{t+k} + v_{t+k} \mid \theta_{t+k}, y_{1:t}) \mid y_{1:t}) \\
&= E(F_{t+k}\theta_{t+k} \mid y_{1:t}) = F_{t+k}E(\theta_{t+k} \mid y_{1:t}) \\
&= F_{t+k}a_t(k),
\end{aligned}$$

$$\begin{aligned}
Q_{t+k} &= \text{Var}(Y_{t+k} \mid y_{1:t}) = \text{Var}(E(Y_{t+k} \mid \theta_{t+k}, y_{1:t}) \mid y_{1:t}) \\
&\quad + E(\text{Var}(Y_{t+k} \mid \theta_{t+k}, y_{1:t}) \mid y_{1:t}) \\
&= \text{Var}(E(F_{t+k}\theta_{t+k} + v_{t+k} \mid \theta_{t+k}, y_{1:t}) \mid y_{1:t}) \\
&\quad + E(\text{Var}(F_{t+k}\theta_{t+k} + v_{t+k} \mid \theta_{t+k}, y_{1:t}) \mid y_{1:t}) \\
&= \text{Var}(F_{t+k}\theta_{t+k} \mid y_{1:t}) + E(v_{t+k} \mid y_{1:t}) \\
&= F_{t+k}\text{Var}(\theta_{t+k} \mid y_{1:t})F'_{t+k} + V_{t+k} \\
&= F_{t+k}R_t(k)F'_{t+k} + V_{t+k}.
\end{aligned}$$

■

Pronósticos en R

En el caso de pronósticos, R proporciona los pronósticos del primer paso mediante la función *dlmFilter*, es decir nos proporciona una predicción al tiempo posterior al que tenemos la información, para predicciones con un horizonte de tiempo mayor, existe la opción de usar la función *dlmForecast*, cuyo argumento es un objeto de clase *dlmFiltered* (la salida producida por la función *dlmFilter*), el número de predicciones que se desea (*nAhead*), y el número de muestras que se quiere de las predicciones (*sampleNew*). Los datos de salida producidos por la función son las medias y varianzas de las distribuciones predictivas de las observaciones (*f* y *Q* respectivamente) y de los estados (*a*, *R* respectivamente), la función también acepta información que contenga *NA*'s, sin necesidad de hacer ajustes. A continuación vemos un pequeño fragmento de código para ver como se usa la función:

```
----- Código en R -----
> # Definimos un Pronóstico de la siguiente forma:
> dlm.Pron <- dlmForecast(dlm.Filt, nAhead = num_predicciones,
                          sampleNew = num_muestras)
> # Donde nAhead es el número de predicciones que queremos
> # obtener, y sampleNew el número de muestras que queremos
> # de las predicciones.
> # Podemos además observar que "dlm.Filt" es un objeto de clase
> # dlmFiltered:
> class(dlm.Filt)
[1] "dlmFiltered"
>
> # Como mencionamos, podemos ver la salida que produce el
> # Pronóstico
> ls(dlm.Pron)
[1] "a"          "f"          "newObs"     "newStates" "Q"
[6] "R"
>
> # La función head(), arroja las primeras partes de un vector,
> # matriz, tabla, data frame o función.
>
> # Podemos acceder a estos campos de la siguiente forma:
> head(dlm.Pron$f)
[1] 607.9 637.9 558.3 537.0 607.9 637.9
-----
```


2.4 Procesos de Innovación

Un aspecto importante en cualquier modelo es la verificación del mismo, o mejor dicho cómo podríamos determinar que un modelo es adecuado en nuestro problema, y algo que nos puede ayudar es comprobar si los supuestos del modelo se cumplen. En ese sentido los procesos conocidos como procesos de innovación son una ayuda:

Para los DLM's ya vimos que podemos obtener la predicción del primer paso, $f_t = E(Y_t | y_{1:t-1})$, entonces definimos al *error de predicción* como:

$$e_t = Y_t - E(Y_t | y_{1:t-1}) = Y_t - f_t.$$

Que puede escribirse, de manera alternativa, en términos del error en la estimación del estado como sigue:

$$\begin{aligned} e_t &= Y_t - F_t a_t = F_t + v_t - F_t a_t \\ &= F_t(\theta_t - a_t) + v_t. \end{aligned}$$

además la sucesión, $\{e_t\}_{t \geq 1}$ de errores de predicción cumple importantes propiedades, a continuación presentamos algunas:

Proposición 2.4 *Sea $\{e_t\}_{t \geq 1}$ la sucesión de errores de predicción de un DLM. Entonces las siguientes propiedades se cumplen:*

- (i) *El valor esperado de e_t es cero.*
- (ii) *El vector aleatorio e_t no está correlacionado con cualquier función de Y_1, \dots, Y_{t-1} .*
- (iii) *Para cualquier $s < t$, e_t y Y_s no están correlacionados.*
- (iv) *Para cualquier $s < t$, e_t y e_s no están correlacionados.*
- (v) *e_t es una función lineal de Y_1, \dots, Y_t .*
- (vi) *$\{e_t\}_{t \geq 1}$ es un proceso Gaussiano.*

DEMOSTRACIÓN.

(i) Tomando la esperanza:

$$\begin{aligned} E(e_t) &= E(E(Y_t - f_t | y_{1:t-1})) \\ &= E(E(Y_t | y_{1:t-1}) - E(f_t | y_{1:t-1})). \\ &= E(f_t - f_t) = E(0) = 0 \end{aligned}$$

(ii) Sea $Z = g(Y_1, \dots, Y_{t-1})$ entonces:

$$\begin{aligned} \text{Cov}(e_t, Z) &= E(e_t Z) - E(e_t)E(Z) \\ &= E(E(e_t Z \mid y_{1:t-1})) = E(ZE(e_t \mid y_{1:t-1})) \\ &= E(0) = 0 \end{aligned}$$

(iii) De que $\forall s < t$ e_t y Y_s no están correlacionados se sigue de (ii), en particular $Z = Y_s$ si las observaciones son univariadas, en caso contrario, se aplica componente a componente

(iv) De que $\forall s < t$ e_t y e_s no están correlacionados se sigue de (ii), en particular $Z = e_s$ si las observaciones son univariadas, en caso contrario, se aplica componente a componente

(v) Recordando que (Y_1, \dots, Y_t) tiene distribución normal y $e_t = Y_t - f_t$, donde $f_t = E(Y_t \mid y_{1:t-1})$ se sigue que f_t es función lineal de (Y_1, \dots, Y_{t-1}) , lo que implica que e_t es función lineal (Y_1, \dots, Y_t)

(vi) Por (v), (e_1, \dots, e_t) es una transformación lineal de (Y_1, \dots, Y_t) , que tiene distribución Normal conjunta, de lo que se sigue que (e_1, \dots, e_t) también tiene distribución Normal. Por lo tanto todo vector (e_1, \dots, e_t) finito tiene distribución Normal, así que el proceso $\{e_t\}_{t \geq 1}$ es Gaussiano

■

Los errores de predicción e_t , son llamados también innovaciones. La representación de $Y_t = f_t + e_t$ justifica la terminología, ya que se piensa en Y_t como la suma de un componente f_t , el cual se puede predecir a partir de las observaciones pasadas, y otro componente e_t , el cual es independiente del pasado, por lo que se considera que proporciona la nueva información a Y_t . Es conveniente trabajar con esta forma el DLM, lo cual se logra tomando como nueva variable de estados los vectores $a_t = E(\theta_t \mid y_{1:t-1})$, entonces la ecuación de las observaciones se deriva de $e_t = Y_t - f_t = Y_t - F_t a_t$:

$$Y_t = F_t a_t + e_t, \tag{2.4a}$$

tomando $a_t = G_t m_{t-1}$, donde m_{t-1} está dado por el filtro de Kalman:

$$a_t = G_t m_{t-1} = G_t a_{t-1} + G_t R_{t-1} F'_{t-1} Q_{t-1}^{-1} e_{t-1}$$

así, la ecuación de estados se convierte en:

$$a_t = G_t a_{t-1} + w_t^*. \quad (2.4b)$$

con $w_t^* = G_t R_{t-1} \tilde{F}'_{t-1} Q_{t-1}^{-1} e_{t-1}$. El sistema 2.4 es la forma “innovada” del DLM. Notemos que así los errores de las observaciones y los errores de los estados ya no son independientes, es decir, la dinámica de los estados ya no es independiente de las observaciones. La ventaja principal de esto, es que en la forma inovada todos los componentes del vector de estados de los que no podemos obtener información de las observaciones son removidos, y en este sentido se hablaría de un modelo mínimo.

Cuando las observaciones son univariadas, la sucesión estandarizada de inovaciones, definida por $\tilde{e} = \frac{e_t}{\sqrt{Q_t}}$, es un ruido blanco. Esta propiedad puede ser usada para verificar los supuestos del modelo, es decir, si el modelo es correcto, la sucesión $\tilde{e}_1, \dots, \tilde{e}_t$ obtenida de los datos debería parecer una muestra de tamaño t de una distribución normal estándar.

Pasando al caso en que las observaciones sean multivariadas, se procede de una manera similar, en ese caso consideramos al vector estandarizado $\tilde{e} = B_t e_t$, donde B_t es una matriz de $p \times p$, tal que $B_t Q_t B_t' = I$. Esto hace a las componentes de \tilde{e}_t independientes e idénticamente distribuidas de acuerdo a una distribución normal estándar, y buscamos así un análisis similar al del caso univariado.

Procesos de Innovación en R

En R las innovaciones estandarizadas pueden ser extraídas de los objetos de clase *dlmFiltered* usando la función *residuals*. A continuación vemos un pequeño fragmento de código para ver como se usa la función:

```
----- Código en R -----
> # La función head(), arroja las primeras partes de un vector,
> # matriz, tabla, data frame o función.
>
> # Habiendo definido un objeto de clase dlmFiltered (dlm.Filt),
> # podemos extraer las innovaciones de la siguiente forma:
> head(residuals(dlm.Filt, sd = FALSE))
2012-01-03 2012-01-04 2012-01-05 2012-01-06 2012-01-09 2012-01-10
-4.784e-06 1.089e-05 2.984e-01 1.061e-01 -5.595e-01 -6.907e-02
-----
```


Parte III

Análisis de Series de Tiempo Univariadas mediante DLM

Capítulo 3

Representación alternativa de los ARIMA

A continuación describiremos una idea muy útil que se puede aplicar con los modelos dinámicos lineales en la práctica, que consiste en modelar cualquier serie de tiempo como el conjunto de varias componentes, y cada una puede representar un modelo dinámico lineal.

Uno de los principales problemas cuando se intenta modelar datos en la realidad es, cómo se puede representar el modelo. En series de tiempo una idea muy intuitiva, es el pensar a la serie que se intenta modelar, como el resultado de combinar distintos componentes que caracterizan a esta serie, como lo pueden ser la tendencia, la estacionalidad u otros componentes. Los modelos dinámicos lineales, permiten utilizar esta idea, y así considerar cada componente de las que nos referimos, como un modelo dinámico lineal. Pero el punto es cómo se logra, o cómo se pueden combinar estas componentes, y qué darían como resultado. Para esto se emplea lo que se conoce como *técnica de descomposición aditiva*, que, como su nombre lo indica, es de forma aditiva a los componentes. Además lo que se obtiene como resultado es un único modelo dinámico lineal, que sería el que modelaría a la serie en cuestión.

Consideremos una serie de tiempo univariada $\{Y_t\}$. Ahora suponemos que la serie puede escribirse como la suma de componentes independientes:

$$Y_t = Y_{1,t} + \dots + Y_{t,h} \quad (3.1)$$

donde $Y_{i,t}$ representa el i -ésimo componente, $i = 1, \dots, h$, y puede ser representado por un DLM de la siguiente forma:

$$\begin{aligned} Y_{i,t} &= F_{i,t}\theta_{i,t} + v_{i,t}, & v_{i,t} &\sim \mathcal{N}_m(0, V_{i,t}) \\ \theta_{i,t} &= G_{i,t}\theta_{i,t-1} + w_{i,t}, & w_{i,t} &\sim \mathcal{N}_P(0, W_{i,t}) \end{aligned}$$

donde los vectores p_i -dimensionales vectores de estados $\theta_{i,t}$ son distintos y las series $(Y_{i,t}, \theta_{i,t})$ y $(Y_{j,t}, \theta_{j,t})$ son mutuamente independientes para toda $i \neq j$. Entonces los DLM's son combinados para obtener el DLM de $\{Y_t\}$. Debido a la independencia de los componentes se tiene que $Y_t = \sum_{i=1}^h Y_{i,t}$, y es descrito por el DLM:

$$\begin{aligned} Y_t &= F_t\theta_t + v_t, & v_t &\sim \mathcal{N}_m(0, V_t) \\ \theta_t &= G_t\theta_{t-1} + w_t, & w_t &\sim \mathcal{N}_P(0, W_t) \end{aligned}$$

donde

$$\theta_t = \begin{bmatrix} \theta_{1,t} \\ \vdots \\ \theta_{h,t} \end{bmatrix}, \quad F = [F_{1,t} \mid \cdots \mid F_{h,t}]$$

G_t y W_t son los bloques de matrices diagonales

$$G_t = \begin{bmatrix} G_{1,t} & & \\ & \ddots & \\ & & G_{h,t} \end{bmatrix}, \quad W_t = \begin{bmatrix} W_{1,t} & & \\ & \ddots & \\ & & W_{h,t} \end{bmatrix},$$

Esta técnica le da al modelo una mayor flexibilidad a comparación de los modelos ARIMA, permitiendo la modelación de series no estacionarias directamente, sin necesidad de una transformación previa.

Antes de pasar a la siguiente sección, mencionaremos una de las peculiaridades de los modelos dinámicos lineales, y es que éstos, se podrían considerar como el caso general de los modelos ARIMA.

3.1 Representación de un ARIMA mediante un DLM

La flexibilidad de los DLM hace posible que cualquier modelo ARIMA pueda ser representado mediante un modelo dinámico lineal, de hecho, se puede demostrar que para cualquier ARIMA es posible encontrar un DLM cuyo proceso $\{Y_t\}$ tiene la misma distribución como la dada por el ARIMA, además de que este espacio de estados no queda determinado de manera única. A continuación presentaremos una forma de hacer esta representación, para ello, consideremos un modelo $ARMA(p, q)$, suponiendo que su nivel (μ) es cero, entonces se puede establecer la siguiente relación:

$$Y_t = \sum_{j=1}^r \phi_j Y_{t-j} + \sum_{j=1}^{r-1} \psi_j \epsilon_{t-j} + \epsilon_t,$$

con $r = \max\{p, q + 1\}$, $\phi_j = 0$ para $j > p$ y $\psi_j = 0$ para $j > q$. Además definimos las siguientes matrices:

$$F = [1 \quad 0 \quad \cdots \quad 0] \tag{3.2}$$

$$G = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ \phi_{r-1} & 0 & \cdots & 0 & 1 \\ \phi_r & 0 & \cdots & 0 & 0 \end{bmatrix}$$

$$R = [1 \quad \psi_1 \quad \cdots \quad \psi_{r-2} \quad \psi_{r-1}]'$$

y se introduce un vector de estados r-dimensional $\theta_t = (\theta_{1,t}, \dots, \theta_{r,t})'$, entonces para el modelo ARMA dado, se tiene la siguiente representación en forma de un DLM:

$$\left\{ \begin{array}{l} Y_t = F\theta_t, \\ \theta_t = G\theta_{t-1} + R\epsilon_t. \end{array} \right\} \tag{3.3}$$

Observe que éste, es un DLM con $V = 0$ y $W = RR'\sigma^2$ donde σ^2 es la varianza de la sucesión de errores $\{\epsilon_t\}$. A continuación se mostrará la equivalencia.

Observe que, de la ecuación de observaciones, se tiene que $y_t = \theta_{1,t}$, y la ecuación de estados queda de la siguiente manera:

$$\begin{aligned}
\theta_{1,t} &= \phi_1\theta_{1,t-1} + \theta_{2,t-1} + \epsilon_t \\
\theta_{2,t} &= \phi_2\theta_{1,t-1} + \theta_{3,t-1} + \psi_1\epsilon_t \\
&\vdots \\
\theta_{r-1,t} &= \phi_{r-1}\theta_{1,t-1} + \theta_{r,t-1} + \psi_{r-2}\epsilon_t \\
\theta_{r,t} &= \phi_r\theta_{1,t-1} + \psi_{r-1}\epsilon_t
\end{aligned} \tag{3.4}$$

sustituyendo la expresión de $\theta_{2,t}$, obtenida de la segunda ecuación en la primera ecuación, obtenemos lo siguiente:

$$\theta_{1,t} = \phi_1\theta_{1,t-1} + \phi_2\theta_{1,t-2} + \theta_{3,t-2} + \psi_1\epsilon_{t-1} + \epsilon_t$$

y procediendo de manera recursiva, tenemos que:

$$\theta_{1,t} = \phi_1\theta_{1,t-1} + \dots + \phi_r\theta_{1,t-r} + \psi_1\epsilon_{t-1} + \dots + \psi_{r-1}\epsilon_{t-r+1} + \epsilon_t$$

donde $r = \max\{p, q + 1\}$ y $y_t = \theta_{1,t}$, con lo que se llega a que es un ARMA. La representación de un modelo $ARIMA(p, d, q)$ mediante un DLM, con $d \in \mathcal{Z} > 0$, puede derivarse como una extensión del caso estacionario. Para ello consideremos a $Y_t^* = \Delta^d Y_t$, y además Y_t^* se puede modelar con un $ARMA(p, q)$, con lo que, basados en lo mencionado anteriormente, se puede representar a este proceso mediante un DLM. Pero si la idea es modelar el proceso original $\{Y_t\}$ se tiene que recuperar de Y_t^* , se puede demostrar que para $d > 0$, considerando $Y_t^* = \Delta^d Y_t$, y que la siguiente igualdad se cumple:

$$\Delta^{d-j} Y_t = Y_t^* + \sum_{i=1}^j \Delta^{d-i} Y_{t-1}, \quad j = 1, \dots, d. \tag{3.5}$$

Definimos ahora al vector de estados de la siguiente forma:

$$\theta_t = \begin{bmatrix} Y_{t-1} \\ \Delta Y_{t-1} \\ \vdots \\ \Delta^{d-1} Y_{t-1} \\ Y_t^* \\ \phi_2 Y_{t-1}^* + \cdots + \phi_r Y_{t-r+1}^* + \psi_1 \epsilon_t = \cdots + \psi_{r-1} \epsilon_{t-r+2} \\ \phi_3 Y_{t-1}^* + \cdots + \phi_r Y_{t-r+2}^* + \psi_2 \epsilon_t = \cdots + \psi_{r-1} \epsilon_{t-r+3} \\ \vdots \\ \phi_r Y_{t-1}^* + \psi_r - 1 \epsilon_t \end{bmatrix} \quad (3.6)$$

$$F = [1 \quad 1 \quad \cdots \quad 1 \quad 0 \quad \cdots \quad 0],$$

$$G = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & 1 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \phi_1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \phi_2 & 0 & 1 & \cdots & 0 \\ & & \cdots & \vdots & \vdots & & \ddots & \\ \cdots & \cdots & \cdots & \phi_{r-1} & 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & \phi_r & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (3.7)$$

$$R = [0 \quad \cdots \quad 0 \quad 1 \quad \psi_1 \quad \cdots \quad \psi_{r-2} \quad \psi_{r-1}]',$$

$$W = RR' \sigma^2$$

Con lo anterior podemos definir un modelo *ARIMA* para $\{Y_t\}$ por medio de la representación de un DLM, de la siguiente manera:

$$\begin{cases} Y_t = F\theta_t, \\ \theta_t = G\theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, W) \end{cases} \quad (3.8)$$

En la sección anterior se mostró que debido a la estructura de los DLM es posible realizar la modelación directa de series de tiempo no estacionarias, ya que se puede considerar el uso de componentes para la tendencia, o la estacionariedad, por ejemplo. En esta sección mostramos que es posible considerar a los modelos *ARIMA* como un caso particular de los DLM, pero éstos, entre otras

cosas, no poseen la flexibilidad de modelar las series de tiempo no estacionarias de manera directa, por lo que es poco usual la representación de un modelo ARIMA mediante un DLM.

A continuación surge la pregunta ¿cómo podemos usar estas ideas en algún problema en particular?. En este sentido los modelos dinámicos lineales tienen herramientas muy útiles para determinados casos, ¿pero por qué para determinados casos?, esto se debe a la idea de poder representar a una serie de tiempo a través de distintas componentes, y cada una representarla mediante un modelo dinámico lineal. En el caso de series de tiempo algunas de las componentes más frecuentes son la tendencia y la estacionalidad. Explicaremos en las siguientes secciones los modelos que comúnmente se utilizan para estos casos. Además, en la última sección mostraremos el modelo para el caso en que se desea modelar una serie de tiempo con factores que tengan alguna correlación mediante la generalización de los modelos de regresión lineal simple.

3.2 Modelos para la Tendencia

Dentro de los DLMS, los conocidos como Polinomiales, son los que comúnmente se utilizan para describir la tendencia de una serie de tiempo, si consideramos que la serie se encuentra en el tiempo t , la tendencia esperada de la serie puede ser vista como el comportamiento esperado de Y_{t+k} para $k \geq 1$, dada la información al tiempo t , es decir, $f_t(k) = E(Y_{t+k} | y_{1:t})$, la función predictiva. Se considera que un modelo polinomial de orden n es un DLM con matrices constantes $F_t = F$, $G_t = G$, y función predictiva:

$$f_t(k) = E(Y_{t+k} | y_{1:t}) = a_{t,0} + a_{t,1}k + \dots + a_{t,n-1}k^{n-1}, \quad k \geq 0, \quad (3.9)$$

donde $a_{t,0}, \dots, a_{t,n-1}$ son funciones lineales de $m_t = E(\theta_t | y_{1:t})$ y son independientes de k , por lo que la función predictiva es un polinomio de orden $n-1$ en k . Aunque se puede pensar en usar una n lo suficientemente grande para aproximar el polinomio a la función predictiva, generalmente en la práctica se utilizan valores pequeños, en particular $n = 1$ (Modelo Uniforme Local) y $n = 2$ (Modelo de Crecimiento Lineal).

Modelo Uniforme Local

La caminata aleatoria más un ruido o Modelo Uniforme Local, está definido por

$$\begin{aligned} Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W), \end{aligned} \quad (3.10)$$

donde la sucesión de errores $\{v_t\}$ y $\{w_t\}$ son independientes, dentro de la secuencia y entre secuencias. Vale la pena también mencionar que bajo este modelo, el comportamiento de $\{Y_t\}$ está influenciado por el cociente de las dos varianzas, $r = W/V$ que se le conoce como el cociente *signal-to-noise*. Con el filtro de Kalman se puede obtener:

$$\begin{aligned} \mu_t | y_{1:t-1} &\sim \mathcal{N}(m_{t-1}, R_t = C_{t-1} + W), \\ Y_t | y_{1:t-1} &\sim \mathcal{N}(f_t = m_{t-1}, Q_t = R_t + V), \\ \mu_t | y_{1:t} &\sim \mathcal{N}(m_t = m_{t-1} + K_t e_t, C_t = K_t V), \end{aligned}$$

donde $K_t = R_t/Q_t$ y $e_t = Y_t - f_t$, además la distribución predictiva a k pasos de este modelo es:

$$Y_{t+k} | y_{1:t} \sim \mathcal{N}(m_t, Q_t(k)), \quad k \geq 1, \quad (3.11)$$

donde $Q_t(k) = C_t + \sum_{j=1}^k W_{t+j} + V_{t+k} = C_t + kW + V$, por lo que se puede ver que la función predictiva $f_t(k) = E(Y_{t,k} | y_{1:t}) = m_t$ es constante como función de k . La incertidumbre en las observaciones futuras es resumida por la varianza $Q_t(k) = C_t + kW + V$ que aumenta conforme el horizonte de tiempo $t+k$ lo hace.

Modelo de Crecimiento Lineal

El modelo local de tendencia lineal o modelo de crecimiento local esta definido por las siguientes ecuaciones:

$$\begin{aligned} Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{t,1}, & w_t &\sim \mathcal{N}(0, \sigma_\mu^2), \\ \beta_t &= \beta_{t-1} + w_{t,2}, & w_t &\sim \mathcal{N}(0, \sigma_\beta^2), \end{aligned} \quad (3.12)$$

con los errores $v_t, w_{t,1}$ y $w_{t,2}$ no correlacionados y el vector de estados $\theta_t =$

$(\mu_t, \beta_t)'$, donde μ_t es interpretado como el nivel y β_t como la tasa de crecimiento. Si consideramos $m_{t-1} = (\hat{\mu}_{t-1}, \hat{\beta}_{t-1})'$, se tiene que:

$$a_t = Gm_{t-1} = \begin{bmatrix} \hat{\mu}_{t-1} + \hat{\beta}_{t-1} \\ \hat{\beta}_{t-1} \end{bmatrix} \quad (3.13a)$$

$$f_t = F_t a_t = \hat{\mu}_{t-1} + \hat{\beta}_{t-1}, \quad (3.13b)$$

$$m_t = \begin{bmatrix} \hat{\mu}_t \\ \hat{\beta}_t \end{bmatrix} = a_t + K_t e_t = \begin{bmatrix} \hat{\mu}_{t-1} + \hat{\beta}_{t-1} + k_{t,1} e_t \\ \hat{\beta}_{t-1} + k_{t,2} e_t \end{bmatrix} \quad (3.13c)$$

Y la función de predicción es:

$$f_t = (k) = \hat{\mu}_t + k\hat{\beta}_t \quad (3.14)$$

que es una función lineal respecto a k , por lo que el modelo de crecimiento lineal es DLM polinomial de orden 2.

Modelo Polinomial de orden n

Tanto el modelo uniforme local como el modelo de crecimiento lineal son un caso particular de los modelos polinomiales de orden n . El modelo general polinomial de orden n tiene un espacio de estados de dimensión n y está descrito por las matrices:

$$F = (1, 0, \dots, 0) \quad (3.15a)$$

$$G = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 1 \\ 0 & \dots & & 0 & 1 \end{bmatrix} \quad (3.15b)$$

$$W = \text{diag}(W_1, \dots, W_n) \quad (3.15c)$$

El modelo también puede ser escrito de la siguiente forma:

$$\begin{cases} Y_t = \theta_{t,1} + v_t \\ \theta_{t,j} = \theta_{t-1,j} + \theta_{t-1,j+1} + w_{t,j} & j = 1, \dots, n-1 \\ \theta_{t,n} = \theta_{t-1,n} + w_{t,n} \end{cases} \quad (3.16)$$

Pero lo usual es usar valores de n pequeños.

3.3 Modelos para la Estacionalidad

A continuación se mostrarán dos maneras de modelar series de tiempo que presentan algún comportamiento cíclico:

Como ejemplo, supongamos que tenemos una serie $\{Y_t\}$ trimestral y asumimos que su media es cero. Más específicamente, sólo consideremos que la serie es estacional, por lo que podemos ver a la serie por medio de desviaciones de la media ($\alpha_i, i = 1, \dots, 4$) las cuales serán distintas en cada ciclo (trimestre). Ahora consideremos que Y_{t-1} es la serie en el primer trimestre, por lo que

$$Y_{t-1} = \alpha_i + v_{t-1} \quad (3.17)$$

$$Y_t = \alpha_t + v_t \quad (3.18)$$

si consideramos a $\theta_{t-1} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$ y $F_t = F = (1, 0, 0, 0)$ lo anterior se convierte en un DLM, con ecuación de observaciones

$$Y_{t-1} = F\theta_{t-1} + v_{t-1}$$

Pero hay que observar que para que quede bien definida la ecuación de estados en el tiempo t , se debe considerar a α_2 por lo que se requiere que $\theta_t = (\alpha_2, \alpha_1, \alpha_4, \alpha_3)'$, y una forma de lograr esto es considerando una matriz de permutación,

$$G = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

La ecuación de estados queda de la forma

$$\theta_t = G\theta_{t-1} + w_t = (\alpha_2, \alpha_1, \alpha_4, \alpha_3)' + w_t$$

En general, una serie de tiempo estacional con periodo s puede ser modelada a través del s - vector de estados θ_t dimensional mediante un DLM, con $F = (1, 0, \dots, 0)$ y G una matriz de permutación. También es común imponer la condición de que la $\sum_{j=1}^s \alpha_j = 0$, lo cual implica que hay solamente $s - 1$ factores estacionales libres, lo que sugiere un representación más parsimoniosa con un vector de estados $(s - 1)$ -dimensional, y en este caso se puede considerar la matriz de permutación

$$G = \begin{bmatrix} -1 & -1 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & & 1 & 0 \end{bmatrix}$$

Cabe mencionar que en el modelo estacional estático, w_t se convierte en un vector de ceros, pero se puede considerar que los efectos cíclicos pueden cambiar a lo largo del tiempo, y en tal caso W_t no será cero. Una variación en los componentes se puede introducir, por ejemplo, considerando $W = \text{diag}(\sigma_w^2, 0, \dots, 0)$.

3.4 Modelos de Regresión

Como mencionamos anteriormente, los DLM pueden ser considerados como una generalización de los modelos de regresión, por lo que la implementación de estos modelos de regresión, a un DLM parece adecuado. Además de tener el extra de poder incluir variables explicativas en la serie, algo que no se podía hacer en los modelos ARIMA.

Recordando, el modelo de regresión lineal simple, tiene la siguiente forma:

$$Y_t = \beta_1 + \beta_2 x_t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_t^2)$$

Debido a que estamos interesados en su aplicación a series de tiempo, el suponer que los errores $\{\epsilon_t\}$ son *i.i.d.*, no es algo razonable. Una solución es considerar la evolución en el tiempo de la relación entre x y y . A estos modelos se les conoce como, *modelos de regresión dinámicos*, y tienen la, forma:

$$Y_t = \beta_{1,t} + \beta_{2,t}x_t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_t^2),$$

Lo que nos define a un DLM, con $F_t = [1 \ x_t]$, $\theta_t = [\beta_{1,t}\beta_{2,t}]'$, $V = \sigma^2$, y una ecuación de estado para θ_t , es decir:

$$\begin{aligned} Y_t &= x_t' \theta_t + v_t, & v_t &\sim \mathcal{N}(0, \sigma_t^2) \\ \theta_t &= G_t \theta_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W_t^2) \end{aligned}$$

donde $x_t' = [x_{1,t}, \dots, x_{p,t}]$ son los valores de la p -ésima variable explicativa al tiempo t . Una elección común es considerar a G_t como la matriz identidad y a W como una matriz diagonal, en otras palabras, modelar los coeficientes de regresión como una caminata aleatoria.

Modelos Dinámicos de Regresión Lineales en R

En R los modelos dinámicos de regresión pueden ser creados mediante la función estándar *dlmModReg*, que básicamente nos permite definir un objeto *dlm* de una manera más directa, y cuyos argumentos son la información, y , en forma de un vector numérico, matriz o serie de tiempo, la varianza del error de la observaciones, dV , los elementos de la diagonal de la matriz de varianzas de los estados, dW , la media a priori $m0$ y la varianza a priori $C0$. Una vez definido el objeto *dlm*, podemos ya sea usar la filtración o suavizamiento como vimos anteriormente. A continuación vemos un pequeño fragmento de código para ver como se usa la función:

```
----- Código en R -----
> # Definimos un Modelo de Regresión Dinámico Lineal de la
> # siguiente forma:
> dlm.ModRegDinLin <- dlmModReg(datos, dV = var.V, dW = var.W)
>
> # Esto nos define un objeto "dlm"
> class(dlm.ModRegDinLin)
[1] "dlm"
```



```
> ls(dlm.ModRegDinLin)
[1] "CO" "FF" "GG" "JFF" "JGG" "JV" "JW" "m0" "V" "W" "X")
>
> # Y podemos usarlo ya sea para filtración o suavizamiento
> dlm.Filt <- dlmFilter(datos, dlm.ModRegDinLin)
> dlm.Smooth <- dlmFilter(datos, dlm.ModRegDinLin)
>
> # La función head(), arroja las primeras partes de un vector,
> # matriz, tabla, data frame o función.
>
> head(dlm.ModRegDinLin$m)
      [,1]      [,2]
2012-01-03 0.00000 0.0000000
2012-01-04 -0.01513 0.0000778
2012-01-05 -0.01531 -0.0343681
2012-01-06 -0.01537 -0.1811737
2012-01-09 -0.01609 -0.2607268
2012-01-09 -0.01706 -0.1994370
```

Parte IV

Implementación de los
DLM

Capítulo 4

DLM en la modelación de rendimientos

El precio de las acciones en el mercado cambia a través del tiempo; surgen nuevas acciones, las existentes pueden desaparecer o combinarse con otras, una gran cantidad de situaciones pueden suceder. Por ello, como implementación de los DLM, usaremos el modelo conocido como el CAPM (por sus siglas en inglés, Capital Asset Pricing Model), el cual intenta estimar el rendimiento de una acción o portafolio, en su sentido más general. Para ello se presentarán los conceptos básicos para su implementación. Cabe mencionar que el CAPM no es el único modelo para la predicción de los rendimientos de una acción, pero es considerado un buen modelo. El CAPM, tiene un rol importante en la selección de portafolios conforme a las preferencias media/varianza, es decir, se busca tener grandes rendimientos con el mínimo riesgos.

4.1 Conceptos básicos

En la medición del desempeño de un portafolio o una colección de inversiones en acciones, lo que nos interesa es el aumento o disminución del valor de estas acciones en un periodo de tiempo. En las distintas teorías económicas existen muchos supuestos en lo referente a los rendimientos de las acciones, al igual de que se consideran distintos tipos de mercado, entre otras cosas. Un supuesto común en muchas de ellas, es el considerar que los individuos de una economía actúan de manera “racional”, es decir, que esperan ganancias racionales y con un cierto grado de aversión al riesgo. Bajo estas circunstancias, los precios y rendimientos pueden ser determinados con modelos que consideren esta idea. Para ello consideremos una serie de tiempo de los precios de una acción, P_1, \dots, P_n , por simplicidad supondremos que no paga dividendos y no consideraremos la inflación, nosotros estamos interesados en calcular el

rendimiento entre dos periodos. Es decir, los cambios en el precio de la acción expresados como una fracción del precio inicial. Con esta idea se define:

Definición 4.1 (Rendimiento Simple) *El rendimiento simple, sobre el periodo de tiempo de $t - 1$ a t , r_t , se define como:*

$$r_t = (P_t - P_{t-1})/P_{t-1}$$

Ahora, con esta sencilla definición podemos comprender la idea del CAPM. El CAPM se basa en dos ideas, la primera es la existencia de un portafolio del mercado M , y la segunda en la noción de un parámetro que se le conoce como *beta*, que vincula a cualquier acción con el mercado. En la teoría, el portafolio del mercado incluye todas las acciones, pero en la práctica, se considera algún índice del sector o área de interés, como el NYSE, el IPC en México, el NASDAQ, o el TOPIX.

Consideremos cualquier acción P con exceso de rendimiento r_P , y al portafolio del mercado M con exceso de rendimiento r_M , donde el exceso de rendimiento es el rendimiento total menos el rendimiento de una acción libre de riesgo, usualmente se considera el rendimiento de algún bono. Entonces se define a la beta de la acción P como:

$$\beta_P = \frac{Cov(r_P, r_M)}{Var(r_M)}$$

Donde observamos que beta es proporcional a la covarianza entre el rendimiento de la acción y el rendimiento del mercado. El nombre de beta es por que el modelo puede ser visto como un modelo de regresión simple del exceso de rendimiento de la acción, $r_P(t)$, en los periodos $t = 1, \dots, T$, es decir:

$$r_P(t) = \alpha_P + \beta_P r_M(t) + v_P(t)$$

La beta en el modelo se usa para separar el riesgo y el rendimiento en dos partes, es decir, si conocemos la beta de alguna acción, podemos descomponer el exceso de rendimiento de esa acción en una componente del mercado, donde consideraremos el riesgo de mercado o sistemático, y una componente residual, donde consideraremos al riesgo residual o específico:

$$r_P = \beta_P r_M + \theta_P$$

de donde obtenemos que:

$$\sigma_P^2 = \beta_P^2 \sigma_M^2 + V_P^2$$

donde V_P^2 es la varianza de v_P .

Además de que en el modelo la componente residual θ_P no está correlacionada con el mercado. La interpretación de la beta se puede hacer similar a la de un modelo de regresión, en este caso, por ejemplo, si el mercado sube, la componente de las acciones relacionada al mercado responden de acuerdo a su beta, si la beta es positiva, cuando el mercado sube, el rendimiento de la acción sube, análogamente, cuando el mercado baja, entonces el rendimiento de la acción baja. Y la magnitud de respuesta de la acción, está determinada por la magnitud de la beta, si beta es uno, la acción tiene el mismo riesgo como el mercado, entonces si el mercado sube 10 %, el rendimiento de la acción sube 10 % en promedio, análogamente, si el mercado baja 10 %, el rendimiento de la acción baja 10 % en promedio. Si beta es mayor que uno, si el mercado sube 10 %, el rendimiento de la acción sube más del 10 % en promedio, es decir, incrementa la volatilidad, por lo que tiene una respuesta mayor a lo que sucede en el mercado. Todas las acciones están relacionadas al mercado, pero cómo cada acción responde al mercado, es diferente. El término α_P que es el promedio de los rendimientos residuales, también tiene su interpretación, además el CAPM plantea algo especial acerca de los rendimientos que no forman parte del mercado, establece que el rendimiento esperado de la componente residual de todas las acciones y cualquier portafolio es igual a cero, es decir, $E(\theta_P) = 0$, lo que implica que el rendimiento esperado de la acción, $E(r_P) = \mu_P$, está determinado enteramente por el rendimiento esperado del mercado, $E(r_M) = \mu_M$, y la beta de la acción, β_P :

$$E(r_P) = \beta_P E(r_M) = \beta_P \mu_M$$

Implícitamente, aquí está uno de los supuestos del modelo, que todos los inversionistas tiene el mismo rendimiento esperado, el cual difiere únicamente por su tolerancia al riesgo. El término α_P también está relacionado con la proporción de información (*IR*, por sus siglas en inglés), usada en la evaluación de portafolios, ya que la idea es que capture la noción de éxito de un portafolio, es decir, qué tan buena fue la selección en el portafolio con respecto a otros inversionistas. Pero volviendo al modelo, la idea es que los inversionistas son

compensados por tomar riesgos necesarios, pero no por tomar riesgos innecesarios. El riesgo en el portafolio del mercado es necesario: No se puede evitar el riesgo del mercado. Por otra parte, el riesgo de la componente residual, o el riesgo específico, es auto - impuesto: todos los inversionistas pueden evitar el riesgo residual. Es importante resaltar que el CAPM es una teoría, y como tal está basada en supuestos que no siempre son tan precisos, por tal motivo existe cierta controversia sobre el modelo, pero no recae en su lógica, si no en las conclusiones a las que se pueden llegar.

4.1.1. Línea del Mercado de Capitales y Línea del Mercado de Títulos

Como complemento del CAPM, una definición importante es lo que se conoce como la línea del mercado de capitales (CML, por sus siglas en inglés), que relaciona el exceso de rendimiento esperado de un portafolio eficiente y su riesgo, la CML, es:

$$\mu_P = \mu_f + \frac{\mu_M - \mu_f}{\sigma_M} \sigma_P$$

Donde P es un portafolio eficiente, M el portafolio del mercado, $\mu_P = E[P]$, el rendimiento esperado del portafolio, μ_f , el rendimiento de la acción libre de riesgo, observe que es constante, $\mu_M = E[M]$, el rendimiento del portafolio del mercado, y σ_M y σ_P , la desviaciones estándar de M y P. La pendiente de la CML, $\frac{\mu_M - \mu_f}{\sigma_M}$, es conocida como el *Índice de Sharpe*, que lo podemos interpretar como cuánto debe incrementar el rendimiento esperado de nuestro portafolio, si el riesgo del portafolio aumenta en una unidad. Veamos un ejemplo, consideremos que $\mu_f = 0.06$, $\mu_M = 0.15$ y $\sigma_M = 0.22$, entonces la CML, está representada Figura 4.1. Pero la CML se deriva para portafolios eficientes, ¿qué sucede en el caso que queramos considerar alguna acción en particular?. Para ello sirve lo que se conoce como la línea del mercado de títulos (SML, por sus siglas en inglés), que análogamente relaciona el exceso de rendimiento y su riesgo (beta). Considerando cualquier acción j , usando el CAPM observamos que:

$$\mu_j - \mu_f = \beta_j(\mu_M - \mu_f)$$

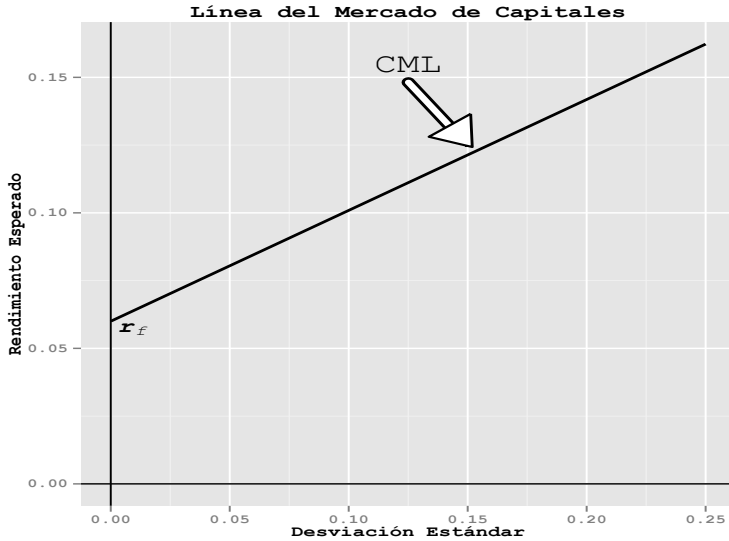


Figura 4.1: Ejemplo de la línea del mercado de capitales.

A esta ecuación, se le conoce como la línea del mercado de títulos, donde μ_j , se considera como una función de β_j . Lo que trata de hacer la SML, es dar la interpretación que mencionamos anteriormente de la beta de forma gráfica, tratando de medir la “*agresividad*” de la acción j , en el sentido de que la SML considera a beta como una medida del riesgo y muestra el beneficio por tomar este riesgo. Por definición, la beta del portafolio de mercado es 1, por lo que una regla usual es considerar:

$$\begin{aligned} \beta_j < 1 & \text{ “agresiva”} \\ \beta_j = 1 & \text{ “riesgo promedio”} \\ \beta_j > 1 & \text{ “no agresiva”} \end{aligned}$$

Cabe mencionar que cuando se usa un modelo de regresión, se le suele llamar *línea característica*, en lugar de línea del mercado de Títulos, ya que la SML no muestra explícitamente que los rendimientos esperados dependen del tiempo. En la Figura 4.2, se muestra un ejemplo de una SML

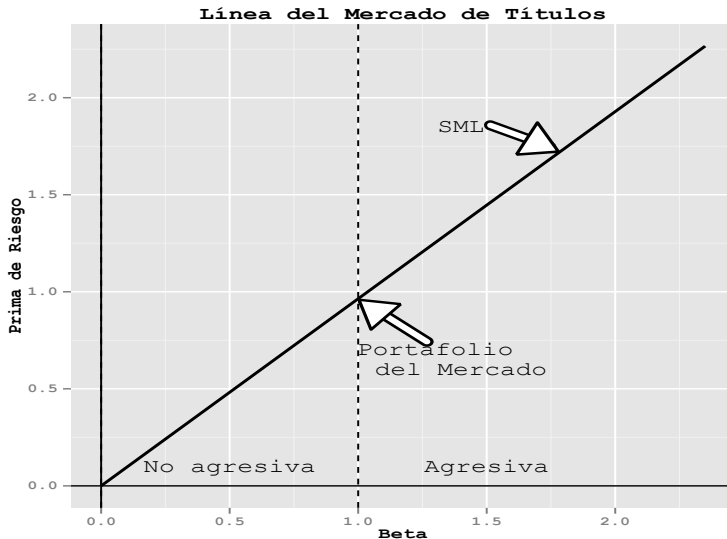


Figura 4.2: Ejemplo de la línea del mercado de Títulos.

4.2 Implementación

4.2.1. Modelo de Regresión Simple

Los datos que utilizaremos abarcan el periodo de: Enero del 2012 a Diciembre del 2012, y son los siguientes:

- Precios de la acción: Apple.
- Precios del portafolio de mercado: *SP&500*.
- Rendimientos de la acción libre de riesgo: Bonos del gobierno (T-Bill).

A continuación podemos observar la gráfica de los precios normalizados, Figura 4.3. Observemos como Apple, si bien no tiene un comportamiento muy similar al del mercado, parece reaccionar hasta cierto punto a los cambios del mercado, es decir, cuando los precios subieron los de Apple también, por tal motivo podríamos esperar una beta positiva, pero también hay que notar que esto no es tan claro. También se podrían observar los rendimientos para tener una idea del posible valor de alfa, pero mejor implementemos el modelo.

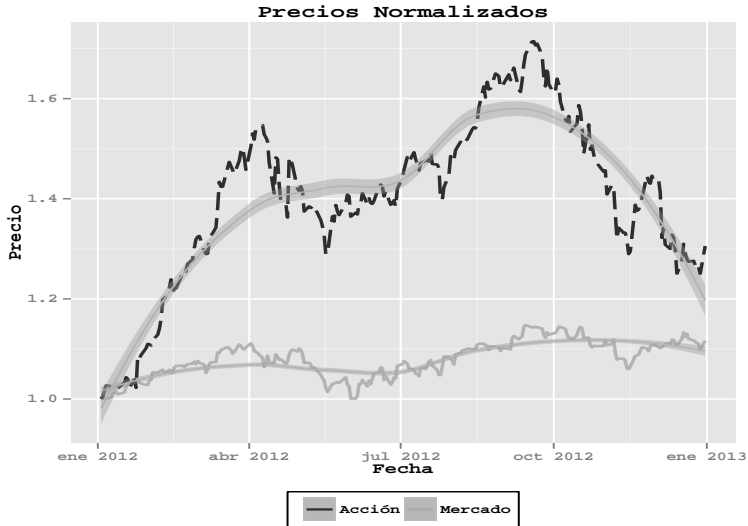


Figura 4.3: Precios Normalizados.

Recordando que el modelo de regresión es: $r_P(t) = \alpha_P + \beta_P r_M(t) + v_P(t)$, usaremos la función `lm`, obteniendo el siguiente resultado:

```
----- Código en R -----
Call:
lm(formula = apple.exc.retornos.diarios ~ sp500.exc.retornos.diarios)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.00393   0.00232    1.69   0.092 .
market.exc.retornos.diarios 1.18709   0.12121    9.79 <2e-16 ***
---
Residual standard error: 0.0157 on 248 degrees of freedom
Multiple R-squared:  0.279,    Adjusted R-squared:  0.276
F-statistic: 95.9 on 1 and 248 DF,  p-value: <2e-16
-----
```

Donde observamos que para Apple la $\hat{\beta} = 1.187$ y $\hat{\alpha} = 0.004$. El estimador de V_P es 0.0157.

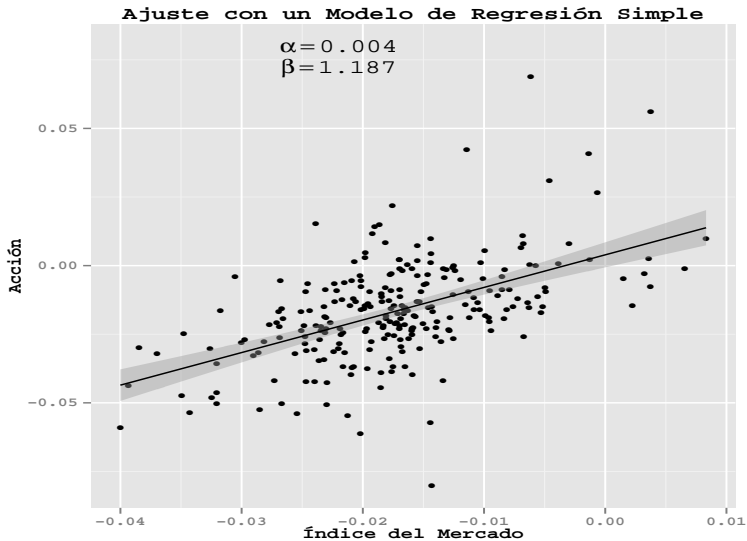


Figura 4.4: Modelo de regresión simple de los rendimientos de Apple.

Note que el valor R^2 para la regresión es del 27.9%, es decir 27.9% del riesgo es debido al riesgo sistemático ($\beta_P^2 \sigma_M^2$), y el restante 72.1% es debido al riesgo específico (V_P^2), en la Figura 4.4 podemos observar la representación gráfica del modelo.

Ahora, se debe tener en cuenta un aspecto que no se ha mencionado, y es que cuando estamos ajustando el modelo, tenemos que considerar el periodo de tiempo que vamos a utilizar, si bien es cierto que entre más largo es el periodo, se dispone de mayor información, también es cierto que, en el modelo de regresión simple, uno de los supuestos que estamos considerando es que β es constante, por lo tanto si usamos un periodo muy amplio, este supuesto no parecería real. Un “método” utilizado para verificar este supuesto, consiste en tomar lo que se le llaman ventanas de tiempo, que consiste en considerar periodos de tiempo más pequeños, usualmente de medio año, y ajustar el modelo, para así observar el recorrido de las betas, y ver qué tan adecuado es suponer que es constante.

4.2.2. Modelo de Regresión Simple Dinámico

Como mencionamos anteriormente, al utilizar el modelo de regresión simple, el supuesto de que α y β son constantes no es el más adecuado, es ahí donde entran los Modelos Dinámicos Lineales al considerar que tanto α como β cambian a través del tiempo. Resaltaremos de nuevo que una de las ventajas de los modelos de espacio de estados, es su flexibilidad, ya que éstos se pueden ir adaptando de acuerdo a las necesidades que se tengan. En el caso de modelar los rendimientos, en particular, se podría utilizar un modelo más general sin especificar una distribución como en el caso de los modelos dinámicos lineales, o si se tiene un conocimiento “profundo”, del comportamiento de las alfas o betas, éste se puede implementar en las ecuaciones de estados.

En nuestro caso mantendremos las cosas sencillas, consideraremos la generalización de los modelos de regresión a través de los modelos dinámicos lineales, con el supuesto de que tanto α como β se comportan como una caminata aleatoria. Con esta idea, podemos presentar a continuación el modelo de regresión dinámico, que es como el modelo de regresión simple presentado antes, pero se le agregó “dynamismo” al modelo:

$$\begin{aligned}r_{P,t} &= \alpha_{P,t} + \beta_{P,t}r_{M,t} + v_t \\ \alpha_{P,t} &= \alpha_{P,t-1} + w_{1,t} \\ \beta_{P,t} &= \beta_{P,t-1} + w_{2,t}\end{aligned}$$

que se puede reescribir como:

$$\begin{aligned}r_{P,t} &= F_t \theta_t + v_t \\ \theta_t &= G \theta_{t-1} + w_t\end{aligned}$$

donde:

$$F_t = [1 \quad r_{M,t}], \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \theta_t = \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}, \quad y \quad w_t = \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}$$

4.2.3. Estimación por Máxima Verosimilitud

Hay que notar que en el modelo existen dos parámetros desconocidos, las matrices de varianzas W_t y V_t , para solucionar esto estimaremos los parámetros usando el método de máxima verosimilitud, utilizando la función *dlnMLE* que viene implementada en *R*. Procediendo de esta manera obtenemos que:

$$v_t = 0.000227 \quad y \quad w_t = \begin{bmatrix} 7.18e^{-11} \\ 0.005353 \end{bmatrix}$$

Con estas estimaciones podemos implementar el modelo. Como ya habíamos mencionado, utilizaremos la función *dln*. Primero comentaremos los resultados obtenidos usando el filtrado. En la Figura 4.5 podemos observar el recorrido de la media de la distribución filtrada de α y β , las gráficas de los resultados tanto de la filtración como del análisis de regresión se mostrarán en las siguientes páginas.

De la gráfica del recorrido de α y β podemos observar que se obtuvo al final del periodo una $\alpha = 0.004055$ y una $\beta = 1.43774$, por lo que cabe mencionar que usando el modelo de regresión simple se obtuvo una beta diferente, de 1.187, mencionando que en el modelo de regresión simple, la alfa era significativamente igual a cero, aunque en este caso, en modelo dinámico también se obtuvo un cambio:

Regresión Simple	Usando DLM	Variación
alfa = 0.00393	alfa = 0.00406	3.202 %
beta = 1.18709	beta = 1.43774	17.434 %

Ahora observemos sus varianzas, esto a través de sus intervalos de confianza, en la Figura 4.6, lo podemos observar, notemos cómo es mayor la varianza al inicio del periodo y conforme pasa el tiempo éstas disminuyen, en un momento comentaremos sobre ello.

Ahora comentemos los resultados usando el análisis retrospectivo. En este caso no es de sorprenderse que al final del periodo se obtuvieron valores de alfa y beta muy similares a los obtenidos con la filtración, en la Figura 4.5 podemos observar su recorrido. Análogamente observemos sus varianzas a través de sus intervalos de confianza, que vemos en la Figura 4.7. En este caso démonos cuenta que su varianza es más estable, si bien en el caso de alfa parece mayor, esto se debe a que los valores de alfa no variaron mucho, pero la varianza es muy parecida al caso de filtración al final del periodo.

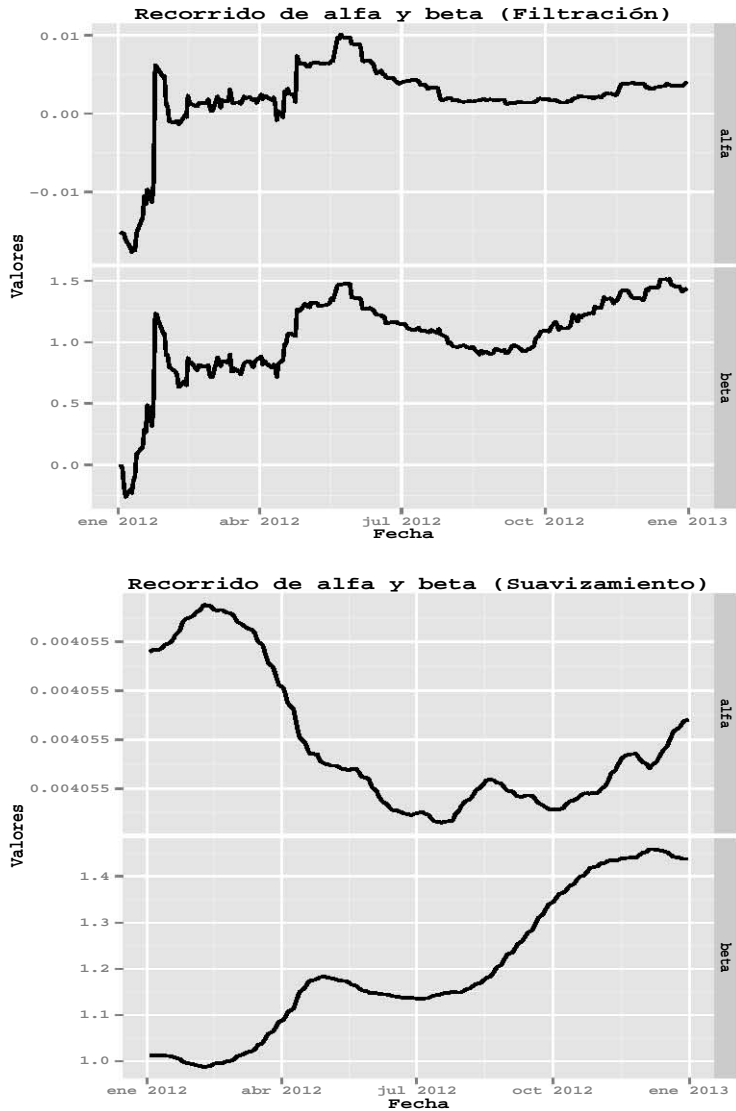


Figura 4.5: Recorrido de alfa y beta en el caso de suavizamiento y filtración. Se puede observar como existe una mayor variación de los valores obtenidos en el caso de filtración, además de que se logra percibir que el supuesto de los Modelos de Regresión Simple de que alfa y beta son constantes a través del tiempo, no es muy adecuado.

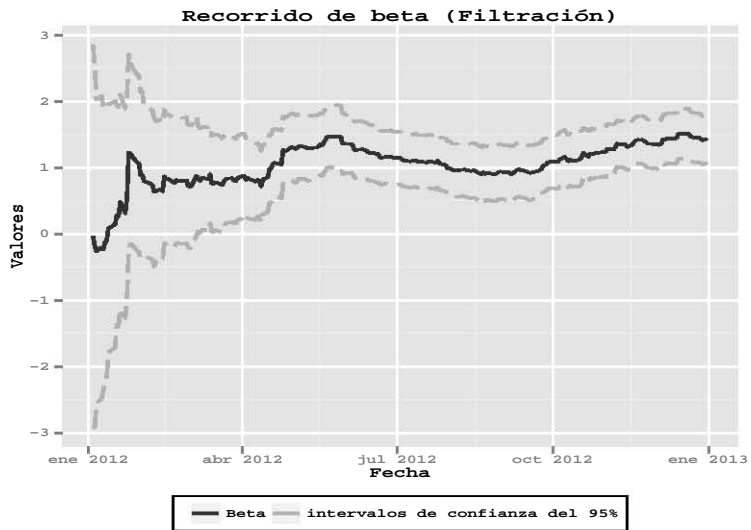
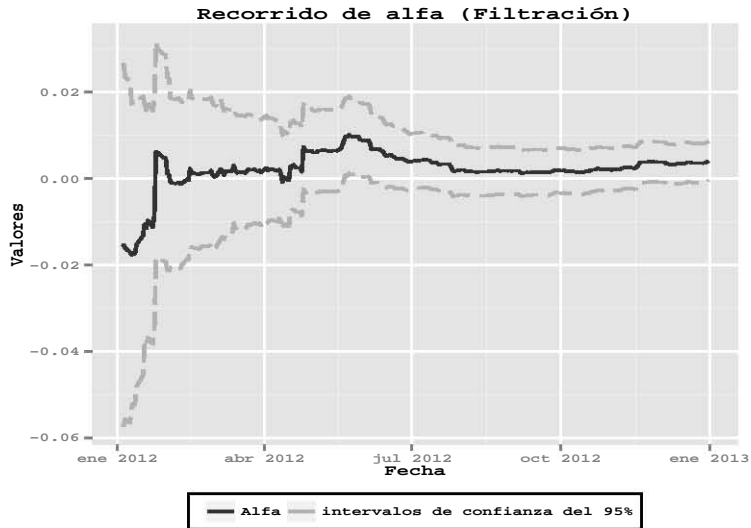


Figura 4.6: Recorrido de alfa y beta con sus intervalos de confianza del 95% usando filtración. Se puede observar como al inicio de los recorridos tanto de alfa como de beta existe una mayor varianza, esto se debe a que en el caso de la filtración, para obtener la distribución a un tiempo t , condicionamos a la información que poseemos a ese periodo.

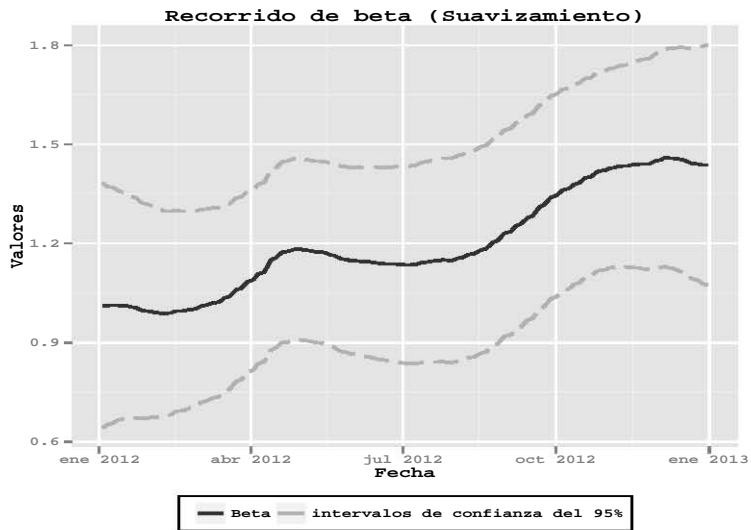
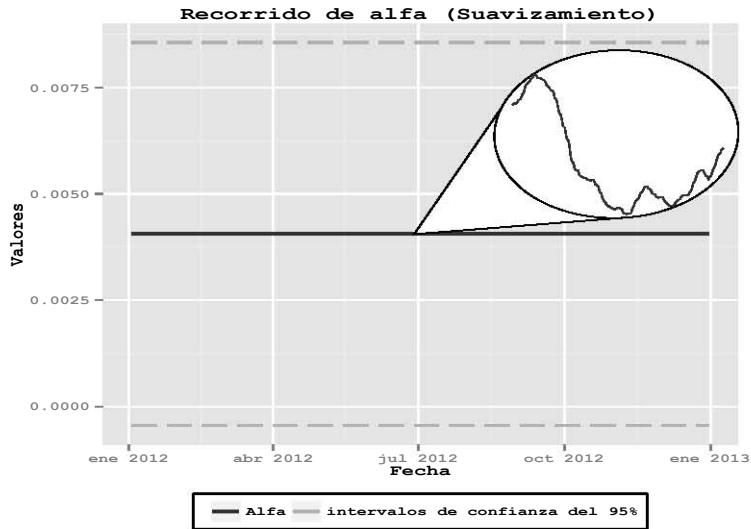


Figura 4.7: Recorrido de alfa y beta con sus intervalos de confianza del 95 % usando suavizamiento, en el caso de alfa se muestra además el recorrido en una escala que permite ver mejor su variación. Se puede observar como las varianzas de los recorridos tanto de alfa como de beta son más estables si las comparamos con el caso de filtración, de hecho se asemejan a la varianza que se tiene al final en el caso de filtración, esto se debe a que en el caso de suavizamiento, para obtener la distribución a un tiempo t , condicionamos a toda información que poseemos.

Observando en conjunto las gráficas podemos hacer algunos comentarios. El primero surge al observar cómo cambia la varianza usando la filtración a usando el análisis retrospectivo, este comportamiento es entendible si recordamos cómo se realizaba la filtración y el suavizamiento, por ejemplo, en el caso de la filtración, la distribución al tiempo 10, por decir algún valor, está condicionada con las primeras diez observaciones, mientras que en el caso del análisis retrospectivo estamos condicionando con respecto al total de las observaciones disponibles, por lo que si observamos al final del periodo sus varianzas son muy similares, es decir, los estados se actualizan por lo que sus varianzas también, por lo que si iniciamos con una varianza muy grande, como es nuestro caso, ésta se va ajustando al obtener información de los datos. Por ejemplo en el caso de alfa (la interpretación del caso de beta es similar), vemos como en el suavizamiento hay muy poca variación en los valores estimados, lo cuál es debido a lo mencionado, ya que aquí condicionamos a toda la información disponible y vamos obteniendo los valores anteriores sucesivamente, es decir con base a la información que disponemos estamos tratando de reconstruir el comportamiento que tuvo alfa a lo largo del tiempo, además de que en el caso de alfa su varianza una vez que condicionamos a toda la información es pequeña, lo cual también se ve reflejado en que los valores varíen poco, en la filtración el enfoque es distinto, tratamos de describir a alfa en un tiempo, en base a la información que disponemos a ese tiempo, por lo que podemos notar que conforme almacenamos más información, el comportamiento en el caso de suavizamiento y filtración se parecen, ya que al transcurrir el tiempo se asemeja la información que poseemos. Lógicamente estos dos enfoques tienen distintos usos, el caso de filtración es importante cuando buscamos monitorear el comportamiento de una serie casi al instante conforme vamos adquiriendo nuevo conocimiento, el caso de suavizamiento cobra importancia cuando, por decir, tenemos información de un evento a cierto tiempo, y buscamos reconstruir su comportamiento previo a ese tiempo, para poder observar algún cambio importante que ya sucedió en la serie, y así aumentar nuestro conocimiento sobre el evento de interés.

4.2.4. Estimación por MCMC

Como hemos venido recordando, un atractivo de los modelos dinámicos es su flexibilidad, a continuación mostraremos uno de los aspectos de los que nos referimos. Muchas veces al tratar de modelar cierto fenómeno en la realidad, nos encontramos con muchas limitantes, una de las que observamos en nuestro caso, es la estimación de ciertos parámetros desconocidos, ahí es donde puede ayudarnos el análisis bayesiano, con la característica de que éste considera

“*todo*” lo desconocido como variable aleatoria, y donde las conocidas como distribuciones iniciales muestran nuestras creencias acerca de los posibles valores de los parámetros desconocidos. La estimación de estos parámetros desconocidos se realiza usando el Teorema de Bayes, que actualiza nuestra creencias conforme nueva información es observada.

Usaremos este enfoque para la estimación de las varianzas en el modelo. Recurriremos a uno de los modelos más comunes para la estimación de varianzas, el modelo conocido como de las *d-gammas-inversas*. A continuación, mostraremos el desarrollo general del caso univariado, cuyos supuestos son que ambas matrices son invariantes ante el tiempo. Específicamente se tiene que:

$$V_t = \phi_y^{-1}, \quad W_t = \text{diag}(\phi_{\theta,1}^{-1}, \dots, \phi_{\theta,p}^{-1})$$

donde ϕ^{-1} ($= 1/\sigma^2$) es la precisión, y $\phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}$ son independientes con distribución gamma, por lo que el vector $(\phi_y^{-1}, \phi_{\theta,1}^{-1}, \dots, \phi_{\theta,p}^{-1})$ es el producto de $d = (p + 1)$ densidades Gammas-Inversas. Las precisiones a priori se pueden expresar de la siguiente manera, $E(\phi_y) = a_y$, y $E(\phi_{\theta,i}) = a_{\theta,i}$, cuya incertidumbre se refleja en las varianzas $\text{Var}(\phi_y) = b_y$, y $\text{Var}(\phi_{\theta,i}) = b_{\theta,i}$, $i = 1, \dots, p$, por lo que:

$$\phi_y \sim \Gamma(\alpha_y, \beta_y), \quad \phi_{\theta,i} \sim \Gamma(\alpha_{\theta,i}, \beta_{\theta,i}), \quad i = 1, \dots, p,$$

con

$$\alpha_y = \frac{a_y^2}{b_y}, \quad \beta_y = \frac{a_y}{b_y}, \quad \alpha_{\theta,i} = \frac{a_{\theta,i}^2}{b_{\theta,i}}, \quad \beta_{\theta,i} = \frac{a_{\theta,i}}{b_{\theta,i}}, \quad i = 1, \dots, p,$$

De las hipótesis de un sistema de estados (A.1) y (A.2) tenemos:

$$(\theta_0, \dots, \theta_n, y_1, \dots, y_n, \psi) \sim \prod_{t=1}^n \pi(y_t | \theta_t, \psi) \pi(\theta_t | \theta_{t-1}, \psi) \pi(\theta_0 | \psi) \pi(\psi)$$

usando este hecho, podemos obtener lo siguiente:

$$\begin{aligned}
\pi(y_{1:T}, \theta_{0:T}, \psi) &= \pi(y_{1:T} \mid \theta_{0:T}, \psi) \pi(\theta_{0:T} \mid \psi) \pi(\psi) \\
&= \prod_{t=1}^n \pi(y_t \mid \theta_t, \phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}) \pi(\theta_t \mid \theta_{t-1}, \phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}) \\
&\quad \pi(\theta_0 \mid \phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}) \pi(\phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}) \\
&= \prod_{t=1}^n \pi(y_t \mid \theta_t, \phi_y) \pi(\theta_t \mid \theta_{t-1}, \phi_{\theta,1}, \dots, \phi_{\theta,p}) \pi(\theta_0) \pi(\phi_y) \prod_{i=1}^p \phi_{\theta,i}
\end{aligned}$$

Ahora la idea es usar el método de Gibbs¹ en el modelo de las $d - \text{gammas} - \text{inversas}$, por lo que necesitamos obtener las distribuciones condicionales completas. En el caso de los estados, se puede usar el algoritmo conocido como FFBS² (Forward Filtering Backward Sampling), que consiste en calcular el filtro de Kalman, para poder simular los estados $\theta_T \sim \mathcal{N}(m_T, C_t)$, y para $t = T - 1, \dots, 0$ calcular θ_t a partir de la distribución de $\pi(\theta_t \mid \theta_{t+1}, y_{1:y})$, ahora para el caso de $\phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p}$, derivaremos un caso, y los demás son análogos:

$$\begin{aligned}
\pi(\phi_y \mid \dots) &= \frac{\pi(y_{1:T}, \theta_{0:T}, \phi_y, \phi_{\theta,1}, \dots, \phi_{\theta,p})}{\pi(y_{1:T}, \theta_{0:T}, \phi_{\theta,1}, \dots, \phi_{\theta,p})} \\
&\propto \prod_{t=1}^T \pi(y_t \mid \theta_t, \phi_y) \pi(\phi_y) \\
&\propto \prod_{t=1}^T \phi_y^{\frac{1}{2}} \exp \left\{ \frac{-\phi_y (y_t - F_t \theta_t)^2}{2} \right\} \phi_y^{\alpha_y - 1} \exp \{-\beta_y \phi_y\} \\
&= \phi_y^{\frac{T}{2}} \exp \left\{ \frac{-\phi_y \sum_{t=1}^T (y_t - F_t \theta_t)^2}{2} \right\} \phi_y^{\alpha_y - 1} \exp \{-\beta_y \phi_y\} \\
&= \exp \left\{ -\phi_y \left(\frac{\sum_{t=1}^T (y_t - F_t \theta_t)^2}{2} + \beta_y \right) \right\} \phi_y^{\left(\frac{T}{2} + \alpha_y\right) - 1}
\end{aligned}$$

¹Para una introducción a los métodos de Monte Carlo se puede consultar: Casella y Robert (2009), *Introducing Monte Carlo Methods with R*, Springer. En el Apéndice B se muestra el algoritmo de Gibbs.

²Para conocer más sobre el algoritmo FFBS se puede consultar: Carter, C. y Kohn, R. (1994), *On Gibbs sampling for state space models*, *Biometrika* 81:541-553.

Por lo tanto:

$$\phi_y | \dots \sim \Gamma \left(\frac{T}{2} + \alpha_y, \beta_y + \frac{-\phi_y \sum_{t=1}^T (y_t - F_t \theta_t)^2}{2} \right)$$

análogamente:

$$\phi_{\theta,i} | \dots \sim \Gamma \left(\frac{T}{2} + \alpha_{\theta,i}, \beta_{\theta,i} + \frac{-\phi_y \sum_{t=1}^T (\theta_{t,i} - G_t \theta_{t-1,i})^2}{2} \right)$$

En nuestro caso podemos escoger $a_y = 4405$ y $a_\theta = (1.393e+10, 186.8)'$, para mostrar una creencia inicial de que los valores son similares a los obtenidos por la estimación con máxima verosimilitud.

A continuación en las siguientes figuras (de la 4.8 a la 4.11) mostramos un breve diagnóstico de los resultados obtenidos de la simulación, donde decidimos realizar 12500 iteraciones con un periodo de calentamiento de 8500, esto ya que realizando varias simulaciones se determinó que a partir de 8000 iteraciones aproximadamente, parecía que la simulación empezaba a converger, se consideraron 12500 iteraciones más, ya que aunque se hubiera deseado realizar más iteraciones para notar si existía algún cambio considerable en la simulación o si se notaba una convergencia clara, esta al aumentar las iteraciones aumentaba también su duración de manera considerable. Mostramos 4 gráficas comparando 4 simulaciones (esto para cada parámetro: V , W_1 , y W_2), la primera muestra los valores obtenidos de la simulación, la segunda las densidades posteriores de los parámetros, la tercera es la gráfica de las medias móviles de los valores de la simulación, es decir, la media hasta cierta iteración, y la última es la de las autocorrelaciones, en todos los casos, se descartaron los valores iniciales de calentamiento. Al observar la gráfica de los valores de las simulaciones y de las densidades posteriores, pareciera que la simulación no tiene problemas en cuanto a la mezcla, pero parece que aún no podríamos decir que converge, en el caso de W_1 y W_2 pareciera que no existe problema, pero en particular en V , podemos notar un problema, más explícitamente en la gráfica de las medias móviles, ya que parece que aún no se estabilizan, algunas de las posibles soluciones son aumentar el número de iteraciones, en nuestro caso es algo costoso en tiempo, ya que la duración para terminar la simulación al aumentar las iteraciones aumenta de manera importante, por lo que se necesitaría revisar la implementación y ver si el algoritmo se puede mejorar, o variar la densidad a priori, esto es útil además para contrastar que tan adecuada fue

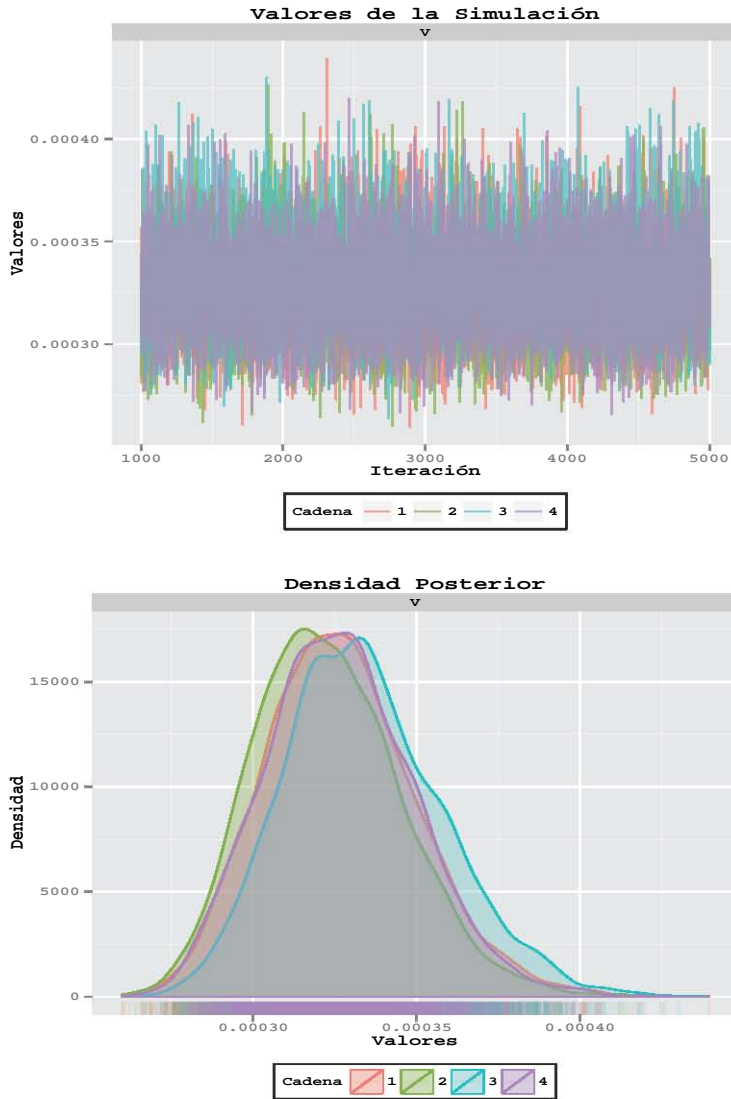


Figura 4.8: Gráficas del diagnóstico de la simulación del parámetro V : Arriba tenemos la de los valores obtenidos de la simulación, y abajo la de las densidades posteriores. En las gráficas podemos observar que parece no haber problema en cuanto a la mezcla de la simulación, tanto por los valores que se obtuvieron en las 4 cadenas, como en sus distribuciones posteriores.

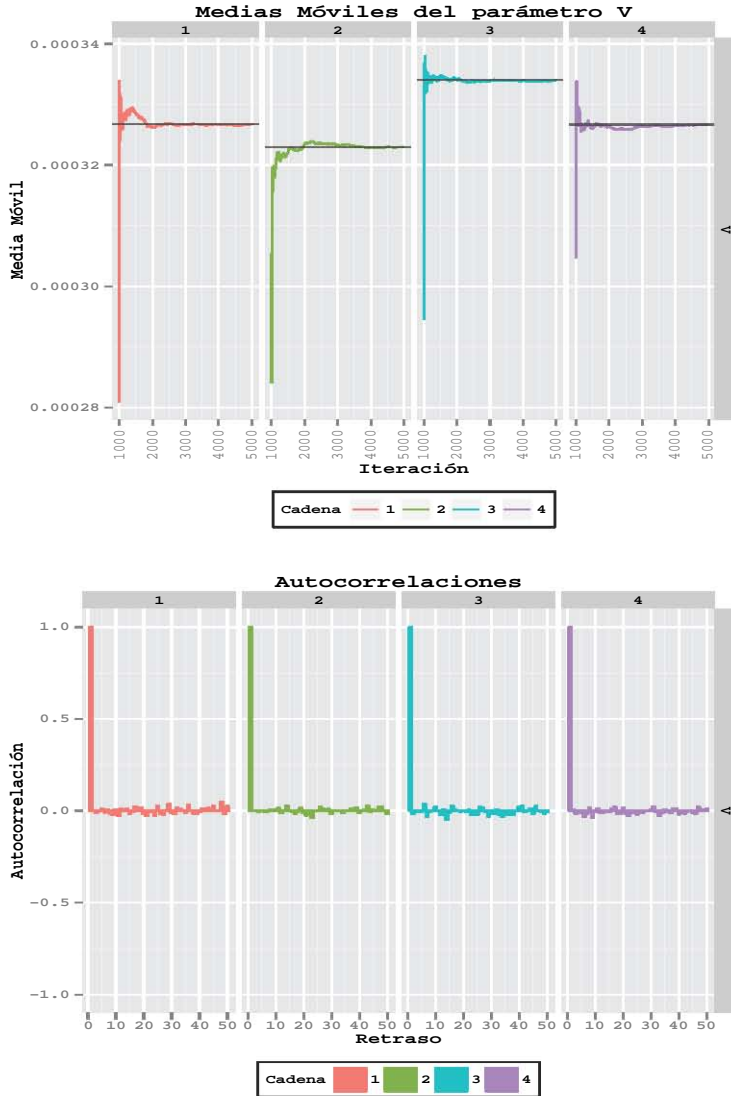


Figura 4.9: Gráficas del diagnóstico de la simulación del parámetro V: Arriba tenemos la de medias móviles de la simulación, y abajo la de las autocorrelaciones. Podemos observar con la gráfica de las medias móviles, que aún no podemos garantizar completamente la convergencia de la simulación, en el caso de las autocorrelaciones no se observa ningún problema.

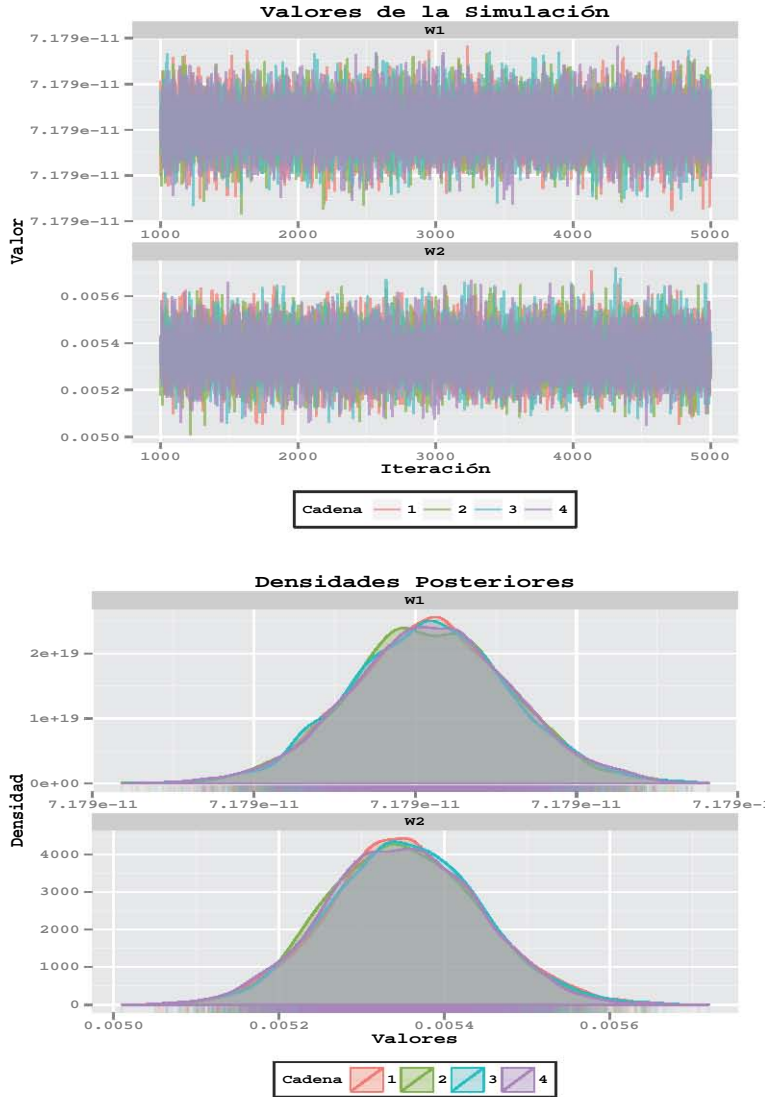


Figura 4.10: Gráficas del diagnóstico de la simulación de los parámetros W_1 y W_2 : Arriba tenemos la de los valores obtenidos de la simulación, y abajo la de las densidades posteriores. En las gráficas podemos observar que parece no haber problema en cuanto a la mezcla de la simulación, tanto por los valores que se obtuvieron en las 4 cadenas, como en sus distribuciones posteriores.

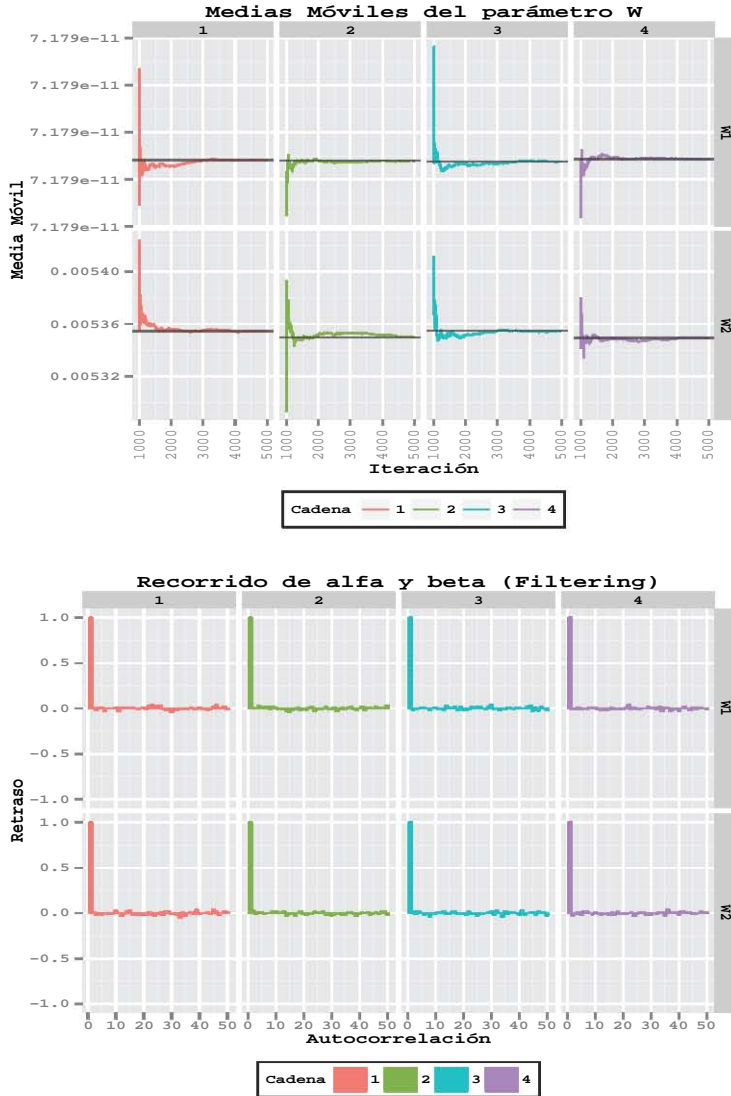


Figura 4.11: Gráficas del diagnóstico de la simulación de los parámetros W_1 y W_2 : Arriba tenemos la de medias móviles de la simulación, y abajo la de las autocorrelaciones. Podemos observar con la gráfica de las medias móviles, que contrario al caso del parámetro V_1 , parece que la simulación está empezando a converger, aunque es importante mencionar que estos criterios son los básicos para poder decir que la simulación converge, en el caso de las autocorrelaciones no se observa ningún problema.

nuestra distribución a priori, lo laborioso para nosotros en este caso es que tendríamos que analizar que distribución a priori es adecuada para los datos, y en base a ella realizar una nueva implementación del algoritmo, pero debido a que estamos enfocados más a la implementación de los DLM, supondremos que podemos en cierta medida estar seguros de la convergencia de la simulación y usaremos estas simulaciones. Cabe resaltar que no existe un criterio absoluto para la evaluación de la convergencia de una cadena de Markov, no importa las realizaciones que hagamos de ésta. Además de que los criterios que mostramos serían los considerados como básicos, por lo que aunque no existiera un problema con las medias móviles en el caso de V , esto no garantizaría que ya se haya alcanzado la convergencia, por ejemplo, existen otros métodos más complejos para evaluar la convergencia, como el factor de reducción (Gelman y Rubin, (1992)), que hacen una comparación de varianzas. Entonces para calcular las medias a posteriori de las varianzas que se desconocen, podemos usar la función *mcmcMean*, teniendo como resultado:

$$v_t = 0.000334 \quad y \quad w_t = \begin{bmatrix} 7.18e^{-11} \\ 0.00535 \end{bmatrix}$$

Observemos que el resultado es parecido al que obtuvimos usando el método de máxima verosimilitud, la razón de esto queda más claro si hacemos cita de dos teoremas:

- **Teorema del límite central para estimadores máximos verosímiles:** Bajo los supuesto adecuados y para una muestra lo suficientemente grande, el estimador de máxima verosimilitud se distribuye aproximadamente normal con media igual al verdadero parámetro y varianza igual a la inversa de la información de Fisher.
- **Teorema de Bernstein-von Mises:** Bajo los supuesto adecuados y para una muestra lo suficientemente grande, la distribución posterior de θ es aproximadamente normal con media igual al verdadero parámetro θ y varianza igual a la inversa de la información de Fisher.

es decir, que el estimador bayesiano y el de máxima verosimilitud tienen la misma distribución muestral, en una muestra lo suficientemente grande. Uno de los supuestos del segundo teorema, es que la distribución a priori no cambia conforme se obtienen nuevos datos, es decir, toda la información que se obtiene proviene de los datos iniciales. Los resultados al implementar las estimaciones de la varianza al modelo, son similares a los de la sección anterior, en este caso el resultado al final del periodo fue; $\alpha = 0.0038$ y $\beta = 1.4091$. A continuación mostramos las figuras (de la 4.12 a la 4.14) de los resultados obtenidos.

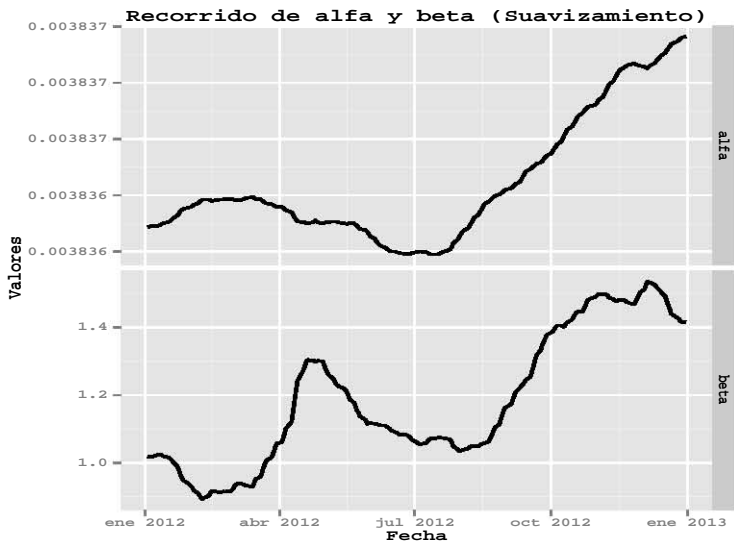
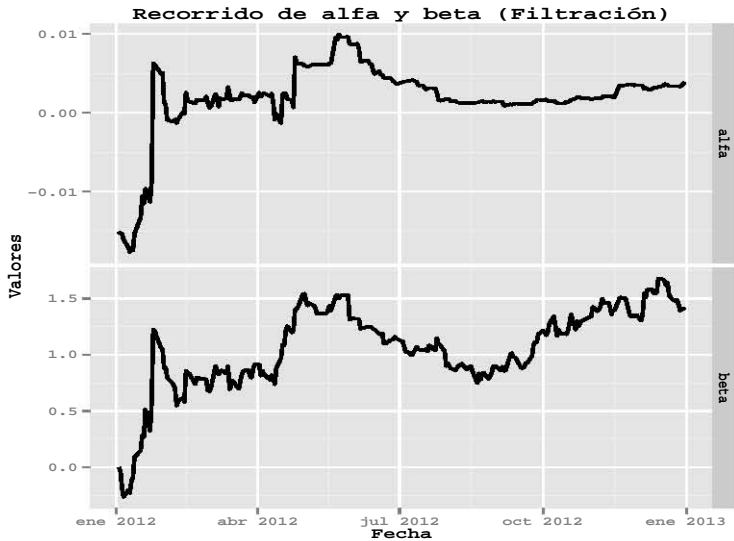


Figura 4.12: Recorrido de alfa y beta en el caso de suavizamiento y filtración. Se puede observar como existe una mayor variación de los valores obtenidos en el caso de filtración, además de que se logra percibir que el supuesto de los Modelos de Regresión Simple de que alfa y beta son constantes a través del tiempo, no es muy adecuado.

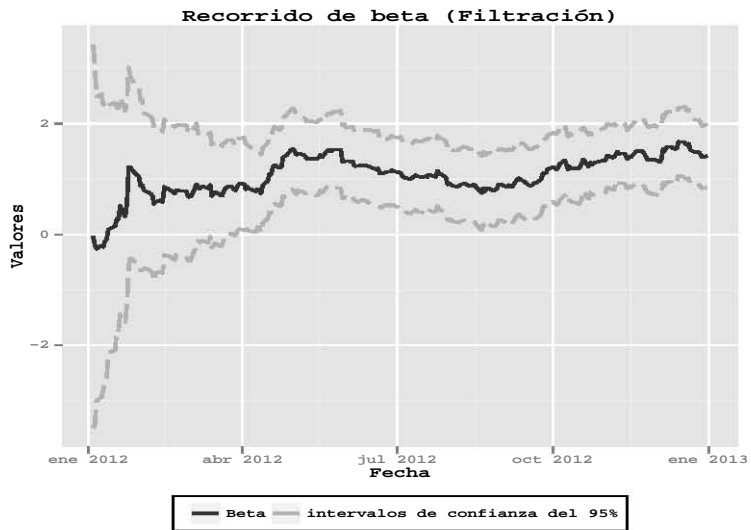
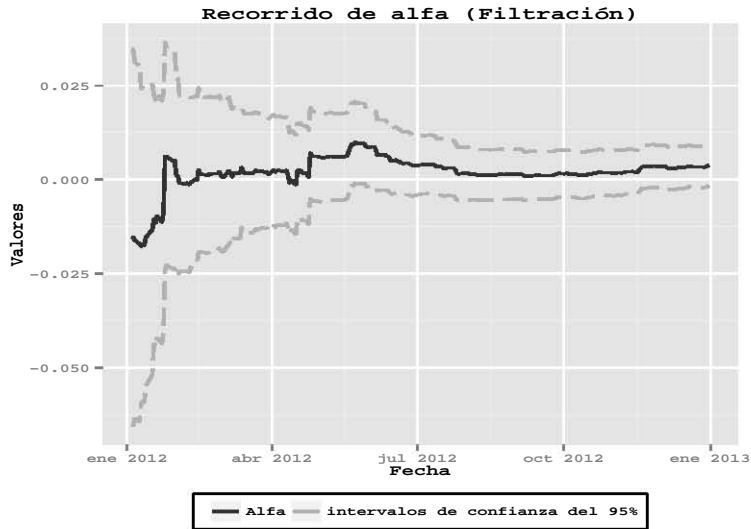


Figura 4.13: Recorrido de alfa y beta con sus intervalos de confianza del 95% usando filtración. Se puede observar como al inicio de los recorridos tanto de alfa como de beta existe una mayor varianza, esto se debe a que en el caso de la filtración, para obtener la distribución a un tiempo t , condicionamos a la información que poseemos a ese periodo.

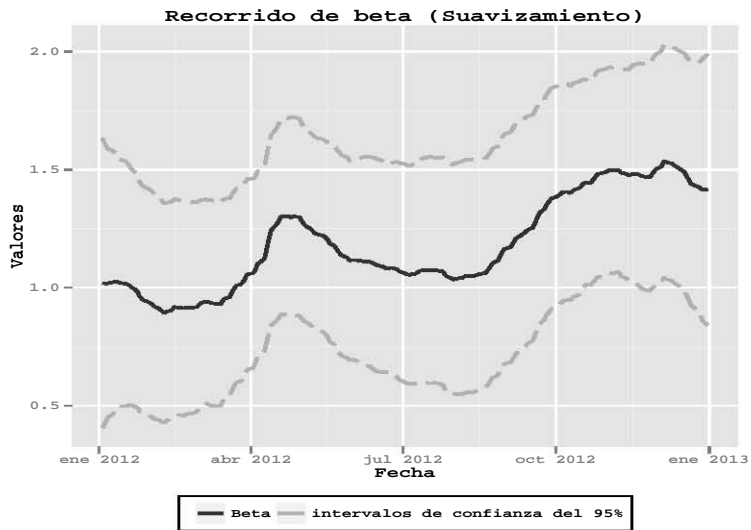
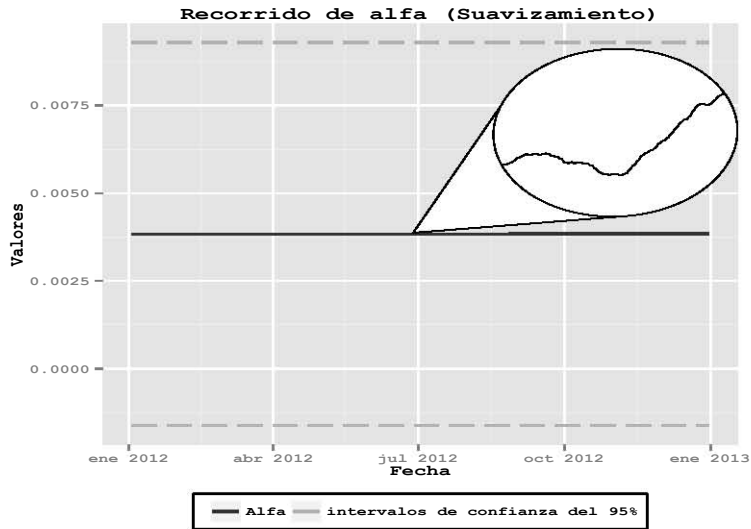


Figura 4.14: Recorrido de alfa y beta con sus intervalos de confianza del 95 % usando suavizamiento, en el caso de alfa se muestra además el recorrido en una escala que permite ver mejor su variación. Se puede observar como las varianzas de los recorridos tanto de alfa como de beta son más estables si las comparamos con el caso de filtración, de hecho se asemejan a la varianza que se tiene al final en el caso de filtración, esto se debe a que en el caso de suavizamiento, para obtener la distribución a un tiempo t , condicionamos a toda información que poseemos.

Cabe mencionar que este método bayesiano para estimar las varianzas que se escogió, sirvió de acuerdo al problema que teníamos, pero usualmente la estimación mediante el enfoque bayesiano usando MCMC, se realiza de acuerdo al problema que se tiene, es decir, esta estimación que usamos no funciona para todos los problemas, por lo que si nos encontramos en otra situación, tendríamos que buscar y adaptar las ideas del enfoque bayesiano a nuestro problema.

También otro punto que no mencionamos fueron las predicciones, puesto que en nuestra implementación del CAPM, no se usa para la predicción de las betas, sino para estimación de acuerdo a los datos con los que se dispone, y así poder utilizar ese conocimiento en distintos problemas, uno muy común es la valuación de proyectos. Independientemente de eso, en la Figura 4.15 podemos ver cómo usando las predicciones que proporcionan la filtración, la varianza aumenta de forma considerable, en la figura se usaron los datos obtenidos con los estimadores bayesianos de las varianzas y un periodo de 3 meses de predicción.

Entonces, por lo que podemos observar de la figura, podemos tener una idea del problema que tendríamos al utilizar predicciones de betas considerando horizontes de tiempo para predicciones medianamente grandes, cuando además

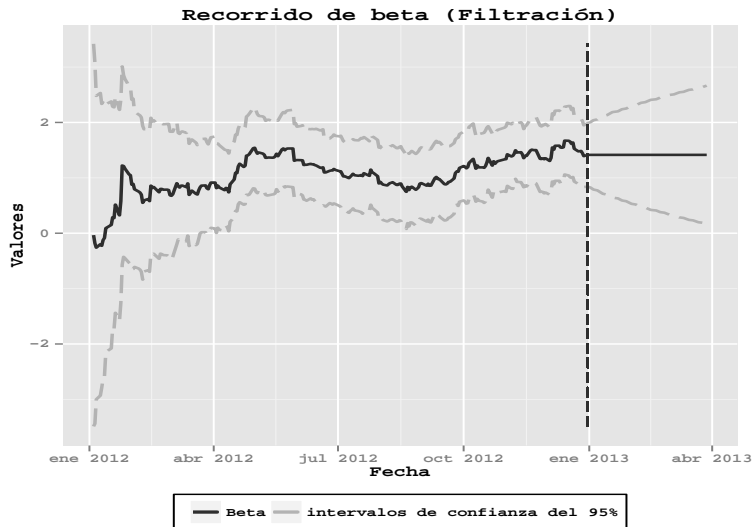


Figura 4.15: Evolución de la varianza, usando filtración.

otro de los puntos por los que no se busca predecir los valores de beta, es también por que los datos se van obteniendo de una manera “*continua*”. Sobre este aspecto de que la información fluya de manera continua, haremos un pequeño comentario en un momento.

Como mencionamos casi al inicio del apartado, es importante mencionar que el CAPM es una teoría y como tal es susceptible a equivocaciones, hacemos mención de esto por que hasta el momento lo que hicimos fue implementar el modelo, ahora mencionaremos algunos aspectos en los que se debe tener cuidado. Es muy importante la selección de los datos, anteriormente comentamos que en el caso de usar un modelo de regresión lineal simple para ajustar los datos, suponía que nuestra beta era constante a través del tiempo, a continuación mostramos una gráfica que nos da una idea de por qué este supuesto se puede considerar erróneo, en la Figura 4.16, se muestra lo que se conoce como betas móviles, la idea es similar a las medias móviles, considerando periodos de 60 días, que lo que hace es ajustar el modelo de regresión en periodos, usualmente se le conocen como ventanas, de 60 días, recorriendo el periodo donde disponemos de datos, es decir, consideramos los primeros 60 datos que corresponden a los días 1 al 60, y ajustamos un modelo de regresión, luego consideramos del día 2 al 61, y ajustamos nuevamente el modelo, así sucesivamente, a partir de los valores obtenidos de este procedimiento podemos observar cómo ajustando el modelo con los primeros 60 días, obtenemos una $\beta = 0.8$, y al ajustarlo con los últimos 60 días, donde disponemos de información, $\beta = 1.78$, y en los periodos intermedios los valores varían. Por lo que no sería correcto suponer que beta es constante. También es importante volver a mencionar que por lo que podemos observar también en la gráfica, el periodo de tiempo que consideremos es importante, en el caso de un modelo de regresión lineal simple, no es adecuado considerar un periodo muy largo, por lo que acabamos de mencionar, en el caso de los DLM, esto nos beneficia, ya que con los datos es como obtenemos la información.

Este no es el único problema con los supuestos, por ejemplo, el suponer que los rendimientos se distribuyen de manera normal. Con una gráfica de sus cuantiles (qq-plot, Figura 4.17) se puede observar que no es adecuado³, a este respecto, en la realidad existen variaciones del CAPM, en algunas de ellas suponen distribuciones que se adecuen más a los datos que se disponen, estas variaciones del CAPM, se deben también a que, el considerar al mercado como único factor para modelar los rendimientos (a estos modelos se les conoce como modelos de índice único), no es lo más adecuado, por lo que usualmente se

³Para una mejor prueba existen por ejemplo la prueba de Anderson-Darling, que ya está implantada en R.

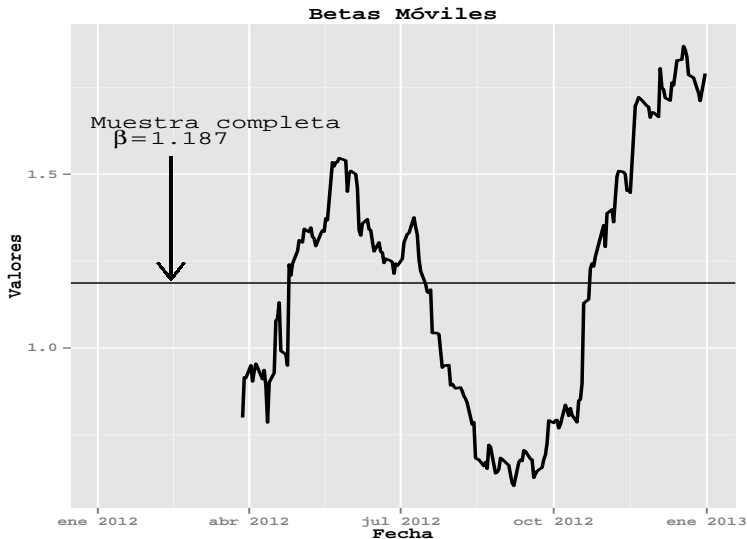


Figura 4.16: Betas Móviles, usando una ventana de 60 días.

consideran múltiple factores (betas), esto lo mencionamos por que uno de las posibles pruebas para validar los modelos dinámicos lineales es que los residuales se distribuyen normalmente de forma independiente, y por lo mencionado anteriormente, en nuestro caso surgen algunos problemas principalmente en lo referente a la normalidad. En el caso de la autocorrelación e independencia podemos observar la Figura 4.17, donde la primera gráfica muestra el qq-plot de los residuos, en la segunda los residuos obtenidos, en la tercera se observa la gráfica de sus autocorrelaciones y en la última los p-values de la prueba de Ljung-Box, cuya hipótesis nula es que los datos se distribuyen de manera independiente. Otro aspecto que tiene que ver con la selección de los datos, es que se tiene que tener en cuenta el índice de mercado que se utiliza, usualmente se suele utilizar uno que se refiera al área en el que estamos tomando nuestra acción. En nuestro caso consideramos como acción Apple, e índice a *SP&500*, pero se observaría un cambio en el ajuste del modelo si consideráramos como índice el *NASDAQ*, que contempla acciones ligadas a la tecnología. Concluimos diciendo que para poder acercarnos a modelos que expliquen mejor la realidad, es necesario comprender el fenómeno que se intenta modelar, para estar conscientes de cómo los cambios en algunos supuestos afectan los resultados.

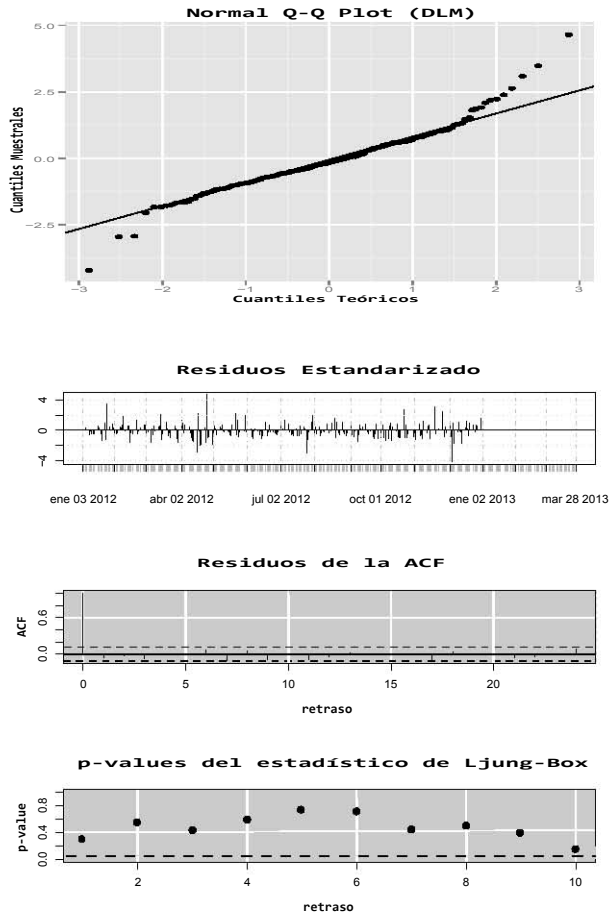


Figura 4.17: Gráficas para el diagnóstico de los residuos del Modelo Dinámico Lineal. Observamos como existe un problema con el supuesto de Normalidad.

Parte V

Conclusiones

Comentarios Finales

Uno de los aspectos más fuertes de los Modelos de Estados, es su flexibilidad. Vimos que a partir de una idea sencilla y general como considerar a los fenómenos que se pueden describir mediante ciertos estados en cierto momento y darle un marco teórico, hace que su aplicación sea muy variada. Es importante mencionar que únicamente nos enfocamos a un caso muy particular de los modelos de espacio de estados, en los modelos dinámicos lineales, los cuales simplifican en gran medida el trabajo de la implementación. Pero no por ello dejan de ser “*poderosos*”, aunque en este caso se intercambia parte de la flexibilidad del modelo, por simplicación. Aún así el modelo hace posible una variedad de modificaciones en las implementaciones. Para mostrarlas rescatemos algunos puntos mencionados en la tesis, es posible modelar la estacionalidad y la tendencia de las series de tiempo, o describirlas como el conjunto de varios factores (modelos de regresión dinámica), pero el modelo va más lejos, mediante la técnica de descomposición aditiva, permite pensar en la serie como la unión de distintos componentes, donde cada componente puede describirse mediante un modelo dinámico; además, esta idea puede trasladarse al caso de series de tiempo multivariadas, donde se busca aprovechar la correlación de dichas series. En el caso de series multivariadas existen otras implementaciones, como los modelos dinámicos lineales jerárquicos, que consisten en considerar que los estados están determinados por otra variable, esta idea se puede extender, es decir, se puede considerar también que esta nueva variable está a su vez determinada por otra variable y así sucesivamente⁴, que no mencionamos en la tesis, pero que proporcionan aún más flexibilidad al modelo.

Es importante decir, que esta flexibilidad también se podría considerar un “*arma de doble filo*” si no se maneja de manera adecuada, su uso inadecuado puede llevar a resultados errados o que no tienen nada que ver con el fenómeno que se intenta modelar, es decir, muchas de las implementaciones que se le pueden hacer al modelo, recaen en qué tanto conocimiento se tiene del fenómeno

⁴Para una introducción a estos modelos se puede consultar: Gamerman, D. y Migon, H. (1993) *Dynamic hierarchical model*, Journal of the Royal Statistical Society, Series B 55: 629-642.

de interés, y cómo se puede usar en el modelo.

En la parte de la implementación, escogimos una implementación simple pero que permitiera notar, de una manera sencilla y general, el uso que se le puede dar a los modelos dinámicos lineales. Un punto llamativo, es que debido a su construcción, es posible calcular los parámetros del modelo de manera recursiva, esta ventaja está más relacionada con la implementación del modelo y con su representación computacional, que consistiría en aplicar una función iterativa a los datos, como el filtro de Kalman. Aunque un aspecto que se podría mejorar es lo referente a la retroalimentación cuando llega nueva información, ya que en este caso, se tendría que correr la función con todos los datos de nuevo, y si el flujo de información es continua, o casi continua, esto no es eficiente. Existe una solución a este problema, que usa las ideas de las MCMC, a estos métodos se les conoce como métodos de Monte Carlo Secuenciales⁵, que no mencionamos en la tesis, pero puede tenerse en cuenta.

En lo referente al caso escogido para la implementación, consideramos a los CAPM por su sencillez pero con un importante uso en la práctica, aunque como vimos, al ser más un teoría, muchos de los supuestos del modelo no son tan “verosímiles”, por lo cual existen variaciones del modelo que lo mejoran. El uso de los modelos dinámicos lineales, como vimos, fue sencillo y de fácil interpretación, muy similar a los modelos de regresión lineal simple, pero se tiene que diferenciar dos aspectos en el uso del modelo, uno es, la facilidad con que se pueden interpretar los resultados, y el otro es su representación, es decir, cómo llevamos la parte teórica a la realidad, en este caso, nosotros obtuvimos gran ayuda de los paquetes de R, pero si lo que se busca es su representación en otros lenguajes, se tendrían que considerar también otros aspectos de programación. En la parte de la interpretación del modelo logramos observar su funcionamiento, pero en la representación, nuestra intervención fue algo limitada.

En conclusión, los Modelos Dinámicos son como una caja de herramientas muy amplia, que tiene tanto herramientas de uso muy sencillo, así como otras que requieren mayor práctica para poder obtener el mayor provecho posible. Nosotros únicamente usamos unas pocas herramientas de un compartimiento, en particular los modelos dinámicos lineales, pero esperamos seguir explorando las restantes.

⁵Para conocer más sobre los métodos de Monte Carlo Secuenciale se puede consultar: Liu J. (2001), *Monte Carlo Strategies in Scientific Computing*, Springer.

Apéndice A

Procedimiento completo de la demostración del inciso (iii), de la proposición 2.1 (Filtro de Kalman);

Consideremos un DLM especificado en 1.1, y supongamos que:

$$\theta_{t-1} \mid y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$$

Entonces se cumple que la distribución filtrada de θ_t dado $y_{1:t}$ es Gaussiana, con parámetros:

$$\begin{aligned} m_t &= E(\theta_t \mid y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t, \\ C_t &= \text{Var}(\theta_t \mid y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t, \end{aligned}$$

donde $e_t = Y_t - f_t$, el error de predicción.

Antes de proceder con el desarrollo, se presentará un pequeño resultado sobre matrices, necesario para la demostración:

Teorema .1 (La Inversa de una Suma) *Sean $A \in \mathcal{M}_m$ y $B \in \mathcal{M}_n$ matrices no singulares y sean C y D matrices de orden $m \times n$ y $n \times m$, respectivamente. Si la matriz $A + CBD$ es no singular, entonces:*

$$(A + CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}$$

DEMOSTRACIÓN. Observemos primero que $(B^{-1} + DA^{-1}C)$ es una matriz no singular:

$$\begin{aligned} \det(B^{-1} + DA^{-1}C) &= \det(B^{-1})\det(I_n + BDA^{-1}C) \\ &= \det(B^{-1})\det(I_m + A^{-1}CBD) \\ &= \det(B^{-1})\det(A^{-1})\det(A + CBD) \neq 0 \end{aligned}$$

Así:

$$\begin{aligned}
& (A+CBD)^{-1}(A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}) \\
&= I_m - C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} + CBDA^{-1} \\
&\quad - CBD^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} \\
&= I_m - C((B^{-1} + DA^{-1}C)^{-1} - B \\
&\quad + BDA^{-1}C(B^{-1} + DA^{-1}C)^{-1})DA^{-1} \\
&= I_m C((I_n + BDA^{-1}C)(B^{-1} + DA^{-1}C)^{-1} - B)DA^{-1} \\
&= I_m - C(B(B^{-1} + DA^{-1}C)(B^{-1} + DA^{-1}C)^{-1} - B)DA^{-1} \\
&= I_m - C(B - B)DA^{-1} = I_m
\end{aligned}$$

■

Ahora continuemos con la demostración, la idea es aplicar el teorema de Bayes utilizando como distribución a priori a $\pi(\theta_t | y_{1:t-1})$ y como función de verosimilitud a $\pi(y_t | \theta_t)$, lo cuál lleva a un problema de inferencia bayesiana, notemos que la función de verosimilitud tiene una distribución normal, más específicamente, $Y_t | \theta_t \sim \mathcal{N}(F_t\theta_t, V_t)$, por lo que se puede expresar de la siguiente manera:

$$\begin{aligned}
\pi(Y_t | \theta_t) &\propto |V|^{-1/2} \exp\left\{-\frac{1}{2}(y_t - F_t\theta_t)'V^{-1}(y_t - F_t\theta_t)\right\} \\
&= \exp\left\{-\frac{1}{2}(y_t'V^{-1}y_t - 2\theta_t'F_t'V^{-1}y_t + \theta_t'F_t'V^{-1}\theta_t F_t')\right\}
\end{aligned}$$

viendola como función de θ_t podemos observar que es proporcional al kernel de una distribución normal $\mathcal{N}(m, C)$, con:

$$\begin{aligned}
m &= (F_t'V^{-1}F_t)^{-1}F_t'V^{-1}y_t \\
C &= (F_t'V^{-1}F_t)^{-1}
\end{aligned}$$

aplicando la fórmula de Bayes, tenemos lo siguiente:

$$\begin{aligned}
\pi(\theta | y_{1:t}) &\propto a \text{ priori} \times \text{verosimilitud} = \pi(\theta | y_{1:t-1})\pi(\theta | y_t) \\
&= \exp\left\{-\frac{1}{2}(\theta_t - a_t)'R_t^{-1}(\theta_t - a_t)\right\}\exp\left\{-\frac{1}{2}(F_t'V^{-1}F_t\theta - 2\theta'F_t'V^{-1}y_t)\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\theta_t'F_t'V^{-1}F_t + R_t^{-1})\theta_t - 2\theta_t'(F_t'V^{-1}y_t + R_t^{-1}a_t)\right\}
\end{aligned}$$

donde fijándose bien, se puede identificar el kernel de una distribución normal $\mathcal{N} \sim (m_t, C_t)$, con:

$$\begin{aligned} m_t &= C_t(F_t'V^{-1}y_t + R_t^{-1}a_t) \\ C_t &= (R_t^{-1} + F_t'V^{-1}F_t)^{-1} \end{aligned}$$

Aplicando el resultado del Teorema .1 a C_t , obtenemos que:

$$\begin{aligned} C_t &= R_t + R_tF_t'(F_tR_tF_t' + V)^{-1}F_tR_t \\ &= R_t + R_tF_t'Q_t^{-1}F_tR_t \end{aligned}$$

y sustituyendo en m_t , tenemos;

$$\begin{aligned} m_t &= C_t(F_t'V^{-1}y_t + R_t^{-1}a_t) \\ &= (R_t + R_tF_t'Q_t^{-1}F_tR_t)(F_t'V^{-1}y_t + R_t^{-1}a_t) \\ &= R_tF_t'V^{-1}\theta_t + a_t - R_tF_t'Q_t^{-1}F_tR_t(F_t'V^{-1}\theta_t) - R_tF_t'Q_t^{-1}F_t a_t \\ &= a_t - R_tF_t'Q_t^{-1}F_t a_t + R_tF_t'(V^{-1} - Q_t^{-1}F_tR_tF_t'V^{-1})\theta_t \end{aligned}$$

se cumple que, $(V^{-1} - Q_t^{-1}F_tR_tF_t'V^{-1}) = Q_t^{-1}$, pues;

$$\begin{aligned} Q_t(V^{-1} - Q_t^{-1}F_tR_tF_t'V^{-1}) &= Q_tV^{-1} - Q_tQ_t^{-1}F_tR_tF_t'V^{-1} \\ &= (F_tR_tF_t' + V)V^{-1} - F_tR_tF_t'V^{-1} \\ &= I + F_tR_tF_t'V^{-1} - F_tR_tF_t'V^{-1} \\ &= I \end{aligned}$$

con lo que llegamos a;

$$\begin{aligned} m_t &= a_t - R_tF_t'Q_t^{-1}F_t a_t + R_tF_t'Q_t^{-1}\theta_t \\ &= a_t + R_tF_t''Q_t^{-1}(\theta_t - a_t) \\ &= a_t + R_tF_t''Q_t^{-1}e_t \end{aligned}$$

y concluimos que:

$$\begin{aligned} m_t &= E(\theta_t | y_{1:t}) = a_t + R_tF_t'Q_t^{-1}e_t, \\ C_t &= Var(\theta_t | y_{1:t}) = R_t - R_tF_t'Q_t^{-1}F_tR_t, \end{aligned}$$

Apéndice B

El algoritmo de Gibbs es uno de los métodos más sencillos de Cadenas de Markov de Monte Carlo (MCMC, por sus sigla en inglés), además de ser un caso especial del algoritmo más general de Metropolis-Hasting. El algoritmo de Gibbs genera variables aleatorias a partir de una distribución (marginal) de manera indirecta, sin necesidad de calcular la densidad, lo que permite descomponer problemas complejos, por ejemplo cuando la función objetivo (la función de la cual se desea obtener una muestra) es de una dimensión grande, en pequeños problemas donde la función objetivo es de menor dimensión.

Supongamos que deseamos obtener una muestra de la variable aleatoria X , y además que para alguna $p > 1$, X puede ser escrita como $X = (X_1, \dots, X_p)$, donde las X_i son variables aleatorias univariadas o multivariadas. Uno de los supuestos del algoritmo, es que podemos simular las densidades condicionadas f_1, \dots, f_p , es decir; $X_i | \sim x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p \sim f_i(x_i | \sim x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$. Entonces para $i = 1, \dots, p$, el algoritmo de Gibbs es el siguiente:

Algoritmo de Gibbs:

Para $t = 1, 2, \dots$, dado $x^{(t)} = (x_1^{(t)}, \dots, x_p^{(t)})$, generamos:

1. $X_1^{(t+1)} \sim f_1(x_1 | \sim x_2^t, \dots, x_p^t)$;
2. $X_2^{(t+1)} \sim f_2(x_2 | \sim x_1^{t+1}, x_3^t, \dots, x_p^t)$;
- \vdots
- p . $X_p^{(t+1)} \sim f_p(x_p | \sim x_1^{t+1}, \dots, x_{p-1}^{t+1})$;

Las densidades f_1, \dots, f_p , son llamadas *densidades condicionales completas*, observe que en el algoritmo son las únicas densidades que se utilizan, pero es necesario poder simular de ellas para poder implementarlas, lo cual no siempre

es posible, por lo que se podrían utilizar métodos como los ARS o ARMS, u otros métodos de Monte Carlo⁶.

⁶Para una introducción a los métodos de Monte Carlo se puede consultar: Casella y Robert (2009), *Introducing Monte Carlo Methods with R*, Springer.

Bibliografía

- [1] Richard C. Grinold y Ronald N. Kahn (1999), “*Active Portfolio Management*”, Irwin Library.
- [2] Scott M. Lynch (2007), “*Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*”, Springer, New York.
- [3] Giovanni Petris, Sonia Petrone, y Patrizia Campagnoli (2009), “*Dynamic Linear Models with R*”, Springer, New York.
- [4] Casella y Robert (2009), “*Introducing Monte Carlo Methods with R*”, Springer, New York.
- [5] David Ruppert (2010), “*Statistics and Data Analysis for Financial Engineering*”, Springer, New York.
- [6] Eric Zivot y Guy Yollin (2012), “*Time Series Forecasting with State Space Models*”, University of Washington.