

2 of 38



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

MODELOS PARA SERIES DE TIEMPO:
UNA PERSPECTIVA BAYESIANA

T E S I S

PARA OBTENER EL TITULO DE

A C T U A R I A

P R E S E N T A

SILVIA SANCHEZ MEXICANO

DIRECTOR DE TESIS:

DR. GUSTAVO JAVIER VALENCIA RAMIREZ

CD. UNIVERSITARIA D. F.

NOVIEMBRE, 1989

TESIS CON
FALLA DE ORIGEN



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 está 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.

INDICE

Págs.

Prólogo

CAPITULO I El problema de predicción

1.1	Conceptos Básicos	1
1.2	Introducción	2
1.2.1	La Importancia de la predicción	2
1.2.2	Características que presentan las predicciones	3
1.3	Métodos de predicción	4
1.3.1	Modelos Matemáticos y Estadísticos	5
1.3.1.1	Modelos Econométricos	6
1.3.1.2	Modelos Determinísticos	6
1.3.1.3	Modelos ARIMA	7
1.3.1.4	Modelos Lineales Dinámicos	8
1.3.2	Otros modelos de predicción	8

CAPITULO II Conceptos Fundamentales en el Análisis de Series de Tiempo

2.1	El concepto de Estacionaridad	10
2.2	El concepto de Autocorrelación	11
2.3	Análisis de Series de Tiempo No Estacionarias	12

CAPITULO III El Enfoque Clásico de Series de Tiempo La Metodología de Box y Jenkins

3.1	Consideraciones Generales	13
3.2	Modelos de Promedios-Móviles	14
3.3	Modelos Autorregresivos	15
3.4	Modelos Autorregresivos y de Promedios-Móviles	16
3.5	Modelos ARIMA y Modelos ARIMA Estacionales	17
3.6	Identificación, Estimación y Predicción a través de los modelos ARIMA	18
3.6.1	Identificación	18
3.6.2	Estimación	20
3.6.3	Predicción	21

CAPITULO IV	El Enfoque Bayesiano de Series de Tiempo	
	Modelos ARMA	
4.1	Generalidades	23
4.2	Modelos Autorregresivos	25
4.3	Procesos de Promedios-Móviles	29
4.4	Procesos Autorregresivos y de Promedios-Móviles	29

CAPITULO V	Modelos Lineales Dinámicos	
5.1	Introducción	32
5.2	El Modelo Lineal Dinámico	33
5.3	El Filtro de Kalman	35
5.4	El Filtro de Kalman y El Modelo Lineal Dinámico	41
5.5	Aígunos modelos que pueden formularse como modelos lineales dinámicos	45
5.5.1	El modelo estático de regresión	45
5.5.2	El modelo dinámico de regresión	48
5.5.3	El modelo steady	46
5.5.4	El modelo lineal de crecimiento	47
5.5.5	El modelo autorregresivo de orden p	47
5.5.6	El modelo de promedios-móviles de orden q	48
5.6	Ejemplo	49

CONCLUSIONES

APENDICE

A.1	El Teorema de Bayes
A.2	Definiciones y Teoremas relevantes sobre probabilidades y esperanzas condicionales
A.3	Tres Teoremas Importantes
A.4	El Filtro de Kalman El esquema de Jazwinski. Demostración.
A.5	El Modelo Lineal Dinámico Demostración de las relaciones de Kalman en la versión de Harrison y Stevens.

BIBLIOGRAFIA

PROLOGO

El estudio de series de tiempo resulta de sumo interés en varias disciplinas; diversos autores han propuesto diferentes modelos con el fin de poder explicar su comportamiento. Entre los modelos que se emplean para tal fin, se encuentran los Modelos Lineales Dinámicos, estos modelos se presentan en un artículo de Harrison y Stevens, el cuál se ha empleado como fuente principal de información en la elaboración de este trabajo.

En la formulación de los Modelos Lineales Dinámicos se emplean las relaciones recursivas dadas a través de los Filtros de Kalman, mismas que permiten la actualización de las distribuciones de probabilidad de los parámetros.

La presentación que de estos modelos se hace, se enmarca dentro de un contexto Bayesiano, en el cuál se utilizan probabilidades subjetivas o personales.

En el Capítulo V de esta tesis, se muestra la formulación de estos modelos, su relación con los Filtros de Kalman y se presentan algunos ejemplos de modelos estadísticos que pueden expresarse como Modelos Lineales Dinámicos, cubriéndose de esta manera el objetivo principal de este trabajo.

Adicionalmente se presentan en los Capítulos III y IV algunos estudios, que sobre Modelos ARIMA, han hecho distintos autores dentro de un contexto Clásico y Bayesiano, respectivamente. Se presentan estos modelos con el fin de ilustrar cómo pueden relacionarse éstos con el establecimiento de Modelos Lineales Dinámicos. Asimismo en el Capítulo II se proporcionan algunas definiciones que son relevantes en el estudio de series de tiempo.

Algunas características que presentan las predicciones, así como ciertas consideraciones que deben hacerse cuando se predice o pronostica, se comentan en el primer capítulo de este trabajo. Además en él se muestran algunos modelos que se han utilizado para pronosticar series de tiempo, explicando a grosso modo su funcionamiento.

En la parte final de ésta tesis se presenta un ejemplo de la aplicación de Modelos Lineales Dinámicos a un serie de tiempo específica. También se plantean las conclusiones que obtuvo la autora sobre la misma. Finalmente, aparece un apéndice en el que se han ubicado algunos tópicos que pueden ser de interés para el lector de este trabajo.

EL PROBLEMA DE PREDICCIÓN

CAPITULO I. El Problema de Predicción

1.1 Conceptos Básicos

Debido a que en este trabajo se utilizan con frecuencia algunos términos, creemos necesario dar una idea general de los mismos.

SERIE DE TIEMPO. Una serie de tiempo es una secuencia de observaciones de una variable particular, ordenadas respecto del tiempo. Algunos ejemplos pueden ser:

- La serie formada por los precios semanales de un artículo.
- La serie compuesta por las temperaturas diarias de una región.
- La serie formada por los índices de inflación mensuales del país.
- La serie compuesta por el número de egresados cada año de una cierta licenciatura.
- etc.

PREDECIR. Entenderemos este término en un sentido un tanto distinto a lo que comúnmente se entiende por él. Generalmente la predicción se asocia con el anuncio de cualquier evento futuro. En este trabajo cuando decimos que vamos a predecir el valor de una variable, nos referimos a que vamos a tratar de averiguar el comportamiento de esa variable en un periodo de tiempo específico y vamos a proporcionar un valor (o rango de valores) que probablemente tenía, tiene o tendrá en ese periodo, en base a la *información* que tenemos de la variable, hasta este momento. Como puede observarse, el periodo específico a que nos referimos no necesariamente tendrá que estar ubicado en el futuro.

PREDICCIÓN O PRONOSTICO. Usaremos el término predicción en dos sentidos: como la acción de predecir y como el producto de esa acción. Además cuando nos referimos al producto, utilizaremos los términos predicción y pronóstico en forma indistinta. Cabe mencionar y hacer hincapié en el hecho de que el ejercicio de predicción no sólo se realiza a futuro.

1.2 Introducción

Cuando se cuenta con información acerca de una variable que nos interesa, resulta atractivo tratar de entender el comportamiento que ha venido mostrando ésta; un problema que resulta interesante, es tratar de averiguar el comportamiento de esta variable en otros periodos de tiempo sobre los cuales no tenemos ningún tipo de información. Nos interesa predecir, posiblemente, porque algunas de las decisiones que vayamos a tomar o que tomarán las personas para las cuales hacemos el estudio, estarán fuertemente relacionadas con aquellas condiciones que tratamos de predecir en tales periodos. Estas decisiones pueden ser tan simples como el decidir que tipo de ropa vestiremos mañana -tal decisión se verá influenciada por el clima que oímos ha sido pronosticado-, o tan complejas como administrar los recursos de alguna institución en base a los pronósticos realizados sobre los ingresos y egresos que esa institución tendrá.

1.2.1 La importancia de la predicción

Pronosticar es una actividad muy importante en muchos tipos de organizaciones, ya que las predicciones pueden incorporarse al proceso de toma de decisiones. La importancia de predecir puede verse de varias formas, por ejemplo: un gobierno puede y debe predecir factores de diversa índole o algunas características de estos factores, que sean de vital importancia para el desarrollo de su país, ya que de esa manera podrá formular los lineamientos que va a seguir en adelante. A una institución educativa podría interesarle pronosticar el número de alumnos que ingresarán en un año determinado para poder tomar decisiones concernientes a los recursos que destinará a cada una de sus áreas de estudio.

Además de lo ya mencionado, podemos agregar que, otro de los propósitos al realizar predicciones es obtener un conocimiento previo sobre los eventos inciertos o desconocidos que son importantes para tomar decisiones a corto o largo plazo, de los cuales conoceremos su valor hasta que transcurra cierto tiempo.

Es necesario destacar que no sólo el problema de predicción es relevante como parte de un proceso de toma de decisiones, en este contexto puede resultar más importante juzgar la *calidad* de una predicción, ya que por ejemplo, conseguir "*buenas predicciones*" ocasionará, generalmente, que las decisiones que tomemos a partir de ellas sean más acertadas.

Para poder preparar una predicción —la cual queremos que sea juzgada como "*buena*"—, necesitaremos del estudio y análisis de la información disponible sobre el fenómeno o factor que deseamos predecir y/o de la opinión de un *experto* o *expertos*. Aún cuando en un proceso de predicción, se consideren como "*buenos*" los pronósticos

que hayamos hecho, es necesario recordar que cuando predecimos, estamos acarreado incertidumbre en todo aquéllo en lo que involucremos a tal predicción, puesto que no tenemos el valor real del fenómeno que deseamos pronosticar. Es por esta razón que resulta conveniente destacar algunas de las características inherentes a las predicciones.

1.2.2 Características que presentan las predicciones

Las predicciones (o pronósticos) tienen características que son inherentes a ellas, características que deben tomarse en cuenta cuando los pronósticos vayan a usarse. Las predicciones son incorrectas en su mayoría. Aunque muchas personas reconocen el hecho, existen otras que hacen proyectos con los pronósticos que tienen y no están alertas ante los posibles errores que pueden ocurrir.

Es por ello recomendable actualizar constantemente la información que se tenga sobre el asunto, reemplazando el valor de la observación que se vaya obteniendo por su correspondiente predicción, además de incorporarlo también al volumen de información que se tiene para hacer las subsecuentes predicciones. Las predicciones deben darse en un rango, de modo tal que, junto con la predicción tentativa, vaya asociada la estimación del error que pueda tenerse. Otra característica relevante es la pérdida de exactitud que se tiene a medida que se realizan los pronósticos para periodos más alejados en el horizonte de tiempo. Paralelamente, podemos esperar que el error de predicción se reduzca cuando hagamos predicciones para periodos de tiempo más próximos.

Algunas sugerencias que hacen Sullivan y Claycombe(1977) en su libro *Fundamentals of Forecasting* son las siguientes:

El analista debe conocer su área, es decir, debe tener un conocimiento adecuado de los hechos que ocurren en su campo de trabajo, del medio ambiente que rodea al evento que desea predecir. Una "buena predicción" involucra conocimiento sobre tendencias históricas, factores geográficos, correlación en algunos indicadores económicos (en algunos casos), factores políticos, psicológicos, técnicos o económicos que influyen en aquello que se desea predecir.

Las predicciones se basan en suposiciones. La persona que predice debe conocer y estar preparada para establecer sus suposiciones. Estas deben referirse a factores sobre los cuales no se tiene control o no se está en posibilidad de hacerlo, así como a aquellos factores que sí pueden controlarse.

Al realizar una predicción debe tenerse un propósito en mente, el cual puede establecerse en términos de las preguntas que se deseen responder. Teniendo la información en la mano, el analista debe desarrollar una hipótesis o solución tentativa,

misma que debe probar en el transcurso de su estudio. La información relacionada con la hipótesis debe ser recolectada cuidadosamente, depurándose posteriormente.

Cuando se posee información numérica, es recomendable graficar los datos, a fin de obtener una visión preliminar del comportamiento de los mismos. Mediante este sencillo análisis se puede averiguar si los datos presentan alguna tendencia, si se notan ciclos o bien si existen observaciones "extrañas". Ninguna predicción debe aceptarse como final. Toda predicción debe revisarse constantemente. Cada nuevo dato u observación contribuye a la confiabilidad o desconfianza en las hipótesis usadas.

1.3 Métodos de predicción

Existen metodologías para poder predecir o pronosticar un evento, en particular el valor futuro de una serie de datos. El método o los métodos que se decidan usar dependerán de la aplicación especial o problema específico que se tenga entre manos. Es probable que, por el monto de información disponible, la gama de modelos a los que se pueda recurrir se limite, ya que las distintas metodologías exigirán, implícita o explícitamente, un grado menor o mayor de información. Los métodos de predicción han sido clasificados por sus características; sin embargo, la clasificación no es estándar, por ejemplo, Bowerman y O'Connell(1979; 16-21) clasifican a los métodos de predicción de la siguiente manera:

A. Métodos Cualitativos. Utilizan la opinión de *expertos* para realizar las predicciones. Dentro de estos métodos se encuentran:

Cualitativos { Ajuste Subjetivo de Curvas
Método Delphi o Técnica Delphos
Comparaciones Tecnológicas

B. Métodos Cuantitativos. Estos métodos necesitan de información numérica para poder realizar las predicciones. En este grupo se encuentran:

Cuantitativos { Modelos de Series de Tiempo
Modelos Causales

Para Bowerman y O'Connell, los métodos cualitativos son aquellos métodos que usan la opinión de "expertos" para predecir eventos futuros. Consideran que son convenientes cuando se tiene acceso a las personas calificadas como "expertos" y sobre

todo cuando la información numérica histórica no está disponible o es escasa. Los métodos cuantitativos consisten en el análisis de información histórica a fin de predecir valores futuros de la variable de interés. Nelson(1973; 1-11), por su parte, clasifica a los métodos de predicción en distintas categorías, a saber:

A. Predicción Subjetiva. Uso de la intuición del que predice cuando cree que ésta es mejor que cualquier función matemática de predicción, o bien, cuando requiere de un pronóstico inmediato; en su caso, cuando no puede costear el uso de otra técnica.

B. Modelos Estadísticos y Matemáticos.

i.) Modelos Estructurales y Econométricos. Un modelo estructural es un conjunto de funciones matemáticas que representan relaciones causales entre un conjunto de variables; cuando el modelo incluye un término de perturbación se conoce como modelo estadístico, y cuando se desarrolla en un contexto económico recibe el nombre de econométrico.

ii.) Modelos Determinísticos. Son aquellos modelos que tratan a la variable que se desea predecir, como una función determinística del tiempo.

iii.) Formulas de Predicción Ad-Hoc. Son aquellos modelos matemáticos que se desarrollan de acuerdo a las necesidades particulares de cada problema, sin tratar de justificar *formalmente* la metodología empleada.

iv.) Análisis de Series de Tiempo. Son modelos en los cuales, los valores que toma la variable de interés en los distintos tiempos (o una transformación de los mismos), son considerados como una función de ellos mismos sólo que en periodos de tiempo distintos, más un término aleatorio o de error.

A continuación describiremos algunas de las técnicas de predicción, tratando de unificar las clasificaciones mencionadas anteriormente; cabe aclarar que no pretendemos dar una nueva clasificación de las técnicas.

1.3.1 Modelos Matemáticos y Estadísticos

Cuando se trabaja con información numérica generalmente se hace un análisis del comportamiento que tienen los datos respecto del tiempo o, si se cuenta con la información, respecto de otras variables. Se trata de identificar, por medio de alguna metodología, una estructura que permita determinar los valores futuros de la variable de interés. Esta estructura se trata de establecer en términos matemáticos.

1.3.1.1 Modelos Econométricos

El uso de estos modelos involucra la *identificación* de variables que están relacionadas con la variable que se desea predecir, y que además puedan, en un momento dado, explicarla. Después de que se ha hecho tal identificación, el siguiente paso es establecer un modelo matemático que represente las relaciones existentes entre tales variables; debido a que se trata de identificar una estructura matemática entre las variables, a estos modelos se les designa, en ocasiones, como *estructurales*.

La etapa posterior es *predecir*, considerando los valores que llegarán a tomar las variables explicativas; en ocasiones, por no contar con los valores reales de estas variables, es necesario introducir valores estimados de las mismas, sin embargo, la validez de las predicciones disminuye si no se ha hecho explícita, en el modelo, la suposición de que se consideran valores estimados.

Debido a que en numerosas ocasiones es difícil contar con los valores reales de las variables explicativas, frecuentemente se recurre a ejercicios de simulación, los cuales permiten analizar diferentes escenarios obtenidos por la variación de los valores de las variables explicativas.

Estos modelos son difíciles de formular en la práctica, pues es necesario contar con bastante información, ya que no sólo se necesita tenerla sobre la variable de interés, sino también sobre las variables explicativas. Por esta razón, en múltiples ocasiones, la habilidad para poder predecir a la variable de interés dependerá de la habilidad para predecir a las otras variables.

Una gama importante de estos modelos la constituyen los modelos de regresión lineal; la utilidad de estos modelos se ha constatado en numerosas ocasiones, sin embargo, cuando el objetivo es predecir series de tiempo, resultan insuficientes para este propósito, debido a que las hipótesis de las cuales parten estos modelos no son satisfechas por las series de tiempo (p. ej., el supuesto de no correlación en los errores)¹.

1.3.1.2 Modelos Determinísticos

A los modelos que consideran que la variable de interés o de respuesta es una función determinística del tiempo se les conoce como Modelos Determinísticos. Estos modelos consideran que si la serie de observaciones hechas en el tiempo se denotan por z_1, z_2, \dots, z_n entonces existe una función f tal que $z_t = f(t)$; $t = 1, 2, 3, \dots, n$.

¹Para mayor información sobre estos modelos véase Draper y Smith (1961).

Una elección común para f , es suponer que f es un polinomio en t , esta elección está motivada por el Teorema de Aproximación de Weierstrass, el cual menciona que toda función continua puede aproximarse por un polinomio.

Uno de los problemas que se presentan, es determinar el grado apropiado del polinomio. Sabemos que si se tienen n observaciones, existe un polinomio de grado, a lo más, $n - 1$ que se ajusta exactamente a los puntos dados por esas observaciones. Sin embargo, puede resultar que algunas de las potencias de t sean poco importantes en la evolución de z_t y por otra parte, si nosotros usamos el polinomio resultante para pronosticar más allá del periodo muestral, tendremos errores en la predicción, de los cuales el modelo no nos proporciona información alguna. En forma alternativa podemos optar por un polinomio de grado menor. En una situación así el polinomio no se ajusta exactamente a todos los puntos generados por las observaciones, por lo que se hace necesario introducir un término de perturbación que refleje esta situación, de este modo recurrimos nuevamente a métodos estadísticos para poder resolver el problema.

Otro modelo determinístico usado con frecuencia, es el modelo de crecimiento exponencial $z_t = f(t) = Ae^{\alpha t}$ donde A y α son constantes. Con este modelo los problemas son similares al anterior y aún pueden ser más agudos debido al acelerado crecimiento que presentan; sin embargo, pueden resultar útiles cuando se desea predecir a intervalos de tiempo muy cortos.

1.3.1.3 Modelos ARIMA

Descritos a *grosso modo* los modelos *ARIMA* consideran que cada observación de la serie de tiempo² es una combinación lineal de *observaciones* y perturbaciones aleatorias ocurridas en periodos distintos, así como de perturbaciones aleatorias tenidas en el periodo presente. En términos matemáticos, un proceso *ARIMA* tiene la forma:

$$z_t = \mu + \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}$$

donde $\{u_t \mid t \in Z\}$ es una sucesión de variables aleatorias independientes con media cero y varianza común σ^2 .

En este trabajo se consideran dos formas de conseguir predicciones a través de estos modelos, las cuales dependen del enfoque de probabilidad que se esté empleando: un Enfoque *Clásico* o un Enfoque *Bayesiano*.

² Esa observación puede ser producto de diferencias sucesivas aplicadas a valores consecutivos de la serie de tiempo en cuestión.

1.3.1.4 Modelos Lineales Dinámicos

Estos modelos se emplean en la predicción de series de tiempo y se definen a través de las relaciones de recurrencia de los *filtros de Kalman*. Se estudian con detalle en el último capítulo de esta tesis. Se ubican dentro del Enfoque Bayesiano.

1.3.2 Otros métodos de predicción

Algunas personas que desean predecir una situación futura, en ciertas ocasiones, por causas que enunciaremos más adelante, recurren a metodologías que utilizan la opinión de *expertos* para poder predecir algún evento futuro. Estos métodos son adecuados en situaciones en las cuales se pueda contar con la ayuda de los "*expertos*", pues en caso contrario la aplicación del método no tendría sentido. Algunas de las razones por las cuales se utilizan estos, son las siguientes:

— Cuando el evento que se intenta predecir no es de índole numérica. Por ejemplo, predecir la política que seguirá un nuevo gobierno, o bien, la forma de vida de sociedades futuras y otros problemas similares.

— Cuando la persona que predice o pronostica considera que recurrir a la opinión de personas calificadas tiene mayor validez que alguna técnica matemática conocida por él, debido a las características de su problema.

— Cuando la información histórica cuantificable que requieren los métodos de predicción matemáticos, es nula, insuficiente o poco confiable; o bien, cuando alguno de los supuestos o hipótesis en los cuales se fundamentan estos métodos no se satisfacen y no se conoce alguna forma justificable para proseguir con el uso de ellos.

— Cuando la persona que predice puede contar con la opinión de *expertos* y desea aplicar la metodología, a fin de complementar las predicciones obtenidas por medio de otras técnicas.

Debido a que no existe un patrón para calificar a una persona como "*experto*", algunos estudios pueden perder su validez —aún cuando el método empleado haya sido aplicado en forma correcta—, a causa de que las personas que se calificaron como *expertas* no sean reconocidas como tales. Este es uno de los principales problemas con los que se enfrentan estos métodos. Un ejemplo de ellos es el método que se conoce como *Método Delphi* o *Técnica Delfos*, el cual consideraremos en los siguientes párrafos³.

³Otros métodos son considerados por Bowerman y O'Connell (1979; 16-20) y por Sullivan y Claycombe (1977; 139-190).

El Método Delphi. En muchas ocasiones, la situación sobre la cual se va a decidir es considerablemente compleja y poco entendida, de modo que no puede esperarse que una sola persona pueda tomar una decisión atinada. Esta es una de las premisas de las cuales parte el método, es decir, se considera que el conocimiento combinado de varias personas es al menos tan bueno como el de una persona. La otra premisa es que las personas que tienen un gran conocimiento sobre una área particular hacen las predicciones más creíbles.

Un acercamiento tradicional que habían tenido los métodos que usan la consulta a *expertos*, era tomar la opinión de éstos a través de un panel de discusiones; sin embargo, los resultados obtenidos de este modo eran, generalmente, poco satisfactorios debido a que la opinión de los *expertos* estaba distorsionada por la influencia de individuos que dominaban el panel y/o porque las opiniones mayoritarias intimidaban a algunos de los participantes.

El Método Delphi intenta superar esas dificultades, convenciendo a los involucrados en el ejercicio de predicción, para que expresen sus opiniones en forma anónima a través de un intermediario.

El intermediario actúa como centro de control de análisis de respuestas de cada ronda de opiniones, reuniendo y retroalimentando la opinión de los participantes en rondas subsiguientes.

Por lo tanto, la técnica Delphi es un procedimiento para solicitar y organizar los pronósticos de *expertos* sobre el evento de interés, por medio del uso de respuestas anónimas e iterativas a una serie de cuestionarios y de una retroalimentación controlada de las opiniones del grupo. Siguiendo este procedimiento, se espera que las respuestas converjan a un pronóstico de consenso que de un buen estimador del resultado real, o que permita la definición de ciertas posturas por parte de los *expertos*⁴.

⁴Para mayor información sobre el empleo de esta técnica, puede consultarse Sullivan y Claycomb(1977; 140-150).

**CONCEPTOS FUNDAMENTALES EN EL
ANALISIS DE SERIES DE TIEMPO**

Este supuesto nos lleva a que la función de distribución marginal para cualquier par de *observaciones* es la misma, lo mismo sucede para la esperanza y la varianza. Si consideramos la covarianza entre ellas resultará, bajo este supuesto, que la covarianza entre cada par de observaciones depende única y exclusivamente del número de periodos que las separa.

El supuesto a que hemos hecho mención es el supuesto de *estacionaridad estricta o fuerte*. Existe un supuesto menos exigente que el anterior y es conocido como *estacionaridad débil*, donde lo que se supone es que:

$$E\{z_t\} = \mu$$

y

$$\text{Cov}\{z_t, z_{t+k}\} = \text{Cov}\{z_{t+m}, z_{t+m+k}\}$$

donde t es cualquier punto en el tiempo y k y m son una pareja de enteros.

En algunas ocasiones se utiliza el término *autocovarianza* para designar a la covarianza entre cada par de observaciones, debido a que son observaciones de la misma serie.

2.2 El concepto de Autocorrelación

Un concepto que puede resultar interesante y que se deriva del concepto de autocovarianza es el de *autocorrelación*.

Hemos mencionado que la covarianza entre dos observaciones en una serie de tiempo, depende solamente del número de periodos que las separa, cuando se tiene la propiedad de estacionaridad. De este modo, al contar con tal propiedad, si conocemos el comportamiento que tiene la serie en un lapso, podremos *inferir* el comportamiento de la misma en otro periodo de tiempo, pues aquél será semejante.

Con el fin de estandarizar los valores de las covarianzas, consideramos las correlaciones entre las observaciones, las cuales reciben el nombre de *autocorrelaciones*, por referirse a observaciones de la misma serie.

Las autocorrelaciones nos auxiliarán en la identificación de aquellas observaciones que son de mayor influencia en una observación dada, pues podemos pensar que aquéllas que muestran una correlación en valor absoluto *más grande* que otras observaciones con distinto periodo de retraso, influirán *más* en el valor de la variable que nos interesa.

No obstante, los valores de las autocorrelaciones, así como otras estadísticas debemos estimarlas debido a que no conocemos a la función de distribución conjunta que explica a las observaciones³.

2.3 Análisis de Series de Tiempo No Estacionarias

Es frecuente que nos encontremos con series de tiempo para las cuales es evidente que el supuesto de estacionaridad carece de validez (p. ej. cuando los valores de la serie de tiempo presentan alguna tendencia). En estos casos, lo que se trata de hacer es transformar los datos de modo tal que la serie resultante pueda suponerse estacionaria⁴.

Generalmente, lo que se hace es considerar la serie compuesta por las diferencias entre valores consecutivos de la serie original, i.e. si z_0, z_1, \dots, z_T son los valores de la serie original entonces la nueva serie resultará ser x_1, x_2, \dots, x_T donde

$$x_i = z_i - z_{i-1}; \quad \text{con } i = 1, 2, \dots, T$$

Si la serie resultante x_1, x_2, \dots, x_T no pudiera suponerse estacionaria, existen otras transformaciones que pueden intentarse con el propósito de poder suponer estacionaridad⁵.

En el siguiente capítulo nos daremos cuenta de la necesidad de poder suponer estacionaria a la serie de tiempo con la que estemos trabajando, ya que la metodología tratada en ese capítulo está orientada a la predicción de series de tiempo con esas características.

Existen otras técnicas estadísticas, como las que se comentan en el Capítulo IV y V, para las que no es necesario suponer estacionaria la serie⁶.

³Una discusión más extensa así como algunos ejemplos pueden encontrarse en Nelson(1973; 23-27 y 70-75) y Bowerman y O'Connell(1979; 341-366).

⁴Como veremos más adelante, existen técnicas para las cuales no es indispensable suponer estacionaridad.

⁵Véase Nelson(1973; 56-59) y Box y Jenkins(1976; 85-87).

⁶La condición de estacionaridad se puede agregar al modelo en el caso que se sepa que la serie tiene esa característica, lo cual significaría que se tiene ese conocimiento "extra" sobre la serie.

EL ENFOQUE CLASICO DE SERIES DE TIEMPO

LA METODOLOGIA DE BOX Y JENKINS

Capítulo III. El Enfoque Clásico de Series de Tiempo

La Metodología de Box y Jenkins

3.1 Consideraciones Generales

Hemos mencionado en los capítulos anteriores que los valores que conforman a una serie de tiempo, están correlacionados entre sí, debido a la dependencia que tienen respecto del tiempo. Esta propiedad se utiliza en la Metodología de Box y Jenkins, ya que en ella se analizan cuáles valores están más correlacionados entre sí, tomando en cuenta el periodo de tiempo que los separa, para poder decidir que valores se usarán en la predicción.

Además, debemos enfatizar que en muchas técnicas de predicción como la que describiremos en este capítulo, la persona que va a predecir, analiza la información que tiene entre manos, con el fin de *identificar* un *patrón* que pueda usar para describir esa misma información. Después extiende ese patrón al periodo que le interesa predecir para llevar a cabo su pronóstico. Esta estrategia se usa frecuentemente y descansa en una suposición básica: *"El patrón que se ha identificado es el mismo tanto para el periodo muestral como para el periodo en el cual se quieren realizar las predicciones"*. Es claro que, si se tiene evidencia en contra de la aseveración anterior, no tenemos porque esperar que la técnica de predicción empleada, nos arroje *buenas predicciones*.

Los modelos que serán tratados en este capítulo pertenecen a una clase más general, la clase de los *procesos estocásticos lineales discretos*; estos procesos pueden expresarse en la forma siguiente:

$$z_t = \mu + u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \dots$$

donde μ y $\psi_i, i = 1, 2, \dots$ son parámetros fijos y la serie de tiempo $(\dots, u_{t-2}, u_{t-1}, u_t, \dots)$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza común desconocida σ^2 . Es común que a esta sucesión de variables aleatorias se les conozca como *ruido blanco*. El proceso es discreto debido a que las observaciones $\{z_t \mid t \in Z\}$ se toman en intervalos de tiempo discretos e igualmente espaciados; es lineal, porque es una combinación lineal de los parámetros.

La condición de estacionaridad (débil) para este tipo de modelos se satisface si las siguientes sumas convergen:

$$\sum_{i=0}^{\infty} \psi_i, \quad \sum_{i=0}^{\infty} \psi_i^2, \quad \sum_{i \neq j} \psi_i \psi_j \quad \text{donde} \quad \psi_0 = 1$$

Debido a que el número de parámetros involucrados en este tipo de modelos es *extremadamente grande*, resulta de mayor utilidad contar con modelos más sencillos que involucren pocos parámetros, sobre todo porque el manejo de los mismos se facilita. En la construcción de modelos, a este principio se le conoce como *parsimonia*. En las siguientes secciones trataremos con este tipo de modelos.

3.2 Modelos de Promedios-Móviles

Un proceso de promedios-móviles de orden q tiene la forma:

$$z_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$$

donde μ y θ_i ; $i = 1, 2, \dots, q$ son parámetros fijos y las variables aleatorias $\{u_t \mid t \in Z\}$ son independientes e idénticamente distribuidas con media cero y varianza común σ^2 ; es claro que se trata de un proceso lineal discreto si se considera que:

$$\psi_0 = 1 \quad ; \quad \psi_i = -\theta_i \quad , \quad i = 1, 2, \dots, q \quad \text{y} \quad \psi_j = 0 \quad \text{si} \quad j > q$$

Una convención que se ha hecho en la literatura sobre series de tiempo es denotar a los procesos de promedios-móviles cuyo orden es q como $MA(q)$. Una propiedad de los $MA(q)$ que resulta interesante es que dada una observación cualquiera, digamos z_t , las observaciones que influirán en el valor que tome z_t serán aquellas que se encuentren q periodos antes o después, i.e. en el valor que tome z_t influirán únicamente z_{t-q}, \dots, z_{t-1} o bien z_{t+1}, \dots, z_{t+q} . Esta situación puede verse de manera más formal a través de la autocorrelación existente entre dos observaciones dadas; después de algunos cálculos pueden verificarse las siguientes expresiones:

$$E[z_t] = \mu$$

$$\gamma_0 = \text{Var}[z_t] = \sigma^2 \sum_{i=0}^q \theta_i^2 \quad \text{donde} \quad \theta_0 = -1,$$

$$\gamma_j = \text{Cov}[z_t, z_{t-j}] = \begin{cases} \sigma^2(-\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q), & \text{si } j = 1, 2, \dots, q; \\ 0 & \text{cuando } j > q \end{cases}$$

$$\rho_j = \text{Corr}[z_t, z_{t-j}] = \begin{cases} \frac{-\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{si } j = 1, 2, \dots, q, \\ 0 & \text{cuando } j > q. \end{cases}$$

3.3 Modelos Autorregresivos

Otros modelos importantes dentro de la teoría de series de tiempo son los modelos autorregresivos, mismos que presentan la siguiente forma:

$$z_t = \eta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + u_t$$

donde $\{u_t \mid t \in Z\}$ es una colección de variables aleatorias idénticamente distribuidas con media cero y varianza común σ^2 .

La razón por la cual estos modelos reciben tal nombre es por su similitud con la ecuación de regresión. El prefijo *auto* se agrega porque los valores de la serie están correlacionados con valores de la misma serie, sólo que en periodos diferentes. Estos modelos pueden verse como una forma reducida del modelo:

$$z_t = \eta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + u_t$$

el cual es producto de una reexpresión del modelo lineal general¹:

$$z_t = \mu + u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \dots$$

al que ya hemos hecho mención. El orden de los modelos autorregresivos, a semejanza de los modelos de promedios-móviles está determinado por el *retraso mayor* sufrido por las variables involucradas, en este caso el orden del modelo es p . Usualmente estos modelos son denotados por $AR(p)$.

Para este tipo de modelos ya no resulta inmediata la forma en la cual se puede *chequear* estacionaridad, sin embargo, esta condición se satisface si las raíces de la *ecuación característica*

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$$

se encuentran fuera del círculo unitario². B es una variable *muda*. Puede probarse que estos modelos satisfacen lo siguiente:

$$E[z_t] = \frac{\eta}{1 - \phi_1 - \dots - \phi_p}$$

además:

$$\gamma_0 = \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma^2$$

$$\gamma_1 = \phi_1 \gamma_0 + \dots + \phi_p \gamma_{p-1}$$

$$\vdots$$

$$\gamma_p = \phi_1 \gamma_{p-1} + \dots + \phi_p \gamma_0.$$

¹Véase Nelson (1973; 37-38).

²Véase Box y Jenkins (1970; 83, 85 y 80-82).

En este caso, es claro que dados los parámetros ϕ_1, \dots, ϕ_p y σ^2 se puede resolver el sistema de ecuaciones y obtener los valores de $\gamma_0, \dots, \gamma_p$, sin embargo, es frecuente que se desconozca el valor de tales parámetros.

Para periodos mayores a p , las covarianzas pueden calcularse recursivamente, a través de la ecuación:

$$\gamma_j = \phi_1 \gamma_{j-1} + \dots + \phi_p \gamma_{j-p}; \quad j > p.$$

Si consideramos las ecuaciones anteriores y las dividimos entre γ_0 , obtenemos el sistema de ecuaciones conocido como ecuaciones de Yule - Walker:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \rho_2 + \dots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p. \end{aligned}$$

3.4 Modelos Autorregresivos y de Promedios-Móviles

Esta clase de modelos resulta ser una mezcla de los modelos autorregresivo y de promedios-móviles. Presentan la siguiente forma:

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$$

los cuales, por abreviación, reciben el nombre de modelos $ARMA(p, q)$; ésta notación, además, resulta mnemotécnica pues recuerda que el grado de la parte autorregresiva es p y el grado de la componente de promedios-móviles es q .

Del mismo modo que para un proceso $AR(p)$, la condición de estacionaridad se satisface si las raíces de la ecuación característica

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$$

se encuentran fuera del círculo unitario.

Un equivalente algebraico del concepto de estacionaridad es el concepto de *invertibilidad*, el cual se define como la capacidad para convertir un modelo $MA(q)$ en una expresión de la forma:

$$z_t = \eta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + u_t.$$

Para poder *invertir* un modelo $MA(q)$ es necesario que las raíces de la ecuación

$$1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0$$

se localicen fuera del círculo unitario. Cabe destacar que aún cuando la propiedad de *invertibilidad* y de *estacionaridad* son algebraicamente muy similares, conceptualmente no lo son, ya que, *p. ej.*, cualquier proceso de promedios-móviles es estacionario pero no necesariamente es invertible y todo proceso autorregresivo es invertible pero no necesariamente estacionario.

3.5 Modelos ARIMA y Modelos ARIMA Estacionales

Se mencionó en el Capítulo II que en algunas series de tiempo, es evidente que no puede suponerse estacionaridad. Se comentó también que, con el fin de suponer estacionaridad se recurre con frecuencia a la transformación de la serie que da por resultado las diferencias entre valores sucesivos de la serie inicial. Si la serie de tiempo que deseamos modelar ha sufrido alguna transformación de esta índole, entonces esa transformación debe considerarse en el modelo. De este modo si se obtiene una serie de tiempo, producto de haber calculado d diferencias a la serie original, los modelos que se consideran en este caso son los siguientes: $ARI(p, d)$, $IMA(d, q)$ y $ARIMA(p, d, q)$ ³. Lo cual implica que a la serie transformada se le ajustará un modelo $AR(p)$, un $MA(q)$ o un $ARMA(p, q)$ respectivamente.

Al considerar algunas series como: venta mensual de juguetes en el D.F., o bien, accidentes automovilísticos semanales en carretera, podríamos pensar, en el primer caso, que la venta de juguetes aumenta en los meses de diciembre y enero, y que este fenómeno se repite año con año; por otra parte, los accidentes en carretera en un país como México, resultan ser más frecuentes en periodos vacacionales, como lo es la semana santa. En la literatura sobre series de tiempo, estas situaciones *periódicas* reciben el nombre de *componente estacional* o *estacionalidad* de la serie de tiempo⁴. Los modelos *ARIMA Estacionales* han incorporado este concepto, sólo que de manera determinística, pues, se considera en ellos que el periodo estacional es fijo, *p. ej.*, se supone que las alzas en ventas se dan *exactamente* en el mismo periodo para todos los años, o bien, en el otro ejemplo, que los accidentes automovilísticos aumentan *exactamente* cada x meses, etc. La expresión de estos modelos es como sigue:

$$w_t = \Delta + \Phi_1 w_{t-s} + \dots + \Phi_p w_{t-s}^p + u_t - \Theta_1 u_{t-s} - \dots - \Theta_q u_{t-s}^q.$$

Podemos observar que en estos modelos, las interacciones entre las observaciones son periodo a periodo, *i.e.*, si las observaciones que tenemos son mensuales y suponemos que el periodo estacional es de un año, las observaciones de un mes x se relacionarán *solamente* con las observaciones de los meses x pero no se considera la correlación entre meses de un mismo periodo estacional. Esta situación ha motivado que se consideren

³ Un desarrollo más amplio puede verse en Nelson (1973; 53-63) y en Box y Jenkins (1970; 85-125).

⁴ No debe confundirse el término con *estacionaridad*.

otros modelos que superen estos inconvenientes⁵. Estos modelos no se considerarán en este trabajo.

3.6 Identificación, Estimación y Predicción a través de los modelos ARIMA

En esta sección nos limitaremos a mostrar algunas de las ideas que Nelson (1973) da para la solución de los problemas de identificación, estimación y predicción. No profundizamos en el estudio de ellos debido a que no es el objetivo principal de esta tesis; sin embargo, es conveniente mencionar que este es un campo en el cual se requiere de una mayor investigación, ya que no resulta del todo claro, al menos dentro del planteamiento de Nelson, cómo y de qué forma se realiza la identificación del modelo, debido a que aplicando la metodología que él propone, la identificación puede llegar a dificultarse y ser confusa aún cuando se trate de series de tiempo simuladas, las cuales *efectivamente* son producto de un proceso como los que se han visto en las secciones anteriores.

3.6.1 Identificación

Nelson considera el problema de identificación de modelos $ARIMA(p, d, q)$. Para abordar el problema, trata por separado la identificación de modelos de promedios-móviles, modelos autorregresivos y modelos mixtos (autorregresivos y de promedios-móviles).

Para poder continuar, necesitaremos introducir algunos conceptos como lo son la *función de autocorrelación* y el *correlograma*. La primera no es una función propiamente, pues, se refiere al conjunto formado por las autocorrelaciones ρ_0, ρ_1, ρ_2 , etc. El correlograma no es más que la gráfica de la función de autocorrelación, es decir, a cada valor $0, 1, 2, \dots$ se le asocia el valor correspondiente $\rho_0, \rho_1, \rho_2, \dots$.

Debido a que en la mayoría de las ocasiones es difícil contar con los valores de $\rho_0, \rho_1, \rho_2, \dots$, los valores que se grafican son los estimadores de estos valores, mismos que están dados por r_0, r_1, r_2, \dots donde

$$r_j = \frac{c_j}{c_0} \quad j = 0, 1, 2, \dots$$

con

$$c_j = \frac{1}{T} \sum_{t=1}^{T-j} (z_t - \bar{z})(z_{t+j} - \bar{z}) \quad \text{y} \quad \bar{z} = \frac{1}{T} \sum_{t=1}^T z_t.$$

⁵Nelson (1973); El Modelo General Estacional Multiplicativo.

El correlograma que se obtiene al graficar los valores estimados, r_0, r_1, r_2, \dots , recibe el nombre de *correlograma muestral*.

Hemos mencionado que cuando la serie en cuestión no se considera estacionaria tenemos como alternativa transformarla mediante diferencias. Sin embargo, una pregunta que surge es: ¿Cuál es el grado apropiado de diferencias que debe obtenerse con el fin de que la serie resultante pueda considerarse estacionaria?. Nelson sugiere que no se calcule ninguna diferencia mientras no se cuente con suficiente evidencia que haga suponer que la serie no es estacionaria. En el caso de que sí se cuente con esa evidencia, él considera que en la mayoría de los casos es suficiente con calcular la primera diferencia; si no es así será necesario analizar las segundas diferencias. En ambos casos, Nelson emplea el correlograma de la serie para decidir si puede considerarla estacionaria o no.

Es importante notar la conveniencia de calcular un grado bajo de diferencias, solamente, pues de lo contrario, perderemos claridad en el ajuste que pretendamos hacer. De no lograr estacionaridad con las primeras o segundas diferencias es recomendable recurrir a otro tipo de transformaciones que tengan la característica de poder *desbaratarse*, ya que de no ser así, tendremos problemas al tratar de predecir.

Para tratar de identificar el grado apropiado de un modelo de promedios-móviles, Nelson recurre al correlograma muestral que ya hemos mencionado, mismo que emplea para distinguir algunos valores tentativos para el grado del modelo⁶, q .

Además él propone una prueba para poder decidir si a partir de un valor dado q^* , los valores estimados de las autocorrelaciones empiezan a ser *suficientemente* parecidos a cero, de modo que pueda suponer que el valor q^* propuesto es el grado de la parte de promedios-móviles.

Por otra parte, para poder identificar el grado apropiado de un modelo autorregresivo, Nelson introduce un nuevo concepto: las *autocorrelaciones parciales*, las cuales se obtienen a partir de las ecuaciones de *Yule - Walker*⁷.

El procedimiento que emplea es el siguiente:

i.) Supone que el grado de la parte autorregresiva es 1, de este modo puede estimar el valor de ϕ_1 .

ii.) Prueba si el valor de ϕ_1 es estadísticamente diferente de cero, para asegurarse que tiene un proceso autorregresivo⁸.

⁶ Véase Nelson (1973; 72-75) para mayores detalles.

⁷ Véase sección 3.3.

⁸ Nótese que puede tratarse de un proceso autorregresivo aún cuando el valor de ϕ_1 sea cero.

iii.) Para ver si el proceso es de orden $p^* = 2$ o más grande, resuelve las ecuaciones de Yule-Walker suponiendo que el grado del modelo es p^* . Si el estimador que obtiene para ϕ_{p^*} es diferente de cero continúa el procedimiento suponiendo ahora que el grado del modelo es $p^* + 1$, en caso contrario, concluye que el grado es $p^* - 1$.

Cuando se trata de identificar modelos mixtos, la situación es todavía más compleja. En este caso Nelson propone analizar las autocorrelaciones y las autocorrelaciones parciales a partir de los correlogramas. La descripción que él hace del procedimiento resulta poco informativa, además queda de manifiesto que la identificación de los modelos sólo se logra en base a la experiencia.

El procedimiento que propone Nelson no es suficiente para identificar modelos ARIMA, por ello resulta necesario hacer una investigación más profunda sobre el tema.

3.6.2 Estimación

Estimar el valor de los parámetros es otro problema importante dentro de Series de Tiempo. La estimación de los parámetros de los modelos ARIMA es otro de los problemas que intenta resolver Nelson; aquí trataremos de mostrar lo que él hace al respecto. Adelantaremos que el problema no queda mucho mejor solucionado que el de identificación.

Uno de los métodos que Nelson reporta está basado en la sustitución de los valores de las autocorrelaciones por sus respectivas estimaciones, en las ecuaciones que relacionan a las autocorrelaciones con los parámetros. Esto no parece ser muy complicado cuando se sustituyen éstos en las ecuaciones de Yule-Walker, en un modelo autorregresivo, pues se tienen p ecuaciones con p incógnitas; sin embargo, cuando se trata de modelos de promedios-móviles, las ecuaciones resultantes no son lineales, esto hace que los cálculos resulten difíciles, ya que, p. ej., cuando el grado es 1, la ecuación es cuadrática, en cuyo caso se tienen dos soluciones. Nelson propone que en tal caso debemos quedarnos con aquella solución que satisfaga la condición de invertibilidad, pero, ¿Qué pasa si las soluciones son complejas?. Por lo anterior, consideramos que el método que él propone no resuelve el problema.

Nelson también comenta que la estimación puede hacerse mediante el cálculo de los estimadores de máxima-verosimilitud, es decir, los valores para $\phi_1, \dots, \phi_p, \eta, \theta_1, \dots, \theta_q$ que minimizan la siguiente expresión:

$$\sum_{t=1}^T (z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} - \eta + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q})^2$$

El cálculo de tales valores puede facilitarse mediante aproximación numérica, sin embargo, debe notarse que en ese caso sólo tendremos los valores tentativos para los

estimadores, pero no podemos atribuirles ninguna de las propiedades de los estimadores de máxima-verosimilitud. Nelson justifica el uso del método de máxima-verosimilitud suponiendo que el número de observaciones con que se cuenta generalmente es *muy grande*.

Debe apuntarse aquí que si el grado de alguna de las componentes autorregresivas o de promedios-móviles no fué adecuadamente identificado, ese error será acarreado cuando se calculen los estimadores de los parámetros pues no se considera en su cálculo que los grados del modelo sean estimados.

3.6.3 Predicción

En el Capítulo I mencionamos algunas de las consideraciones que deben tenerse en mente cuando se trabaja con predicciones. Cuando se modela⁹, se trata de encontrar el proceso que *genera* la información que estamos analizando, sin embargo, el modelo que obtenemos resulta ser una aproximación de aquél; de ahí que el éxito en este paso dependerá de la habilidad que tengamos en la identificación y estimación de los parámetros del modelo.

Si suponemos que el modelo que hemos identificado es el correcto, se puede demostrar que la esperanza de la observación que deseamos predecir condicionada sobre las observaciones de que disponemos nos arrojará la predicción con la menor suma de cuadrados del error, es decir, si definimos a:

$$\hat{z}_t(l) = E\{z_{t+l} \mid z_t, z_{t-1}, z_{t-2}, \dots\}$$

entonces

$$E\{(z_{t+l} - \hat{z}_t(l))^2 \mid z_t, z_{t-1}, \dots\} < E\{(z_{t+l} - P)^2 \mid z_t, z_{t-1}, \dots\}$$

donde P es cualquier predicción para z_{t+l} . Es por ello que se utiliza a la esperanza condicional como función de predicción. Así si

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$$

entonces

$$\hat{z}_t(l) = \begin{cases} \delta + \phi_1 \hat{z}_t(l-1) + \dots + \phi_p \hat{z}_t(l-p) - \theta_1 u_t - \dots - \theta_q u_{t-q+l} & \text{si } l \leq q \\ \delta + \phi_1 \hat{z}_t(l-1) + \dots + \phi_p \hat{z}_t(l-p) & \text{en otro caso} \end{cases}$$

donde $\hat{z}_t(l-i) = z_{t+l-i}$ si $i \geq l$.

⁹Dentro del enfoque clásico.

En este caso debe notarse que para poder realizar la predicción debemos contar con los valores de $u_t, u_{t-1}, \dots, u_{t-g+l}$ mismos que por su naturaleza no son observables. No obstante lo anterior, el obstáculo puede salvarse si notamos que

$$z_{t-i} - \hat{z}_{t-i-1} = u_{t-i} \quad i = 0, 1, 2, 3, \dots$$

y hacemos la sustitución correspondiente.

Algunos ejemplos de la aplicación de modelos *ARIMA* en la predicción de series de tiempo pueden encontrarse en Nelson(1973; 147-157).

**EL ENFOQUE BAYESIANO DE
SERIES DE TIEMPO**

MODELOS ARMA

CAPITULO IV. El Enfoque Bayesiano de Serles de Tiempo

Modelos ARMA

El objetivo del presente capítulo es describir algunos modelos para series de tiempo que se han desarrollado bajo el *Enfoque Bayesiano* o *Corriente Bayesiana* de probabilidad. En forma breve trataremos de resaltar las principales diferencias que existen entre los tratamientos *Bayesiano* y *Clásico*¹ para series de tiempo. No discutiremos respecto de las diferencias conceptuales e ideológicas de cada corriente, debido a que escapa de los límites de este trabajo.

Los modelos que se presentan en esta sección, matemáticamente, tienen la misma estructura que los modelos propuestos por Box y Jenkins que aparecen en el capítulo anterior; sin embargo existen diferencias de fondo entre ambos enfoques, por ejemplo, la forma en la cual se ataca el problema de *estimación* de los valores de los parámetros es sustancialmente diferente en el Enfoque Bayesiano puesto que los parámetros, en general, por el conocimiento que sobre ellos tiene el interesado, son considerados variables aleatorias y como tales tienen asociado un espacio de probabilidad, mientras que en el Enfoque Clásico se suponen fijos en todo momento y se estiman a través de los valores de las observaciones.

A continuación presentamos una descripción de algunos conceptos que se manejan dentro de la Corriente Bayesiana y que se utilizarán en lo sucesivo.

4.1 Generalidades

Dentro de la *Corriente Bayesiana* se argumenta que la aleatoriedad de un fenómeno está dada por el conocimiento que se tiene de él, es decir, un fenómeno se considera aleatorio cuando su comportamiento no se conoce en forma *casi determinística*; cuando esto sucede, el fenómeno trata de explicarse a través de un modelo probabilístico, mismo que se propone en base a lo que se conoce del fenómeno; de este modo se hace una asignación inicial de probabilidades, construyéndose así la *distribución inicial* de los parámetros del modelo en base al conocimiento *a priori* que se tiene sobre ellos.

¹Corriente Frecuentista de Probabilidad.

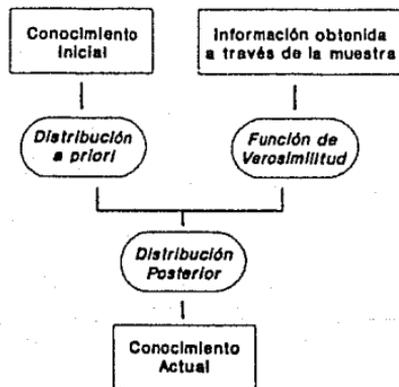
Cuando el conocimiento sobre el fenómeno aumenta², esa asignación inicial de probabilidades cambia o se refuerza, por ello dentro de este enfoque, las probabilidades son *dinámicas* y *personales*, puesto que cambiarán de acuerdo al nuevo conocimiento que adquiera el interesado³.

La actualización de las probabilidades se realiza mediante el uso del Teorema de Bayes⁴. Este teorema se usa con bastante frecuencia en Estadística Bayesiana y es, básicamente, la forma en la cual se incorpora la nueva información sobre el fenómeno.

El mecanismo de actualización que provee el Teorema de Bayes está dado a través del siguiente diagrama:

El conocimiento inicial que el interesado tiene sobre el fenómeno aunado a la información proveniente de la muestra, conforman el conocimiento actual que el interesado tiene sobre el fenómeno⁵.

$$p(\theta | X) \propto p(\theta)p(X | \theta)$$



² El conocimiento sobre un fenómeno aumenta por diferentes causas, tanto indirectas como directas, por ejemplo, la lectura de un libro, la asistencia a una conferencia, o bien, la obtención de una muestra y su análisis contribuyen al aumento en el conocimiento sobre un fenómeno dado.

³ Llamaremos *interesado* a la persona o grupo de personas que desean estudiar al fenómeno y que conjuntamente con el estadístico realizarán el estudio y modelaje del mismo, pudiendo ser el propio estadístico el interesado.

⁴ Consulte el anexo A.1 para mayor información sobre su estructura matemática.

⁵ $p(X | \theta)$ recibe el nombre de *verosimilitud* debido a que vista como función de θ , resulta ser la función de verosimilitud de la Estadística Clásica.

De esta forma si la distribución inicial para θ es $p(\theta)$ y se obtiene un primer conjunto de datos X_1 con función de probabilidad $p(X_1 | \theta)$, entonces la distribución posterior para θ dada esa nueva información será:

$$p(\theta | X_1) \propto p(\theta)p(X_1 | \theta)$$

Si se obtiene un segundo conjunto de datos independiente de X_1 , digamos X_2 , con función de probabilidad $p(X_2 | \theta)$, podemos encontrar la función de distribución posterior para θ en base a la información que proporcionan esos dos conjuntos de datos. Haciendo uso del teorema de Bayes se obtiene que:

$$p(\theta | X_1, X_2) \propto p(\theta | X_1)p(X_2 | \theta)$$

o bien

$$p(\theta | X_1, X_2) \propto p(\theta)p(X_1 | \theta)p(X_2 | \theta)$$

mediante esta última expresión podemos notar cómo la nueva información se va incorporando a la distribución inicial de probabilidad.

En varias ocasiones la información inicial que el interesado tiene sobre el fenómeno es poco precisa debido a que ésta se ha combinado con la información proveniente de la muestra, por lo cual es muy difícil distinguir cuál era el conocimiento del interesado antes y después de haber visto la muestra; en otras ocasiones la información inicial del interesado llega a ser vaga, en ambos casos es necesario introducir una *distribución inicial de referencia*⁶, la cual se utilizará para iniciar el proceso. Otra razón por la cual pueden introducirse distribuciones no informativas, se debe a que no se desea mezclar el volumen de información que el interesado tiene sobre el fenómeno con la información proveniente de las observaciones. La distribución que se escoja como inicial no informativa, al combinarse con la *verosimilitud*, deberá producir una distribución posterior que sólo reflejará la información que proviene de la muestra. De este modo, la distribución inicial tendrá como único objetivo *arrancar* el proceso que permitirá la actualización de las probabilidades, pero no deberá influir en la distribución posterior que se obtenga.

En la siguiente sección se describen algunos desarrollos Bayesianos de modelos ARMA que han realizado Zellner, Broemeling y Alegría.

4.2 Modelos Autorregresivos

Zellner(1971; 186-200) hace un estudio de los procesos autorregresivos normales de primer orden, los cuales tienen la forma

$$z_t = \eta + \phi_1 z_{t-1} + u_t ; \quad t = 1, 2, \dots, T$$

⁶Llamada también no informativa o difusa.

como puede observarse, estos modelos tienen la misma estructura que los que se vieron en el Capítulo II, en este caso los supuestos que involucra el modelo son los siguientes: η y ϕ_1 son parámetros reales desconocidos, $\{u_t\}$ son términos de perturbación o ruido explicados a través de una variable normal, independientes con media cero y varianza σ^2 . Zellner hace la aclaración de que el subíndice t se refiere, bajo este contexto, al t -ésimo valor de una variable. Nótese que en este caso el dominio de t es finito, ya que sólo se hace referencia a la información disponible hasta el periodo T .

Suponiendo que z_0 es la observación al tiempo cero y que la distribución inicial conjunta no informativa de los parámetros η , ϕ_1 y σ está dada por:

$$f(\eta, \phi_1, \sigma) \propto \frac{1}{\sigma}, \quad \sigma > 0$$

Zellner calcula la verosimilitud y obtiene que:

$$f(z, z_0 | \eta, \phi_1, \sigma) \propto \frac{1}{\sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T (z_t - \eta - \phi_1 z_{t-1})^2 \right\}$$

donde $z = \{z_1, z_2, \dots, z_T\}$.

Teniendo en cuenta lo anterior, Zellner obtiene la distribución posterior para η , ϕ_1 y σ , la cual está dada por la expresión siguiente:

$$f(\eta, \phi_1, \sigma | z, z_0) \propto \frac{1}{\sigma^{n+1}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T (z_t - \eta - \phi_1 z_{t-1})^2 \right\}.$$

La distribución marginal posterior conjunta para η y ϕ_1 , se obtiene integrando sobre el dominio de σ . De forma similar se consigue la distribución marginal para σ . A partir de la distribución posterior conjunta para η y ϕ_1 se obtienen las distribuciones marginales para cada uno de esos parámetros, éstas resultan ser de la forma de una t -Student⁷. Zellner comenta que la función predictiva para z_{T+1} también es una t -Student.

Además del modelo autorregresivo de primer orden, Zellner desarrolla, siguiendo el mismo método, el modelo autorregresivo de segundo orden⁸, obteniendo finalmente, la distribución conjunta para los parámetros del modelo, misma que resulta ser una t bivariada. Zellner no hace comentario alguno sobre la forma de obtener la distribución predictiva para z_{T+1} .

Para ambos modelos, en ningún momento él hace mención sobre la forma en la cual podría llevarse a cabo la predicción de observaciones para un horizonte de tiempo más lejano.

⁷En Zellner(1971; 186-188) se pueden ver con detalle la obtención de las expresiones a las que hacemos referencia.

⁸Su modelo no incluye un término constante.

Una contribución importante del método desarrollado por Zellner es el hecho de que no necesita hacer la suposición de estacionaridad de la serie para poder aplicarlo. Para el caso en el cual el interesado sabe que la serie de tiempo es estacionaria, Zellner desarrolla una variante del método anterior, haciendo explícita la condición estacionaria de la serie⁹.

En los casos anteriores se observa que Zellner obtiene la distribución posterior para los parámetros a partir de una distribución inicial no informativa, en este caso la distribución inicial considerada es proporcional al inverso de la raíz cuadrada de la varianza. Zellner comenta que pueden utilizarse otras distribuciones iniciales, él propone la *distribución no informativa de Jeffreys* e ilustra cómo puede obtenerse esa distribución. La distribución no informativa de Jeffreys está dada por la raíz cuadrada del determinante de la matriz de información de Fisher.

Pudimos notar que el método que sigue Zellner termina con la obtención de la distribución posterior para los parámetros, sin conseguir una expresión explícita de la distribución predictiva para ambos procesos autorregresivos, por los problemas analíticos que ello representa. Broemeling(1985; 182) plantea el problema fundamental con el que se enfrentan los investigadores al tratar modelos ARMA bajo el Enfoque Bayesiano. Su punto de vista se presenta a continuación: *"La verosimilitud es intratable analíticamente para muchos modelos ARMA y esto representa un problema, no sólo para los Bayesianos sino también para aquéllos interesados en estimación por máxima-verosimilitud. ... Si la función de verosimilitud puede representarse de tal forma que se produzcan distribuciones posteriores tratables analíticamente, es posible un Análisis Bayesiano completo."*

Como una extensión del trabajo de Zellner, Broemeling presenta su tratamiento de los modelos autorregresivos. A diferencia de Zellner, quien en todo momento trabaja con distribuciones iniciales no informativas, Broemeling utiliza distribuciones conjugadas para los parámetros y no sólo considera el caso de los modelos autorregresivos de primero y segundo orden, sino que trabaja con modelos de cualquier orden, es decir, los modelos que considera tienen la forma:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + u_t$$

donde $t = 1, 2, \dots, T$; los términos de perturbación, $\{u_t\}$, tienen media cero y precisión¹⁰ τ , con $\tau > 0$; por lo cual la verosimilitud es:

$$f(z_1, \dots, z_p | \phi_1, \dots, \phi_p, \tau) \propto \tau^{T/2} \exp\left\{-\frac{\tau}{2} \sum_{t=1}^T (z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p})^2\right\}$$

⁹ Consulte Zellner(1971; 188-191).

¹⁰ La precisión es el inverso de la varianza. Se usa frecuentemente como parámetro en lugar de la varianza porque facilita las operaciones algebraicas.

en consecuencia, la distribución inicial de los parámetros está dada por:

$$f(\phi | \tau) \propto \tau^{p/2} \exp\left\{-\frac{\tau}{2}(\phi - \mu)^T P(\phi - \mu)\right\}$$

y

$$f(\tau) \propto \tau^{\alpha-1} \exp\{-\tau\beta\}$$

con $\phi = (\phi_1, \dots, \phi_p)$ y $\tau > 0$. Los parámetros μ, P, α y β se suponen conocidos, $\mu \in R^p$; $\alpha, \beta > 0$ y P es una matriz definida positiva de p -ésimo orden.

Utilizando el Teorema de Bayes, Broemeling obtiene la distribución posterior conjunta para los parámetros ϕ y τ , la cual es una normal-gama, nuevamente. Al integrar con respecto a τ , se obtiene la distribución posterior para el vector de parámetros ϕ , la cual resulta ser una densidad t . Broemeling también encuentra la expresión para la distribución predictiva de z_{T+1} ; sin embargo advierte que, al tratar de calcular la predictiva para dos observaciones no obtiene una t biviada, por lo que es necesario recurrir a métodos numéricos para poder realizar un análisis completo.

Como un punto aparte, Broemeling comenta que en el caso en el que se supone estacionaria la serie, el conjunto de distribuciones que él usa no es el más adecuado, ya que la familia normal-gama maneja un dominio muy extenso para los valores de los parámetros. El considera que si se desea incorporar al modelo la suposición de estacionaridad, una aproximación al problema, usando el método que el propone, es hacer un ajuste a los parámetros iniciales, de modo que pueda aproximarse la estacionaridad, o bien, ignorar el supuesto y continuar con el método. Ambos casos representan una aproximación al problema, sin embargo lo más adecuado es encontrar una distribución inicial que contemple esta circunstancia. No obstante, él advierte que las distribuciones posteriores llegan a ser intratables analíticamente por lo cual debe recurrirse a métodos numéricos.

Por su parte, Alegría(1988) retoma la sugerencia de Zellner(1971), considerando como distribución inicial no informativa a la distribución de Jeffreys. Alegría explica cómo puede obtenerse esta distribución para utilizarla más adelante en el cálculo de la distribución posterior conjunta del vector de parámetros de un proceso autorregresivo de orden p . De esta manera, lo novedoso en el trabajo de Alegría respecto a los procesos autorregresivos de orden p , es la derivación de un Método Bayesiano que permite calcular las distribuciones posteriores para los parámetros partiendo de distribuciones iniciales no informativas. El encuentra que la distribución marginal para el vector de parámetros del proceso es una t multivariada no central, así como también obtiene la distribución predictiva para la variable un periodo adelante dados sus valores anteriores.

Al tratar de predecir a z_{T+k} con $k > 1$, Alegría se encuentra con nuevos problemas, debido a que la distribución predictiva resultante no tiene un forma explícita;

sin embargo, comenta que algunos métodos numéricos pueden auxiliar en el cálculo de esas distribuciones, tanto predictivas como marginales¹¹.

4.3. Procesos de Promedios-Móviles

Broemeling(1985) y Alegría(1988) hacen un estudio de los modelos de promedios-móviles. Las dificultades que presentan estos modelos son mayores que aquéllas que se observan para los modelos autorregresivos, bajo el tratamiento de estos autores.

Debido a las dificultades que se le presentan a Broemeling al intentar obtener de forma explícita la distribución predictiva, él muestra únicamente cómo puede llevarse a cabo la predicción de z_{T+1} en los modelos de promedios-móviles de primer orden, esta predicción la efectúa a través de una función distinta de la predictiva.

En el caso del proceso de promedios-móviles de orden 1, Broemeling encuentra que la función que le permitirá realizar las predicciones es la distribución predictiva condicionada sobre los parámetros del modelo, es decir, $f(z_{T+1} | z_1, \dots, z_T, \theta_1)$. A partir de ella calcula la esperanza de z_{T+1} condicionada sobre los valores z_1, \dots, z_T . Esta esperanza, comenta, le puede servir como una predicción puntual de z_{T+1} . Asimismo intenta calcular la varianza de z_{T+1} condicionada sobre z_1, \dots, z_T , pero la varianza resultante depende de ϕ_1 , por lo cual nuevamente se tiene la necesidad de recurrir a métodos numéricos para poder realizar la predicción. Broemeling señala que con ese tipo de métodos, puede obtenerse la distribución predictiva de z_{T+1} , así como su media y varianza. Cabe hacer notar que las distribuciones que se encuentran al estudiar a los modelos de promedios-móviles no son distribuciones estándar¹².

Alegría(1988) estudia a los modelos de promedios-móviles aplicando el procedimiento que el mismo desarrolla para modelos ARMA(p,q), viéndolos como un caso particular de éstos, es decir, como procesos ARMA(0,q). Este procedimiento se comenta en la siguiente sección.

4.4 Procesos Autorregresivos y de Promedios-Móviles

Los modelos mixtos — autorregresivos y de promedios-móviles — al igual que sus predecesores, presentan problemas cuando se intenta obtener la distribución predictiva que permitirá hacer los pronósticos correspondientes por la dificultad al intentar expresar tales distribuciones. Sin embargo, queda de manifiesto que a través del uso de

¹¹ Para mayor información consúltese a Alegría(1988:38-73).

¹² Consulte Broemeling(1985: 187-196).

métodos numéricos puede realizarse un análisis completo.

Broemeling(1985; 197-200) considera un modelo ARMA(1,1). Recurre a la expresión para la verosimilitud que él encontró en el caso de un proceso de promedios-móviles de orden 1; a partir de ella obtiene la verosimilitud para el proceso ARMA(1,1). Supone que las distribuciones iniciales para los parámetros son informativas. La distribución inicial conjunta la descompone de la siguiente manera y es así como la ocupa:

$$p(\theta_1, \phi_1, \tau) \propto p(\phi_1 | \tau)p(\tau)p(\theta_1)$$

donde ϕ_1 es el parámetro de la parte autorregresiva y θ_1 de la parte de promedios-móviles.

Utilizándolas, obtiene la distribución marginal posterior para el parámetro de *promedios-móviles*, que no resulta ser una distribución conocida. La distribución para el parámetro *autorregresivo* está condicionada sobre el parámetro de promedios-móviles y las observaciones, pero ésta sí es una distribución conocida, es una *t - Student*. No obstante, debido a las características de las distribuciones encontradas, el análisis debe complementarse vía métodos numéricos, como se comentó con anterioridad. Sin embargo, para el caso de la distribución predictiva de Z_{T+1} esto no es necesario pues es una función *t* no central.

Broemeling intenta modificar su método con el fin de poder incluir los supuestos de invertibilidad y estacionaridad en el modelo. Su principal dificultad es que al intentar dar una distribución inicial para el parámetro autorregresivo, los parámetros de esa distribución también son parámetros de la distribución inicial de la precisión, lo cual es un supuesto bastante fuerte que no se puede sustentar.

Alegría, por su parte, considera un proceso estacionario e invertible ARMA(*p, q*). Realizando una transformación de este modelo mediante el uso de los operadores de retraso¹³ *B* y las soluciones teóricas de las ecuaciones características:

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$$

$$1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q = 0$$

obtiene una serie de tiempo auxiliar, misma que le permite darse cuenta que una aproximación a la matriz de información de Fisher para los parámetros de un proceso ARMA(*p, q*) es la matriz de información de un proceso autorregresivo de orden *p + q*.

¹³El operador de retraso a *j* pasos, se define como:

$$B^j(y_t) = y_{t-j}$$

En base a lo anterior, Alegría puede obtener una aproximación de la distribución inicial no informativa de Jeffreys asociada a un proceso $ARMA(p, q)$, a través de la cual puede calcular la distribución posterior para los parámetros de un proceso autorregresivo y de promedios-móviles.

En lo referente a procesos ARMA, Alegría concluye su trabajo en este punto. El no hace mención de la manera en la cual pueden calcularse las distribuciones predictivas para modelos con estas características.

Hemos observado las dificultades a las que se enfrentan estos autores cuando estudian los modelos $ARMA$ bajo el Enfoque Bayesiano, algunas de estos problemas se pueden subsanar mediante técnicas de análisis numérico y otros necesitan de un análisis más profundo.

En el siguiente capítulo estudiaremos a los *Modelos Lineales Dinámicos*, los cuales proporcionan una Metodología Bayesiana que permite realizar la predicción de series de tiempo mediante el uso de *Filtros de Kalman*.

La forma en que se ataca el problema difiere de los métodos presentados en este capítulo; sin embargo, la esencia del Enfoque Bayesiano queda de manifiesto en todo momento. Los alcances de los modelos lineales dinámicos son considerables, ya que varios modelos estadísticos para series de tiempo se pueden ver como casos particulares de éstos.

Como se podrá apreciar, el uso del filtro de Kalman le proporciona al modelo la flexibilidad necesaria para incorporar nuevas observaciones que le darán ese dinamismo al que hace referencia su nombre, dinamismo que se reflejará en los parámetros del modelo y que conllevará a una actualización en las predicciones.

MODELOS LINEALES DINAMICOS

Capítulo V. Modelos Lineales Dinámicos

5.1 Introducción

En el primer capítulo de esta tesis, se proporcionó, a *grosso modo*, una descripción de algunas técnicas de pronóstico de series de tiempo; entre las cuales se mencionaron a los modelos lineales dinámicos. Asimismo, se comentó que estos modelos se han desarrollado dentro de un marco Bayesiano, al cual nos referimos en el Capítulo IV.

En este capítulo, nuestro objetivo es dar a conocer a los *Modelos Lineales Dinámicos*, describir su relación con los Filtros de Kalman, así como presentar un ejemplo en el cual podremos darnos cuenta de las potencialidades que estos modelos representan.

A través de este capítulo podremos notar la *generalidad* que representa un Modelo Lineal Dinámico, ya que varios de los modelos estadísticos de predicción pueden expresarse en términos de éste, algunos modelos que tienen esta característica son: los métodos de regresión lineal, los de suavizamiento exponencial¹ y algunos otros modelos lineales.

Observaremos también que, a diferencia de otros modelos de predicción, los modelos lineales dinámicos permiten la incorporación de información por parte del usuario del modelo, al mismo tiempo que ofrecen la posibilidad de incorporar la información que cada *nueva* observación provea. Esto último representa una de las características más relevantes del modelo ya que, como se verá más adelante, el Filtro de Kalman permite tomar a las observaciones *una a una*, actualizando fácilmente los valores de los parámetros, así como las predicciones, conforme nuevas observaciones se tienen disponibles.

Otra ventaja que nos proporcionan los modelos lineales dinámicos es que se puede iniciar el proceso de predicción aún cuando no se tiene disponible ninguna observación, a través de información *a priori*. Es importante recalcar que el estudio de estos métodos se encuentra bajo el Enfoque Bayesiano, al cual ya se hizo referencia en el capítulo anterior.

Además de lo anterior, otra característica importante que muestran los modelos lineales dinámicos es que se han planteado de tal forma que pueda realizarse la predicción de varias series de tiempo a la vez, es decir, se puede hacer análisis multivariado de series

¹Véase Bowerman y O'Connell(1979; 121-132,137-198)

de tiempo con ellos. Sin embargo, debido a que en este trabajo sólo hacemos mención de técnicas univariadas para series de tiempo, no realizaremos un análisis profundo respecto de esta característica del modelo.

Como se comentó al principio de esta tesis, la predicción no se referirá únicamente a la predicción de eventos futuros, sino que también involucrará otros horizontes de tiempo. En este sentido, no resulta difícil extender el horizonte de predicción en los modelos lineales dinámicos.

Los resultados que se plasman en este capítulo, en su mayoría han sido tomados del artículo de Harrison y Stevens(1978).

5.2 El Modelo Lineal Dinámico

Debido a que el Modelo Lineal Dinámico será el objeto de estudio de este capítulo, resulta apremiante proporcionar una definición del mismo²:

Modelo Lineal Dinámico

$$\text{Ecuación de Observación: } y_t = F_t \theta_t + v_t$$

$$\text{Ecuación del Sistema: } \theta_t = G \theta_{t-1} + w_t$$

con $v_t \sim N(0, V_t)$, $w_t \sim N(0, W_t)$ y $t = 1, 2, 3, \dots$

donde:

- t es el índice de tiempo.
- y_t es el vector de observaciones del proceso realizadas al tiempo t de dimensión $n \times 1$.
- θ_t es el vector de parámetros del proceso al tiempo t de dimensión $p \times 1$.
- F_t es la matriz de variables independientes conocidas al tiempo t , cuya dimensión es $n \times p$.
- G es la matriz del sistema de dimensión $p \times p$ (conocida).

²Esta estructura la presentan Harrison y Stevens.

v_t, w_t son vectores aleatorios de dimensiones $n \times 1$ y $p \times 1$ explicados, respectivamente, a través de una normal con media cero y matriz de varianzas y covarianzas V_t y W_t conocidas al tiempo t .

$$V_t = E[v_t v_t^T]$$

$$W_t = E[w_t w_t^T]$$

t_0 es el tiempo inicial.

θ_{t_0} es el estado inicial del sistema.

y_{t_1} es la primera observación.

El Modelo Lineal Dinámico (MLD) es un sistema de ecuaciones que se compone de dos grupos. Al primer grupo se le denomina *Ecuación de Observación* y al segundo *Ecuación del Sistema*. En algunas ocasiones, llamaremos a la ecuación del sistema, *Modelo Paramétrico* o *Espacio de Estados*. Las ecuaciones especifican cómo están dependiendo las observaciones del proceso de los actuales parámetros y cómo éstos evolucionan en el tiempo.

La Ecuación de Observación explica la dependencia estocástica de la variable de observación sobre los parámetros actuales del proceso que son desconocidos. El término v_t explica lo que el estado del sistema, a través de la relación lineal, no puede explicar sobre el vector de observaciones.

Al vector θ_t , algunas veces se le denomina *Estado del Sistema al tiempo t* , es por ello, que la Ecuación del Sistema, indica la relación que existe entre el estado anterior del sistema y el estado actual. A la matriz G se le suele llamar *matriz de transición* pues a través de ella se da el cambio de un estado a otro.

Harrison y Stevens comentan que el modelo lo establecieron suponiendo intervalos de tiempo discretos e igualmente espaciados y mencionan que no es difícil extenderlo para intervalos desiguales o a un modelo de observaciones faltantes. Debido a que dentro del Enfoque Bayesiano se predice con la información que se tiene *disponible* en el momento en que se realiza el análisis, puede pensarse que el modelo que considere la irregularidad en tamaño de los intervalos de tiempo o que considere la falta de algunas observaciones, matemáticamente tendrá la misma estructura; lo que cambiará será la forma en la cual la predicción se lleva a cabo, debido a que ahora la predicción se construirá a partir del estado *más reciente* del que se tenga conocimiento.

Cuando trabajamos con Modelos Lineales Dinámicos debemos hacer uso de ciertas relaciones recursivas, estas relaciones se obtienen a través de los Filtros de Kalman. Por este motivo, en la sección siguiente se presentan a los filtros lineales dinámicos, haciendo uso de diferentes planteamientos encontrados en la literatura.

5.3 El Filtro de Kalman

En esta sección presentaremos tres planteamientos que se dan en la literatura sobre Filtros de Kalman. La razón por la cual se proporcionan estos planteamientos no está motivada por el hecho de que éstos sean los más comunes, sino más bien porque fueron los que se tuvieron al alcance. La estructura matemática que presentan es muy semejante; sin embargo, conceptualmente no representan exactamente lo mismo.

Uno de los esquemas que se presentan, con el fin de mostrar las relaciones de Kalman, ha sido obtenido a partir de una simplificación del planteamiento que hacen Goodwin y Sin(1984; 248-249), el cual en su versión original es, en algunos sentidos, más general que el que plantean Harrison y Stevens(1976; 210-211), pero también más restrictivo en otros. Las diferencias entre ambos se discuten más adelante. La versión que dan Goodwin y Sin se presenta en su versión simplificada, porque a nuestro juicio, permite ver de una forma más clara las relaciones recursivas dadas por el Filtro de Kalman. La demostración de las relaciones de Kalman derivadas de esta versión se presenta dentro de esta sección. Otro planteamiento que se presenta en la literatura y que aquí presentamos es el esquema que proporcionan Harrison y Stevens, éste se presenta con la finalidad de realizar la comparación con la versión de Goodwin y Sin a la que ya se ha hecho mención. En la siguiente sección se discutirá con mayor amplitud. Finalmente, se menciona el planteamiento que proporciona Jazwinski(1970; 200-201). La demostración para los dos últimos planteamientos puede verse en el apéndice.

El planteamiento de Goodwin y Sin

El modelo original que Goodwin y Sin(1984) consideran tiene la siguiente estructura:

$$x(t+1) = Ax(t) + Bu(t) + v_1(t) \quad (1)$$

$$y(t) = Cx(t) + v_2(t) \quad (2)$$

donde $x(t_0)$ tiene media \bar{x}_0 y matriz de varianzas y covarianzas Σ_0 y $\{v_1(t)\}$ y $\{v_2(t)\}$ son procesos estacionarios de ruido blanco con media cero y matriz de covarianzas dada por:

$$E\left\{\begin{pmatrix} v_1(t) \\ v_2(t) \end{pmatrix} \begin{pmatrix} v_1(s)^T & v_2(s)^T \end{pmatrix}\right\} = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta(t-s);$$

$\delta(\cdot)$ es la delta de Kronecker, $Q \geq 0$ y $R > 0$. Goodwin y Sin, en el planteamiento de lo que ellos llaman *El filtro lineal óptimo*, no hacen alguna suposición sobre la distribución de las sucesiones de ruido, la suposición de *normalidad* la hacen explícita cuando muestran las relaciones de Kalman.

Con el fin de facilitar la comparación del planteamiento de Goodwin y Sin con la versión de Harrison y Stevens, incluimos a continuación esta última.

El planteamiento de Harrison y Stevens

$$\begin{aligned} \text{Ecuación de Observación: } & y_t = F_t \theta_t + v_t \\ \text{Ecuación del Sistema: } & \theta_t = G \theta_{t-1} + w_t \end{aligned}$$

con $v_t \sim \mathcal{N}(0, V_t)$ y $w_t \sim \mathcal{N}(0, W_t)$

Al comparar la estructura anterior con el planteamiento que dan Goodwin y Sin, observamos lo siguiente:

- La ecuación de observación de Harrison y Stevens coincide con la ecuación (2) anterior, si consideramos que $x(t) = \theta_t$, $v_2(t) = v_t$ y $C = F$. De este modo observamos que en la ecuación (2) la matriz C es invariante en el tiempo; en este sentido puede decirse que el modelo de Harrison y Stevens es más general ya que, en el modelo lineal dinámico, se permite que la matriz F dependa del tiempo.

- Al tratar de conciliar la ecuación del sistema con la ecuación (1) de Goodwin y Sin, observamos que los modelos considerados son semejantes pero no puede verse a uno como caso particular del otro, ya que al tratar de hacer la analogía entre ambas ecuaciones notamos la necesidad de tomar a $\theta_t = x(t+1)$, sin embargo esto va en contra de lo que se supuso anteriormente, es decir que, $x(t) = \theta_t$ en la ecuación de observación.

- Además de lo que ya hemos mencionado, podemos notar que el modelo dado por Goodwin y Sin considera que las componentes aleatorias v_t y w_t pueden estar correlacionadas, mientras que el modelo de Harrison y Stevens (1976) no lo considera, sin embargo, el modelo de Harrison y Stevens permite que las matrices de varianzas y covarianzas de estas componentes aleatorias varíen en el tiempo, mientras que Goodwin y Sin no lo permiten.

Si consideramos que $Bu(t) + v_1(t) = w_t$, $\theta_t = x(t)$, $A = G$, $C = F$ y $S = 0$, entonces el modelo que obtenemos es el siguiente:

$$\begin{aligned} \theta_{t+1} &= G\theta_t + w_t \\ y_t &= F\theta_t + v_t \end{aligned}$$

$w_t \sim \mathcal{N}(0, Q)$ y $v_t \sim \mathcal{N}(0, R)$

El modelo anterior se utilizará para mostrar las relaciones recursivas de Kalman; en lo sucesivo nos referiremos a él como modelo simplificado de Goodwin y Sin.

TEOREMA. Consideremos el sistema:

$$\begin{aligned} \theta_{t+1} &= G\theta_t + w_t \\ y_t &= F\theta_t + v_t \end{aligned}$$

Sea θ_{i0} el estado inicial del sistema, con media $\bar{\theta}_0$ y con matriz de varianzas y covarianzas Σ_0 . Supongamos que $E\{w_t w_t^T\} = Q$ y $E\{v_t v_t^T\} = R$, donde $Q \geq 0$ y $R > 0$. Además

supongamos que el estado inicial θ_{t_0} y las variables $\{v_s \mid s = t_0, \dots, t\}$ y $\{w_s \mid s = t_0, \dots, t\}$ son independientes y conjuntamente se explican normales. Denotemos por $\hat{\theta}_{t+1}$ a la media condicional de θ_{t+1} dadas las observaciones $\{y_s \mid s = t_0, \dots, t\}$. Bajo las suposiciones anteriores $\hat{\theta}_{t+1}$ satisface las siguientes relaciones recursivas (El filtro de Kalman):

$$\begin{aligned}\hat{\theta}_{t+1} &= G\hat{\theta}_t + K_t[y_t - F\hat{\theta}_t] \\ \hat{\theta}_{t_0} &= \bar{\theta}_0\end{aligned}$$

donde K_t es la ganancia del filtro y está dada por:

$$K_t = G\Sigma_t F^T [F\Sigma_t F^T + R]^{-1}$$

y Σ_t es la matriz de covarianzas del error del estado, es decir,

$$\Sigma_t := E\{[\hat{\theta}_t - \theta_t][\hat{\theta}_t - \theta_t]^T \mid y_{t-1}, y_{t-2}, \dots, y_{t_0}\}$$

Además Σ_t satisface las siguientes ecuaciones:

$$\begin{aligned}\Sigma_{t+1} &= G\Sigma_t G^T + Q - K_t[F\Sigma_t F^T + R]K_t^T \\ \Sigma_{t_0} &= \Sigma_0\end{aligned}$$

Demstración. Por inducción. Notemos que el resultado es válido para t_0 , ya que por hipótesis, al tiempo t_0 , θ_0 tiene media $\bar{\theta}_0$ y matriz de varianzas y covarianzas Σ_0 por lo cual:

$$\begin{aligned}\hat{\theta}_{t_0} &= \bar{\theta}_0 \\ \Sigma_{t_0} &= \Sigma_0\end{aligned}$$

Ahora, supondremos que la distribución condicional para θ_t , dadas $y_{t-1}, y_{t-2}, \dots, y_{t_0}$, es normal con media $\hat{\theta}_t$ y matriz de varianzas y covarianzas Σ_t ; con ella determinaremos la distribución condicional de θ_{t+1} dadas $y_t, y_{t-1}, \dots, y_{t_0}$.

Observemos que el sistema

$$\begin{aligned}\theta_{t+1} &= G\theta_t + w_t \\ y_t &= F\theta_t + v_t\end{aligned}$$

puede escribirse de la siguiente manera:

$$\begin{pmatrix} \theta_{t+1} \\ y_t \end{pmatrix} = \begin{pmatrix} G & I_p & 0 \\ F & 0 & I_n \end{pmatrix} \begin{pmatrix} \theta_t \\ w_t \\ v_t \end{pmatrix}$$

donde I_p y I_n son matrices identidad de dimensiones $p \times p$ y $n \times n$, respectivamente. Para poder obtener las expresiones que a continuación se muestran, será necesario recurrir a

algunos resultados sobre variables aleatorias normales, mismos que se presentan en el apéndice A.3, en la parte final de este trabajo.

Utilizando el Teorema I y II obtenemos que, la distribución conjunta de $\begin{pmatrix} \theta_{t+1} \\ y_t \end{pmatrix}$ dadas las observaciones $y_{t-1}, y_{t-2}, \dots, y_{t_0}$, es normal con media $\begin{pmatrix} G\hat{\theta}_t \\ F\hat{\theta}_t \end{pmatrix}$ y matriz de varianzas y covarianzas

$$\left(\begin{array}{c|c} G\Sigma_t G^T + Q & G\Sigma_t F^T \\ \hline F\Sigma_t G^T & F\Sigma_t F^T + R \end{array} \right)$$

Por lo tanto, si calculamos la distribución condicional de θ_{t+1} dadas $y_t, y_{t-1}, \dots, y_{t_0}$, a partir de esta última expresión observamos que³:

$$\theta_{t+1} | y_t, y_{t-1}, \dots, y_{t_0} \sim \mathcal{N}(\hat{\theta}_{t+1}, \Sigma_{t+1})$$

donde

$$\begin{aligned} \hat{\theta}_{t+1} &= G\hat{\theta}_t + G\Sigma_t F^T [F\Sigma_t F^T + R]^{-1} (y_t - F\hat{\theta}_t) \\ \Sigma_{t+1} &= G\Sigma_t G^T + Q - G\Sigma_t F^T [F\Sigma_t F^T + R]^{-1} F\Sigma_t G^T \end{aligned}$$

A partir de la expresión para Σ_t obtenemos lo siguiente:

$$\begin{aligned} \Sigma_{t+1} &= G\Sigma_t G^T + Q - G\Sigma_t F^T [F\Sigma_t F^T + R]^{-1} F\Sigma_t G^T \\ &= G\Sigma_t G^T + Q - G\Sigma_t F^T [F\Sigma_t F^T + R]^{-1} [F\Sigma_t F^T + R] [F\Sigma_t F^T + R]^{-1} [G\Sigma_t F^T]^T \end{aligned}$$

Luego, si denotamos por K_t a:

$$K_t := [G\Sigma_t F^T] [F\Sigma_t F^T + R]^{-1}$$

entonces, las expresiones para $\hat{\theta}_{t+1}$ y Σ_{t+1} serán las siguientes:

$$\begin{aligned} \hat{\theta}_{t+1} &= G\hat{\theta}_t + K_t (y_t - F\hat{\theta}_t) \\ \Sigma_{t+1} &= G\Sigma_t G^T + Q - K_t [F\Sigma_t F^T + R] K_t^T \end{aligned}$$

Con lo cual queda probado el resultado. •

Como puede observarse el filtro de Kalman proporciona una forma de estimar el estado θ_t del sistema. El filtro tiene las siguientes propiedades⁴:

³ Véase el Teorema III del apéndice A.3, al final de esta tesis.

⁴ Véase Goodwin y Sin (1984; págs. 246-251).

- Bajo la suposición de *normalidad* el filtro arroja el estimador insesgado de varianza mínima, i.e., el filtro evalúa la media de θ_t dada la información pasada $y_{t-1}, y_{t-2}, \dots, y_{t_0}$. En términos matemáticos se expresa como:

$$\hat{\theta}_t = E[\theta_t | y_{t-1}, y_{t-2}, \dots, y_{t_0}]$$

o bien,

$$E[\hat{\theta}_t - \theta_t | y_{t-1}, y_{t-2}, \dots, y_{t_0}] = 0$$

y

$$\Sigma_t := E\{[\hat{\theta}_t - \theta_t][\hat{\theta}_t - \theta_t]^T | y_{t-1}, y_{t-2}, \dots, y_{t_0}\} \leq \Sigma_e$$

donde Σ_e es la covarianza de error dada por cualquier otro estimador de θ_t .

- La ganancia del Filtro de Kalman es que K_t y la covarianza del error Σ_t son independientes de $y_{t-1}, y_{t-2}, \dots, y_{t_0}$ dado que las matrices G, F, R y Q son todas independientes de $y_{t-1}, y_{t-2}, \dots, y_{t_0}$.

- Si no se pide *normalidad*, el filtro arroja el estimador *lineal* del estado, de varianza mínima, i.e., proporciona entre todos los estimadores lineales a aquél que tiene la más pequeña covarianza del error. En este caso el error $(\theta_t - \hat{\theta}_t)$ no está correlacionado con $y_{t-1}, y_{t-2}, \dots, y_{t_0}$, o sea que,

$$E\{(\theta_t - \hat{\theta}_t)y_{t-i}\} = 0; \quad i = 1, 2, \dots, t - t_0$$

Ahora bien, si definimos a I_t como

$$I_t := y_t - F\hat{\theta}_t$$

se puede ver que I_t tiene las siguientes propiedades:

$$E\{I_t | y_{t-1}, y_{t-2}, \dots, y_{t_0}\} = 0$$

lo cual nos conduce a que

$$\begin{aligned} y_t &= F\hat{\theta}_t + I_t \\ &= E\{y_t | y_{t-1}, y_{t-2}, \dots, y_{t_0}\} + I_t \end{aligned}$$

Si nos fijamos en la última igualdad podemos observar que los valores I_t representan la nueva información que no está contenida en $y_{t-1}, y_{t-2}, \dots, y_{t_0}$. Por este motivo a los valores de I_t se les conoce como *sucesión de innovaciones*.

Además de lo que ya hemos mencionado, resulta interesante notar que la media del estado del sistema, $\bar{\theta}_t := E[\theta_t]$, y la covarianza del mismo, $\bar{C}_t := E\{[\theta_t - \bar{\theta}_t][\theta_t - \bar{\theta}_t]^T\}$, satisfacen las siguientes relaciones recursivas:

$$\begin{aligned} \bar{\theta}_{t+1} &= G\bar{\theta}_t \\ \bar{C}_{t+1} &= G\bar{C}_tG^T + Q \end{aligned}$$

Resulta conveniente recordar que los resultados anteriormente comentados pueden encontrarse en el libro de Goodwin y Sin.

A continuación, presentamos sin demostración, las relaciones de Kalman en la versión de Jazwinski.

El planteamiento de Jazwinski

La estructura matemática que usa Jazwinski (1970; 200-201) con el fin de mostrar qué son y en qué consisten las relaciones recursivas de Kalman, se presenta a continuación. Presentamos la versión que él establece haciendo uso de la notación que él emplea. Cabe mencionar que Jazwinski denomina a su modelo *Filtro Discreto*.

El sistema lineal discreto se describe mediante las siguientes expresiones:

$$x_{k+1} = \Phi(t_{k+1}, t_k)x_k + \Gamma(t_k)w_{k+1}, \quad k = 0, 1, \dots$$

donde x_k es el estado del sistema al tiempo t_k , de dimensión $n \times 1$, Φ es la matriz de transición del estado de dimensión $n \times n$ no-singular, Γ es una matriz de $n \times r$, y $\{w_k, k = 1, 2, \dots\}$ es una sucesión de ruido, cada w_k es un vector columna de dimensión r tal que $w_k \sim \mathcal{N}(0, Q_k)$. Las observaciones están dadas por:

$$y_k = M(t_k)x_k + v_k.$$

$x_0, \{w_k\}$ y $\{v_k\}$ se suponen independientes. Si denotamos por:

$$Y_r := \{y_1, \dots, y_r\}$$

$$\hat{x}_s^r := E\{x_s | Y_r\}$$

y

$$P_s^r := E\{[(x_s - \hat{x}_s^r)(x_s - \hat{x}_s^r)^T] | Y_r\}$$

entonces, el filtro lineal óptimo (de varianza mínima) tiene como media condicional y matriz de covarianzas a:

$$\begin{aligned} \hat{x}_{k+1}^k &= \Phi(t_{k+1}, t_k)\hat{x}_k^k \\ P_{k+1}^k &= E\{[(\Phi(x_k - \hat{x}_k^k) + \Gamma w_{k+1})(\Phi(x_k - \hat{x}_k^k) + \Gamma w_{k+1})^T] | Y_k\} \\ &= \Phi(t_{k+1}, t_k)P_k^k\Phi^T(t_{k+1}, t_k) + \Gamma(t_k)Q_{k+1}\Gamma^T(t_k). \end{aligned}$$

además:

$$\begin{aligned} \hat{x}_k^k &= \hat{x}_k^{k-1} + K(t_k)(y_k - M(t_k)\hat{x}_k^{k-1}), \\ P_k^k &= P_k^{k-1} - K(t_k)M(t_k)P_k^{k-1}, \end{aligned}$$

donde

$$K(t_k) = P_k^{k-1} M^T(t_k) [M(t_k) P_k^{k-1} M^T(t_k) + R_k]^{-1}$$

es la ganancia de Kalman. De este modo, el esquema de predicción se complementa con las condiciones iniciales (\hat{x}_k^k, P_k^k) .

Lo anterior representa la versión que Jazwinski establece para presentación de los Filtros de Kalman. Cabe aclarar que él no da demostración alguna del resultado que presenta. La demostración de los resultados que él establece pueden verse en el apéndice A.4 de esta tesis.

5.4 El Filtro de Kalman y El Modelo Lineal Dinámico

A continuación presentamos la interrelación que existe entre el filtro de Kalman y el Modelo Lineal Dinámico, describiendo primero la manera en la cual se determina la distribución posterior del estado del sistema y haciendo después la descripción del proceso de predicción.

Consideremos el MLD, presentado en la sección 5.2. Supongamos que la información concerniente al parámetro θ_0 puede modelarse a través de una distribución normal con media m_0 y matriz de covarianzas C_0 ; supongamos también que la primera observación de que disponemos es y_{t_1} .

Con estas hipótesis en mente y mediante las relaciones de recurrencia del filtro de Kalman obtenemos la distribución posterior para θ_t :

$$\theta_t | y_t, y_{t-1}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1} \sim \mathcal{N}(m_t, C_t)$$

Los valores m_t y C_t se obtienen de la siguiente manera⁵:

$$m_t = G m_{t-1} + A e$$

$$C_t = R - A \hat{Y}_t A^T$$

donde:

$$A = R F_t^T \hat{Y}_t^{-1}$$

$$\hat{Y}_t = F_t R F_t^T + V_t$$

$$R = G C_{t-1} G^T + W_t$$

$$e = y_t - \hat{y}_t$$

$$\hat{y}_t = F_t G m_{t-1}$$

⁵ El artículo de Harrison y Stevens (1978) no incluye una demostración de los resultados que se mencionan en esta sección, sin embargo, una aportación de este trabajo es aclarar cómo se obtienen tales resultados. Consulte el Apéndice A.5.

Como podemos observar, la naturaleza recursiva del algoritmo es importante, debido a que nos permite calcular la distribución posterior del estado del sistema a partir del valor de la observación más reciente, y_t , del valor de la matriz de variables independientes, F_t , de las varianzas de las perturbaciones y el ruido de las observaciones actuales V_t y W_t y a través de la distribución de

$$\theta_{t-1} \mid y_{t-1}, y_{t-2}, \dots, y_1, F_{t-1}, F_{t-2}, \dots, F_1.$$

La mayor ventaja que ofrecen los modelos lineales dinámicos, en la opinión de Harrison y Stevens, es que, por primera vez, se tiene una estructura coherente y razonable para predecir en situaciones en las cuales no existen observaciones o las observaciones son escasas, y también en situaciones en las cuales se necesitan desarrollar sistemas que requieren modelos estructurados que faciliten la comunicación entre hombre y método de predicción.

Hasta ahora hemos mostrado únicamente cómo se obtiene la distribución posterior del vector de parámetros del proceso; sin embargo, aún queda una pregunta en el aire, ¿Cómo vamos a predecir los valores de las observaciones futuras? A continuación veremos cómo las relaciones recursivas mostradas anteriormente nos permitirán inferir la distribución de las observaciones futuras, y_{t+k} , para $k = 1, 2, \dots$

Las predicciones que arroja este modelo son de naturaleza distribucional. Se derivan de la varianza del ruido de las observaciones futuras V_{t+k} , y de la varianza de las perturbaciones del sistema W_{t+k} , por lo cual es indispensable que estas matrices se conozcan.

De las ecuaciones del Modelo Lineal Dinámico podemos derivar las ecuaciones correspondientes al valor futuro de la variable de observación y al estado futuro del sistema:

$$\begin{aligned} y_{t+k} &= F_{t+k}\theta_{t+k} + v_{t+k} \\ \theta_{t+k} &= G\theta_{t+k-1} + w_{t+k}. \end{aligned}$$

Suponiendo que las varianzas de los vectores aleatorios v_{t+k} y w_{t+k} se conocen, la predicción de los valores futuros de y_{t+k} requiere de la inferencia del valor futuro del vector de parámetros θ_{t+k} y de la matriz de variables independientes F_{t+k} . Sea

$$\begin{aligned} \hat{m}_{t+k}^i &= E[\theta_{t+k} \mid y_t, y_{t-1}, \dots, y_1, F_t, F_{t-1}, \dots, F_1] \\ \hat{C}_{t+k}^i &= \text{Var}[\theta_{t+k} \mid y_t, y_{t-1}, \dots, y_1, F_t, F_{t-1}, \dots, F_1] \quad k = 1, 2, \dots \end{aligned}$$

definiendo a

$$\begin{aligned} \hat{m}_t^i &= m_t = E[\theta_t \mid y_t, y_{t-1}, \dots, y_1, F_t, F_{t-1}, \dots, F_1] \\ \hat{C}_t^i &= C_t = E[\theta_t \mid y_t, y_{t-1}, \dots, y_1, F_t, F_{t-1}, \dots, F_1] \end{aligned}$$

donde m_t y C_t se conocen a través de las relaciones de Kalman. Hemos modificado ligeramente la notación que Harrison y Stevens utilizan porque creemos que la que proponemos es mnemotécnica; el subíndice representa el periodo que queremos predecir, mientras que el supraíndice indica el periodo más reciente para el cual contamos con información.

Utilizando la ecuación del sistema para el tiempo $t+k$ podemos calcular los valores de \hat{m}_{t+k}^t y de \hat{C}_{t+k}^t :

$$\begin{aligned}\hat{m}_{t+k}^t &= G\hat{m}_{t+k-1}^t \\ \hat{C}_{t+k}^t &= G\hat{C}_{t+k-1}^t G^T + W_{t+k} \quad k = 1, 2, \dots\end{aligned}$$

de modo que el valor de la media y la varianza del vector de parámetros θ_{t+k} dada la información disponible hasta el momento t , puede calcularse recursivamente usando los valores conocidos iniciales de \hat{m}_t^t y \hat{C}_t^t , así como el valor de W_{t+k} .

Denotemos por \hat{y}_{t+k}^t y \hat{Y}_{t+k}^t a la media y a la varianza, de la observación del proceso k periodos adelante dada la información disponible hasta el tiempo t , es decir:

$$\begin{aligned}\hat{y}_{t+k}^t &= E[y_{t+k} | y_t, y_{t-1}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1}] \\ \hat{Y}_{t+k}^t &= \text{Var}[y_{t+k} | y_t, y_{t-1}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1}] \quad k = 1, 2, \dots\end{aligned}$$

Para poder calcular estos valores es necesario distinguir dos casos, uno cuando F_{t+k} se conoce al tiempo t y el otro cuando no se conoce. En el caso cuando F_{t+k} se conoce, podemos expresar a \hat{y}_{t+k}^t y a \hat{Y}_{t+k}^t de la siguiente forma:

$$\begin{aligned}\hat{y}_{t+k}^t &= F_{t+k} \hat{m}_{t+k}^t \\ \hat{Y}_{t+k}^t &= F_{t+k} \hat{C}_{t+k}^t F_{t+k}^T + V_{t+k} \quad k = 1, 2, \dots\end{aligned}$$

Es interesante observar la relativa facilidad con que podemos predecir el valor de las observaciones al tiempo $t+k$ cuando conocemos el valor de F_{t+k} , ya que \hat{m}_{t+k}^t y \hat{C}_{t+k}^t se obtienen a través de las relaciones recursivas mencionadas con anterioridad.

La situación se vuelve compleja cuando se desconoce el valor de F_{t+k} ya que de alguna manera debemos reflejar en el modelo tal desconocimiento. Recordemos el significado de la matriz⁶ F y hagamos un pequeño paréntesis. La matriz F contiene la información relacionada con las variables independientes o explicativas del vector de observaciones. Es común que dentro del análisis de series de tiempo se trate de explicar el comportamiento de una serie de tiempo dada, con otra serie de tiempo; si la serie de tiempo *explicativa* es lo *suficientemente* conocida entonces se puede esperar que la predicción de la serie de tiempo a *explicar* no se dificulte a causa de la serie de tiempo

⁶ Omitimos los subíndices porque queremos referirnos a la matriz F para cualquier tiempo

explicativa, sin embargo, cuando el conocimiento sobre el comportamiento futuro de la serie de tiempo *explicativa* es escaso, es comprensible que se tenga mayor dificultad en la predicción de la serie de tiempo *objetivo* debido a que trataremos de predecir en primera instancia el valor de la serie de tiempo *explicativa*, para luego poder predecir el comportamiento de la serie de tiempo que nos interesa explicar. Por esta razón es comprensible que el análisis se dificulte cuando la serie de tiempo explicativa no es suficientemente conocida. Dentro de este contexto conviene recordar también que el modelo lineal dinámico está planteado de modo que se pueda realizar la predicción de varias series de tiempo a través de un conjunto de variables conocidas interrelacionadas con aquéllas; en este caso, es evidente que las dificultades serán mucho mayores a las que se presentan en los casos univariados.

Harrison y Stevens indican cómo pueden llevarse a cabo las predicciones aún cuando se desconoce el valor de la matriz F_{t+k} . Suponen que F_{t+k} puede expresarse como un valor esperado \hat{F}_{t+k}^t y un término aleatorio¹ δF_{t+k}^t , i. e.:

$$F_{t+k} = \hat{F}_{t+k}^t + \delta F_{t+k}^t = \begin{pmatrix} \hat{F}_{t+k_1}^t + \delta F_{t+k_1}^t \\ \vdots \\ \hat{F}_{t+k_n}^t + \delta F_{t+k_n}^t \end{pmatrix}$$

donde cada $\hat{F}_{t+k_i}^t$ y $\delta F_{t+k_i}^t$, para $i = 1, 2, \dots, n$ son vectores de dimensión $1 \times p$, componentes² de las matrices \hat{F}_{t+k}^t y $\delta \hat{F}_{t+k}^t$ respectivamente, donde

$$\hat{F}_{t+k}^t = E[F_{t+k} | F_t, F_{t-1}, \dots, F_{t_1}]$$

sus covarianzas son conocidas y están dadas por:

$$\phi_{ij;t+k}^t = \text{Cov}[\delta F_{t+k_i}^t, \delta F_{t+k_j}^{tT}] \\ i, j = 1, 2, \dots, p$$

Ellos aseguran que \hat{F}_{t+k}^t y $\phi_{ij;t+k}^t$ conforman la distribución inicial al tiempo t , del valor futuro de la matriz F_{t+k} . Comentan que con las definiciones anteriores, ellos obtienen lo que a continuación se presenta:

$$\hat{Y}_{t+k}^t = \hat{F}_{t+k}^t \hat{m}_{t+k}^t \\ \hat{Y}_{t+k}^t = \hat{F}_{t+k}^t \hat{C}_{t+k}^t \hat{F}_{t+k}^{tT} + V_{t+k} + Z$$

con

$$Z = Z_{ij} \quad i, j = 1, 2, \dots, n$$

y el elemento

$$Z_{ij} = \text{traza}\{\phi_{ij;t+k}^t (\hat{m}_{t+k_i}^t \hat{m}_{t+k_j}^{tT} + \hat{C}_{t+k}^t)\}$$

¹ De aquí en adelante se adoptará parcialmente la notación que ellos utilizan, anteponiendo la letra griega δ a las perturbaciones que ocurren en una variable específica.

² El subíndice i indica que se trata de la i -ésima columna.

El término Z representa la incertidumbre adicional en la predicción, incertidumbre originada por el desconocimiento de la matriz F_{t+k} .

Harrison y Stevens citan a Feldstein(1971) como autor del resultado anterior. El lector interesado puede obtener la referencia sobre este trabajo al final de esta tesis.

En la siguiente sección mostramos cómo algunos modelos pueden expresarse como modelos lineales dinámicos.

5.5 Algunos modelos que pueden formularse como modelos lineales dinámicos

Desde el punto de vista de Harrison y Stevens, los modelos lineales dinámicos son poco conocidos por las personas que trabajan con modelos; en virtud de esa opinión, ellos muestran, en su artículo, cómo algunos modelos que nos resultan familiares pueden establecerse como modelos lineales dinámicos.

Siguiendo la idea de Harrison y Stevens, en esta sección mostramos cómo algunos modelos *conocidos* pueden formularse como modelos lineales dinámicos. Es conveniente aclarar que en todos los casos se supondrá que la matriz de varianzas y covarianzas de las perturbaciones se conoce al tiempo t .

5.5.1 El modelo estático de regresión

Los modelos de regresión resultan ser bastante familiares para la mayor parte de las personas que han trabajado con modelos estadísticos. El modelo de regresión se expresa comúnmente en los siguientes términos:

$$y_t = X_t \beta + \epsilon_t; \quad t = 1, 2, 3, \dots$$

donde y_t es una observación univariada, X_t es un vector de variables independientes de dimensión $1 \times p$ y β es un vector de coeficientes desconocidos de dimensión $p \times 1$. El modelo estático de regresión es equivalente al siguiente modelo, el cual tiene la forma de un modelo lineal dinámico:

$$\begin{aligned} y_t &= X_t \beta_t + \epsilon_t \\ \beta_t &= I \beta_{t-1} \quad t = 1, 2, 3, \dots \end{aligned}$$

donde I es la matriz identidad de dimensión $p \times p$.

El algoritmo de filtraje dado por el modelo lineal dinámico nos proporciona un método que nos permite la actualización de los estimadores de los coeficientes de regresión cada vez que disponemos de nueva información; por otra parte, no tiene la restricción de que la varianza del ruido sea constante, por lo cual no existe la necesidad de buscar transformaciones que lo consigan.

5.5.2 El modelo dinámico de regresión

Cuando se trabaja con modelos lineales estáticos, es común estimar los coeficientes de estos modelos ajustando información pasada sin actualizar los valores encontrados conforme nueva información se encuentra disponible; esto suscita algunas dificultades debido a que no existe una razón convincente que haga suponer que los coeficientes estimados en este momento permanecerán sin cambio en el transcurso del tiempo. Asimismo cuando deseamos hacer estimaciones consideramos que si tenemos *bastantes* observaciones entonces los estimadores que consigamos a través de esas observaciones serán más confiables que si los hubiésemos conseguido con pocas, sin embargo, si tomamos en cuenta lo que mencionamos con anterioridad, llegamos a un punto en el cual no está claro cómo debemos proceder debido a que podemos obtener estimadores *anticuados* o totalmente *engañosos* cuando trabajamos con información abundante pero poco relevante; es en este momento cuando debemos preguntarnos ¿Cómo y de qué manera detectamos que la información de que disponemos es o deja de ser relevante para nuestro estudio?.

El modelo estático no puede hacer frente a situaciones como las que acabamos de describir pero existe otro modelo que ofrece algunos medios para poder superar esas dificultades: El modelo dinámico de regresión.

$$y_t = X_t \beta_t + \epsilon_t$$
$$\beta_t = \beta_{t-1} + \delta \beta_t \quad t = 1, 2, 3, \dots$$

Este modelo puede formularse como un modelo lineal dinámico si consideramos que $n = 1$, $\theta_t = \beta_t$, $F_t = X_t$ y $G = I_{p,p}$. Harrison y Stevens comentan que la varianza de $\delta \beta_t$ será propuesta por el modelador en base al conocimiento que tenga sobre la estructura del proceso.

5.5.3 El modelo *steady*

El modelo *steady*³ puede escribirse en la siguiente forma:

$$y_t = \mu_t + \epsilon_t$$
$$\mu_t = \mu_{t-1} + \delta \mu_t \quad t = 1, 2, 3, \dots$$

donde μ_t es un variable dependiente del tiempo alrededor de la cual fluctúa el valor de y_t . Este proceso es una de las formas más simples del modelo lineal dinámico que existen, ya que $n = p = 1$, $\theta_t = \mu_t$, $F_t = G = 1$, $V_t = \text{Var}[\epsilon_t]$, y $W_t = \text{Var}[\delta \mu_t]$.

³ Utilizamos la palabra en inglés debido a que, en nuestra opinión, ninguna de las traducciones de la misma resultan adecuadas para ilustrar la forma en la que opera este modelo.

5.5.4 El modelo lineal de crecimiento

El modelo lineal de crecimiento es una extensión del modelo *steady*. Las ecuaciones del proceso son las siguientes:

$$\begin{aligned}y_t &= \mu_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + \beta_t + \delta\mu_t \\ \beta_t &= \beta_{t-1} + \delta\beta_t \quad t = 1, 2, 3, \dots\end{aligned}$$

La interpretación que comúnmente se le da a μ_t es la del nivel del proceso al tiempo t . Al reexpresar este modelo en términos de un modelo lineal dinámico obtenemos lo que a continuación se muestra:

$$\begin{aligned}y_t &= (1 \ 0) \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} + \epsilon_t \\ \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} \delta\mu_t + \delta\beta_t \\ \delta\beta_t \end{pmatrix} \quad t = 1, 2, 3, \dots\end{aligned}$$

En este caso la varianza del vector de perturbación en la ecuación del sistema, deberá incluir la covarianza entre los términos $\delta\mu_t$ y $\delta\beta_t$ debido a que éstos aparecen sumados en una de las entradas del vector.

5.5.5 El modelo autorregresivo de orden p

Los modelos autorregresivos de orden p , $AR(p)$, se definen generalmente como

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + z_{t-p} + u_t \quad t = 1, 2, 3, \dots$$

Como hemos visto en las secciones precedentes, uno de los objetivos principales de algunos métodos es estimar los coeficientes $\phi_1, \phi_2, \dots, \phi_p$ y la varianza del término de perturbación u_t . En estos métodos comúnmente, las estimaciones obtenidas no se sujetan a una revisión frecuente. Dentro de esta metodología, el modelo autorregresivo se plantea de la siguiente manera:

$$\begin{aligned}z_t &= (z_{t-1}, z_{t-2}, \dots, z_{t-p})\phi_t + u_t \\ \phi_t &= \phi_{t-1} + \delta\phi_t\end{aligned}$$

donde

$$\phi_t = \begin{pmatrix} \phi_{1t} \\ \vdots \\ \phi_{pt} \end{pmatrix} \quad t = 1, 2, 3, \dots$$

De esta manera, teóricamente se superan los problemas emanados por la no-revisión de los parámetros. En la opinión de Harrison y Stevens se recomienda que

al tratar de obtener las predicciones para este modelo se recurra a la estimación de F_{t+k} que ellos proponen ya que, en este caso, es obvio que esta matriz es desconocida. Otra ventaja que señalan Harrison y Stevens es que mediante este modelo pueden realizarse cambios a los coeficientes $\{\phi_i\}$, ya que la ecuación del sistema deja abierta esa posibilidad.

5.5.6 El modelo de promedios-móviles de orden q

Los modelos de promedios-móviles $MA(q)$, presentan problemas al tratar de expresarlos como modelos lineales dinámicos debido a la no linealidad de este modelo. No obstante, se han sugerido algunos métodos para superar estas dificultades, al menos parcialmente.

El primer método propone que el modelo de promedios-móviles sea aproximado mediante un proceso autorregresivo. Se comenta que el grado del modelo autorregresivo puede obtenerse por ensayo y error. Un segundo método presupone que los parámetros del modelo se conocen y que a través de ellos se tratará de obtener información sobre los términos de perturbación, efectuando estimaciones de estos términos y revisándolos a través del planteamiento de un modelo lineal dinámico. Otro método parecido al anterior, supone que los términos de perturbación se obtendrán como diferencia entre una observación real y su correspondiente estimación, la cual se calculará con la información que se tiene hasta un periodo atrás y a diferencia del método anterior, considera que con éstas estimaciones se tratarán de obtener las estimaciones de los parámetros, mismas que se irán actualizando mediante un modelo lineal dinámico.

Como puede observarse, en todas las opciones que se plantean existen dificultades al intentar estudiar a los modelos de promedios-móviles. En nuestra opinión, las dificultades que se presentan dentro de este contexto se deben, principalmente a la necesidad de obtener por parte del interesado, la información necesaria para modelar una suma de términos aleatorios dependientes del tiempo, situación que se presenta bastante complicada.

A continuación presentamos un ejemplo que nos permite vislumbrar las posibilidades que nos puede proporcionar un modelo lineal dinámico en la predicción de series de tiempo.

5.6 Ejemplo

Con el fin de mostrar las características de los modelos lineales dinámicos a continuación presentamos un ejemplo de éstos. En este ejemplo se utilizará una serie de tiempo que, por sus características, representa una buena prueba para observar cuál es la sensibilidad del modelo a los cambios bruscos, corroborando así su dinamismo.

La serie de tiempo que consideraremos para tal efecto, es la serie que está compuesta por las cifras de inflación mensual de la República Mexicana, es decir, la serie de las variaciones mensuales en el Índice Nacional de Precios al Consumidor que publica el Banco de México.

Para tal efecto hemos considerado el periodo que va de enero de 1980 a junio de 1989; en consecuencia la serie está compuesta por 114 datos.

Cabe mencionar que el ejemplo que presentamos se proporciona con el fin de ilustrar cómo pueden utilizarse los métodos lineales dinámicos. A continuación se presenta la tabla que contiene los valores de la serie *inflacionaria*, así como la gráfica que muestra el comportamiento de ésta durante el periodo mencionado (Fig 1 y 2.)¹.

Como puede observarse en la gráfica, la serie no es estacionaria, situación que complicaría el análisis si tratáramos de aplicar otra metodología, ya que tendríamos que transformar los datos hasta conseguir que la serie tuviera esa característica. En este caso, no es necesario transformar la serie debido a que el método no lo requiere.

Esta serie como sabemos, está influenciada por factores de índole económica y social, entre otros, razón por la cual la serie presenta cambios importantes en su comportamiento, que ocasionan dificultades al tratar de modelarla. Debido a la escasa información que tenemos de ella, trataremos de utilizar un modelo sencillo, pues de lo contrario, caeríamos en una falacia ya que ello implicaría un mayor conocimiento de la serie de tiempo.

El modelo que consideramos es el modelo *steady* expresado como modelo lineal dinámico, mismo que presentamos en la sección 5.5.3 de este trabajo.

Utilizamos este modelo ya que no sólo nos proporciona una representación sencilla de la serie, sino que además podemos pensar, por el conocimiento que tenemos, que

¹Se presentan al final de esta sección.

este modelo tiene las características adecuadas para predecir una serie de esta índole, sobretudo para periodos próximos, ya que a menos de que haya un cambio importante en el comportamiento económico del país como podría ser una devaluación de la moneda o un cambio en las tasas de interés del mercado, entre otros factores, que no tuvieramos previstos; podemos suponer que los valores de éstas variaciones no cambiarán bruscamente de un periodo a otro. Cabe señalar que si tenemos información sobre este tipo de situaciones, esa información la podemos incorporar al modelaje de la serie en cuestión.

El modelo considerado presenta la siguiente estructura:

$$\begin{aligned} y_t &= \mu_t + v_t \\ \mu_t &= \mu_{t-1} + w_t \quad t = 1, 2, 3, \dots \end{aligned}$$

donde $v_t \sim \mathcal{N}(0, V_t)$ y $w_t \sim \mathcal{N}(0, W_t)$. La distribución para μ_0 es una normal con media m_0 y varianza C_0 .

Para poder iniciar el proceso, necesitamos estimar los términos m_0 , C_0 , W_t y V_t . Con este objetivo en mente, procedemos a estimar los valores m_0 y C_0 . Para tal efecto consideremos los datos del primer semestre sobre el cual tenemos información, es decir, las cifras de enero a junio de 1980. La media y varianza de este conjunto de datos nos permitirá tener un conocimiento previo sobre el parámetro, conocimiento que se reflejará en una distribución inicial del mismo. La estimación puede ser un tanto burda, sin embargo, se espera que mejore a medida que incorporemos las siguientes observaciones al modelo, a causa de que tendremos más conocimiento.

Otros valores que es necesario conocer son las varianzas de los términos de perturbación. En este caso, lo adecuado sería construir estas varianzas en base al conocimiento que se tiene sobre el fenómeno; sin embargo, debido a que no tenemos conocimiento alguno sobre estos términos, para efectos del ejemplo, consideraremos que éstos se explican a través de una distribución normal estándar. Si obtuviéramos posteriormente información sobre algún fenómeno que afecte al valor que tome la variable en cuestión, esa nueva información podríamos reflejarla en estos términos de perturbación.

Como podrá apreciar el lector, un problema que presentan estos métodos es el hecho de que se necesitan conocer, en todo momento, las matrices de varianzas y covarianzas de los términos de perturbación de la ecuación del sistema y de la ecuación de observación, situación que no se aprecia fácil. Una posibilidad es estimar el valor de estas matrices; sin embargo es necesario realizar una nueva investigación en la que se analicen cuáles pueden ser las formas posibles de llevar a cabo la estimación de estas matrices, así como las consecuencias de introducirlas dentro del modelo.

En base a lo anterior, aplicamos el método que Harrison y Stevens han desarrollado para encontrar la distribución posterior de μ_t . En este caso, las relaciones de

Kalman se simplifican considerablemente ya que, no es difícil probar que si

$$\mu_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$$

entonces

$$\begin{pmatrix} \mu_t \\ y_t \end{pmatrix} \sim \mathcal{N} \left\{ \begin{pmatrix} m_{t-1} \\ m_{t-1} \end{pmatrix}, \begin{pmatrix} C_{t-1} + W_t & C_{t-1} + W_t \\ C_{t-1} + W_t & C_{t-1} + W_t + V_t \end{pmatrix} \right\}.$$

En consecuencia la distribución posterior para μ_t es

$$\mu_t \sim \mathcal{N}(m_t, C_t)$$

donde

$$m_t = m_{t-1} + \frac{C_{t-1} + W_t}{C_{t-1} + W_t + V_t} (y_t - m_{t-1})$$

y

$$C_t = C_{t-1} + W_t - \frac{(C_{t-1} + W_t)^2}{C_{t-1} + W_t + V_t}.$$

Para predecir los valores futuros de las observaciones utilizamos las expresiones siguientes:

$$\hat{m}_{t+k}^t = \hat{m}_{t+k-1}^t$$

por lo tanto

$$\hat{m}_{t+k}^t = m_t$$

y

$$\hat{C}_{t+k}^t = \hat{C}_{t+k-1}^t + W_t.$$

En consecuencia la media, \hat{y}_{t+k} , y la varianza, \hat{Y}_{t+k} , de las predicciones, tienen las expresiones siguientes:

$$\hat{y}_{t+k}^t = \hat{m}_{t+k-1}^t$$

de donde

$$\hat{y}_{t+k}^t = m_t$$

y

$$\hat{Y}_{t+k}^t = \hat{C}_{t+k-1}^t + W_t + V_t.$$

En base a las consideraciones que hemos hecho con anterioridad, podemos calcular ahora la distribución posterior del parámetro cada vez que se tenga disponible una nueva observación; con esta distribución podemos calcular la distribución de las predicciones.

A continuación presentamos una tabla de valores que nos permite visualizar cómo varía la media y la varianza del parámetro y cómo son la media y la varianza de las predicciones que se obtienen a partir de esa distribución posterior del parámetro. La tabla que se presenta a continuación, se obtuvo utilizando la hoja de cálculo *Lotus 1-2-3*.(Fig. 3).

Podemos observar que al sustituir el valor de V_t y W_t las expresiones se reducen ya que, en ese caso, la varianza del parámetro en el periodo $t + k$ se incrementará en k unidades respecto a la varianza del mismo en el periodo t , considerando que sólo se tiene información hasta ese periodo. De la misma manera ocurrirá con las predicciones, sólo que el incremento en ellas será mayor (de $2k$), debido a que se tiene mayor incertidumbre respecto de su valor.

Asimismo podemos darnos cuenta que la media del proceso en este caso será la misma, tanto para el parámetro como para las predicciones. No obstante, puede observarse la dinámica que presenta ésta a medida que se incorporan nuevas observaciones al proceso. Con el fin de que se pueda apreciar ese dinamismo se presenta la gráfica de las medias del proceso en cada periodo, calculadas a partir de la información disponible hasta un periodo atrás.(Fig. 4.)

INFLACION MENSUAL

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
ENERO	4.84%	3.22%	4.97%	10.88%	6.35%	7.42%	8.84%	8.10%	15.46%	2.45%
FEBRERO	2.31%	2.46%	3.93%	5.37%	5.28%	4.15%	4.45%	7.21%	8.34%	1.36%
MARZO	2.08%	2.14%	3.65%	4.84%	4.27%	3.88%	4.65%	6.81%	5.12%	1.08%
ABRIL	1.75%	2.25%	5.42%	6.33%	4.33%	3.08%	5.22%	8.75%	3.08%	1.50%
MAYO	1.63%	1.51%	5.62%	4.34%	3.32%	2.37%	5.56%	7.54%	1.93%	1.38%
JUNIO	1.98%	1.40%	4.82%	3.79%	3.62%	2.50%	6.42%	7.24%	2.04%	1.21%
JULIO	2.79%	1.76%	5.15%	4.94%	3.28%	3.48%	4.99%	8.10%	1.67%	
AGOSTO	2.07%	2.06%	11.22%	3.88%	2.84%	4.37%	7.97%	8.17%	0.92%	
SEPTIEMBRE	1.11%	1.86%	5.34%	3.08%	2.98%	3.99%	6.00%	6.59%	0.57%	
OCTUBRE	1.51%	2.22%	5.18%	3.32%	3.49%	3.80%	5.72%	8.33%	0.76%	
NOVIEMBRE	1.74%	1.93%	5.06%	5.87%	3.43%	4.61%	6.76%	7.93%	1.34%	
DICIEMBRE	2.62%	2.69%	10.68%	4.28%	4.25%	6.81%	7.89%	14.77%	2.09%	

Fig. 1

INFLACION MENSUAL
Enero 1980 - Junio 1989

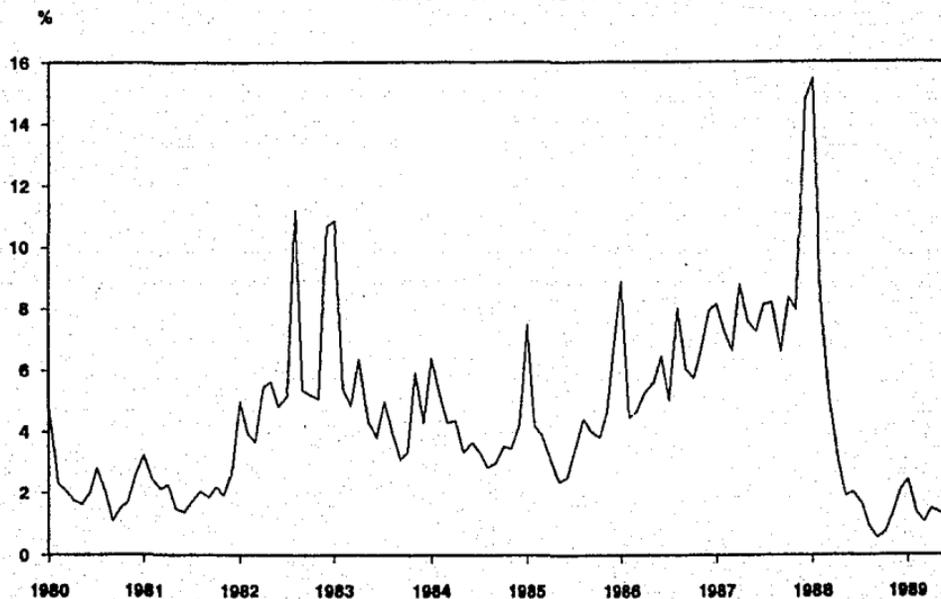


Fig. 2

Medias y varianzas del parámetro y de las predicciones.

MES	y_t
ENE 1980	4.84
FEB	2.31
MAR	2.06
ABR	1.75
MAY	1.63

Información Inicial

t	MES	y_t	\hat{m}_{t+K}^t	\hat{c}_t^t	\hat{c}_{t+1}^t	\hat{c}_{t+2}^t	\hat{c}_{t+3}^t	\hat{c}_{t+K}^t	\hat{v}_{t+1}^t	\hat{v}_{t+2}^t	\hat{v}_{t+3}^t
0			2.43	1.2107							
1	JUL 1980.	2.79	2.68	0.6885	1.6885	2.6885	3.6885	2.68	2.6885	3.6885	4.6885
2	AGO	2.07	2.30	0.6281	1.6281	2.6281	3.6281	2.30	2.6281	3.6281	4.6281
3	SEP	1.11	1.56	0.6195	1.6195	2.6195	3.6195	1.56	2.6195	3.6195	4.6195
4	OCT	1.51	1.53	0.6182	1.6182	2.6182	3.6182	1.53	2.6182	3.6182	4.6182
5	NOV	1.74	1.66	0.6181	1.6181	2.6181	3.6181	1.66	2.6181	3.6181	4.6181
6	DIC	2.62	2.25	0.6180	1.6180	2.6180	3.6180	2.25	2.6180	3.6180	4.6180
7	ENE 1981	3.22	2.85	0.6180	1.6180	2.6180	3.6180	2.85	2.6180	3.6180	4.6180
8	FEB	2.46	2.61	0.6180	1.6180	2.6180	3.6180	2.61	2.6180	3.6180	4.6180
9	MAR	2.14	2.32	0.6180	1.6180	2.6180	3.6180	2.32	2.6180	3.6180	4.6180
10	ABR	2.25	2.28	0.6180	1.6180	2.6180	3.6180	2.28	2.6180	3.6180	4.6180
11	MAY	1.51	1.80	0.6180	1.6180	2.6180	3.6180	1.80	2.6180	3.6180	4.6180
12	JUN	1.40	1.55	0.6180	1.6180	2.6180	3.6180	1.55	2.6180	3.6180	4.6180
13	JUL	1.76	1.68	0.6180	1.6180	2.6180	3.6180	1.68	2.6180	3.6180	4.6180
14	AGO	2.06	1.92	0.6180	1.6180	2.6180	3.6180	1.92	2.6180	3.6180	4.6180
15	SEP	1.86	1.88	0.6180	1.6180	2.6180	3.6180	1.88	2.6180	3.6180	4.6180
16	OCT	2.22	2.09	0.6180	1.6180	2.6180	3.6180	2.09	2.6180	3.6180	4.6180
17	NOV	1.93	1.99	0.6180	1.6180	2.6180	3.6180	1.99	2.6180	3.6180	4.6180
18	DIC	2.69	2.42	0.6180	1.6180	2.6180	3.6180	2.42	2.6180	3.6180	4.6180

Fig. 3

t	MES	y_t	\hat{m}_{t+k}^+	\hat{C}_t^+	\hat{C}_{t+1}^+	\hat{C}_{t+2}^+	\hat{C}_{t+3}^+	\hat{y}_{t+k}^+	\hat{y}_{t+k}^+	\hat{y}_{t+2}^+	\hat{y}_{t+3}^+
19	ENE 1982	4.97	4.00	0.6180	1.6180	2.6180	3.6180	4.00	2.6180	3.6180	4.6180
20	FEB	3.93	3.96	0.6180	1.6180	2.6180	3.6180	3.96	2.6180	3.6180	4.6180
21	MAR	3.65	3.77	0.6180	1.6180	2.6180	3.6180	3.77	2.6180	3.6180	4.6180
22	ABR	5.42	4.79	0.6180	1.6180	2.6180	3.6180	4.79	2.6180	3.6180	4.6180
23	MAY	5.62	5.30	0.6180	1.6180	2.6180	3.6180	5.30	2.6180	3.6180	4.6180
24	JUN	4.82	5.00	0.6180	1.6180	2.6180	3.6180	5.00	2.6180	3.6180	4.6180
25	JUL	5.15	5.09	0.6180	1.6180	2.6180	3.6180	5.09	2.6180	3.6180	4.6180
26	AGO	11.22	8.88	0.6180	1.6180	2.6180	3.6180	8.88	2.6180	3.6180	4.6180
27	SEP	5.34	6.69	0.6180	1.6180	2.6180	3.6180	6.69	2.6180	3.6180	4.6180
28	OCT	5.18	5.78	0.6180	1.6180	2.6180	3.6180	5.78	2.6180	3.6180	4.6180
29	NOV	5.06	5.33	0.6180	1.6180	2.6180	3.6180	5.33	2.6180	3.6180	4.6180
30	DIC	10.68	8.64	0.6180	1.6180	2.6180	3.6180	8.64	2.6180	3.6180	4.6180
31	ENE 1983	10.88	10.02	0.6180	1.6180	2.6180	3.6180	10.02	2.6180	3.6180	4.6180
32	FEB	5.37	7.15	0.6180	1.6180	2.6180	3.6180	7.15	2.6180	3.6180	4.6180
33	MAR	4.84	5.72	0.6180	1.6180	2.6180	3.6180	5.72	2.6180	3.6180	4.6180
34	ABR	6.33	6.10	0.6180	1.6180	2.6180	3.6180	6.10	2.6180	3.6180	4.6180
35	MAY	4.34	5.01	0.6180	1.6180	2.6180	3.6180	5.01	2.6180	3.6180	4.6180
36	JUN	3.79	4.26	0.6180	1.6180	2.6180	3.6180	4.26	2.6180	3.6180	4.6180
37	JUL	4.94	4.68	0.6180	1.6180	2.6180	3.6180	4.68	2.6180	3.6180	4.6180
38	AGO	3.88	4.19	0.6180	1.6180	2.6180	3.6180	4.19	2.6180	3.6180	4.6180
39	SEP	3.08	3.50	0.6180	1.6180	2.6180	3.6180	3.50	2.6180	3.6180	4.6180
40	OCT	3.32	3.39	0.6180	1.6180	2.6180	3.6180	3.39	2.6180	3.6180	4.6180
41	NOV	5.87	4.92	0.6180	1.6180	2.6180	3.6180	4.92	2.6180	3.6180	4.6180
42	DIC	4.28	4.53	0.6180	1.6180	2.6180	3.6180	4.53	2.6180	3.6180	4.6180

Fig. 3 (Continuación).

t	MES	y_t	\hat{m}_{t+L}^+	\hat{c}_t^+	\hat{c}_{t+1}^+	\hat{c}_{t+L}^+	\hat{c}_{t+8}^+	\hat{v}_{t+k}^+	\hat{v}_{t+1}^+	\hat{v}_{t+2}^+	\hat{v}_{t+3}^+
43	ENE 1984	6.35	5.65	0.6180	1.6180	2.6180	3.6180	5.65	2.6180	3.6180	4.6180
44	FEB	5.28	5.42	0.6180	1.6180	2.6180	3.6180	5.42	2.6180	3.6180	4.6180
45	MAR	4.27	4.71	0.6180	1.6180	2.6180	3.6180	4.71	2.6180	3.6180	4.6180
46	ABR	4.33	4.48	0.6180	1.6180	2.6180	3.6180	4.48	2.6180	3.6180	4.6180
47	MAY	3.32	3.76	0.6180	1.6180	2.6180	3.6180	3.76	2.6180	3.6180	4.6180
48	JUN	3.62	3.67	0.6180	1.6180	2.6180	3.6180	3.67	2.6180	3.6180	4.6180
49	JUL	3.28	3.43	0.6180	1.6180	2.6180	3.6180	3.43	2.6180	3.6180	4.6180
50	AGO	2.84	3.07	0.6180	1.6180	2.6180	3.6180	3.07	2.6180	3.6180	4.6180
51	SEP	2.98	3.01	0.6180	1.6180	2.6180	3.6180	3.01	2.6180	3.6180	4.6180
52	OCT	3.49	3.31	0.6180	1.6180	2.6180	3.6180	3.31	2.6180	3.6180	4.6180
53	NOV	3.43	3.38	0.6180	1.6180	2.6180	3.6180	3.38	2.6180	3.6180	4.6180
54	DIC	4.25	3.92	0.6180	1.6180	2.6180	3.6180	3.92	2.6180	3.6180	4.6180
55	ENE 1985	7.42	6.08	0.6180	1.6180	2.6180	3.6180	6.08	2.6180	3.6180	4.6180
56	FEB	4.15	4.89	0.6180	1.6180	2.6180	3.6180	4.89	2.6180	3.6180	4.6180
57	MAR	3.88	4.27	0.6180	1.6180	2.6180	3.6180	4.27	2.6180	3.6180	4.6180
58	ABR	3.08	3.53	0.6180	1.6180	2.6180	3.6180	3.53	2.6180	3.6180	4.6180
59	MAY	2.37	2.81	0.6180	1.6180	2.6180	3.6180	2.81	2.6180	3.6180	4.6180
60	JUN	2.50	2.62	0.6180	1.6180	2.6180	3.6180	2.62	2.6180	3.6180	4.6180
61	JUL	3.48	3.15	0.6180	1.6180	2.6180	3.6180	3.15	2.6180	3.6180	4.6180
62	AGO	4.37	3.90	0.6180	1.6180	2.6180	3.6180	3.90	2.6180	3.6180	4.6180
63	SEP	3.99	3.96	0.6180	1.6180	2.6180	3.6180	3.96	2.6180	3.6180	4.6180
64	OCT	3.80	3.86	0.6180	1.6180	2.6180	3.6180	3.86	2.6180	3.6180	4.6180
65	NOV	4.61	4.32	0.6180	1.6180	2.6180	3.6180	4.32	2.6180	3.6180	4.6180
66	DIC	6.81	5.86	0.6180	1.6180	2.6180	3.6180	5.86	2.6180	3.6180	4.6180

Fig. 3 (Continuación).

t	MES	y_t	\hat{m}_{t+k}^t	\hat{c}_t^t	\hat{c}_{t+1}^t	\hat{c}_{t+2}^t	\hat{c}_{t+3}^t	\hat{y}_{t+k}^t	\hat{y}_{t+1}^t	\hat{y}_{t+2}^t	\hat{y}_{t+3}^t
67	ENE 1986	8.84	7.70	0.6180	1.6180	2.6180	3.6180	7.70	2.6180	3.6180	4.6180
68	FEB	4.45	5.69	0.6180	1.6180	2.6180	3.6180	5.69	2.6180	3.6180	4.6180
69	MAR	4.65	5.05	0.6180	1.6180	2.6180	3.6180	5.05	2.6180	3.6180	4.6180
70	ABR	5.22	5.15	0.6180	1.6180	2.6180	3.6180	5.15	2.6180	3.6180	4.6180
71	MAY	5.56	5.41	0.6180	1.6180	2.6180	3.6180	5.41	2.6180	3.6180	4.6180
72	JUN	6.42	6.03	0.6180	1.6180	2.6180	3.6180	6.03	2.6180	3.6180	4.6180
73	JUL	4.99	5.39	0.6180	1.6180	2.6180	3.6180	5.39	2.6180	3.6180	4.6180
74	AGO	7.97	6.98	0.6180	1.6180	2.6180	3.6180	6.98	2.6180	3.6180	4.6180
75	SEP	6.00	6.38	0.6180	1.6180	2.6180	3.6180	6.38	2.6180	3.6180	4.6180
76	OCT	5.72	5.97	0.6180	1.6180	2.6180	3.6180	5.97	2.6180	3.6180	4.6180
77	NOV	6.76	6.46	0.6180	1.6180	2.6180	3.6180	6.46	2.6180	3.6180	4.6180
78	DIC	7.89	7.34	0.6180	1.6180	2.6180	3.6180	7.34	2.6180	3.6180	4.6180
79	ENE 1987	8.10	7.81	0.6180	1.6180	2.6180	3.6180	7.81	2.6180	3.6180	4.6180
80	FEB	7.21	7.44	0.6180	1.6180	2.6180	3.6180	7.44	2.6180	3.6180	4.6180
81	MAR	6.61	6.93	0.6180	1.6180	2.6180	3.6180	6.93	2.6180	3.6180	4.6180
82	ABR	8.75	8.05	0.6180	1.6180	2.6180	3.6180	8.05	2.6180	3.6180	4.6180
83	MAY	7.54	7.74	0.6180	1.6180	2.6180	3.6180	7.74	2.6180	3.6180	4.6180
84	JUN	7.24	7.43	0.6180	1.6180	2.6180	3.6180	7.43	2.6180	3.6180	4.6180
85	JUL	8.10	7.84	0.6180	1.6180	2.6180	3.6180	7.84	2.6180	3.6180	4.6180
86	AGO	8.17	8.05	0.6180	1.6180	2.6180	3.6180	8.05	2.6180	3.6180	4.6180
87	SEP	6.59	7.15	0.6180	1.6180	2.6180	3.6180	7.15	2.6180	3.6180	4.6180
88	OCT	8.33	7.88	0.6180	1.6180	2.6180	3.6180	7.88	2.6180	3.6180	4.6180
89	NOV	7.93	7.91	0.6180	1.6180	2.6180	3.6180	7.91	2.6180	3.6180	4.6180
90	DIC	14.77	12.15	0.6180	1.6180	2.6180	3.6180	12.15	2.6180	3.6180	4.6180

Fig. 3 (Continuación).

t	MES	y_t	\hat{m}_t^A	\hat{c}_t^A	\hat{c}_t^B	\hat{c}_t^C	\hat{c}_t^D	\hat{c}_t^E	\hat{c}_t^F	\hat{c}_t^G	\hat{c}_t^H
91	ENE 1968	15.46	14.20	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
92	FEB	8.34	10.58	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
93	MAR	5.12	7.20	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
94	ABR	3.08	4.66	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
95	MAY	1.93	2.97	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
96	JUN	2.04	2.40	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
97	JUL	1.67	1.95	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
98	AGO	0.92	1.31	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
99	SEP	0.57	0.85	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
100	OCT	0.76	0.80	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
101	NOV	1.34	1.13	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
102	DIC	2.09	1.72	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
103	ENE 1969	2.45	2.17	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
104	FEB	1.36	1.67	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
105	MAR	1.08	1.31	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
106	ABR	1.50	1.43	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
107	MAY	1.38	1.40	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180
108	JUN	1.21	1.28	0.6180	1.6180	2.6180	3.6180	4.6180	5.6180	6.6180	7.6180

Fig. 3 (Continuación).

INFLACION MENSUAL
Evolución de la media del parámetro
Julio 1980 - Junio 1989

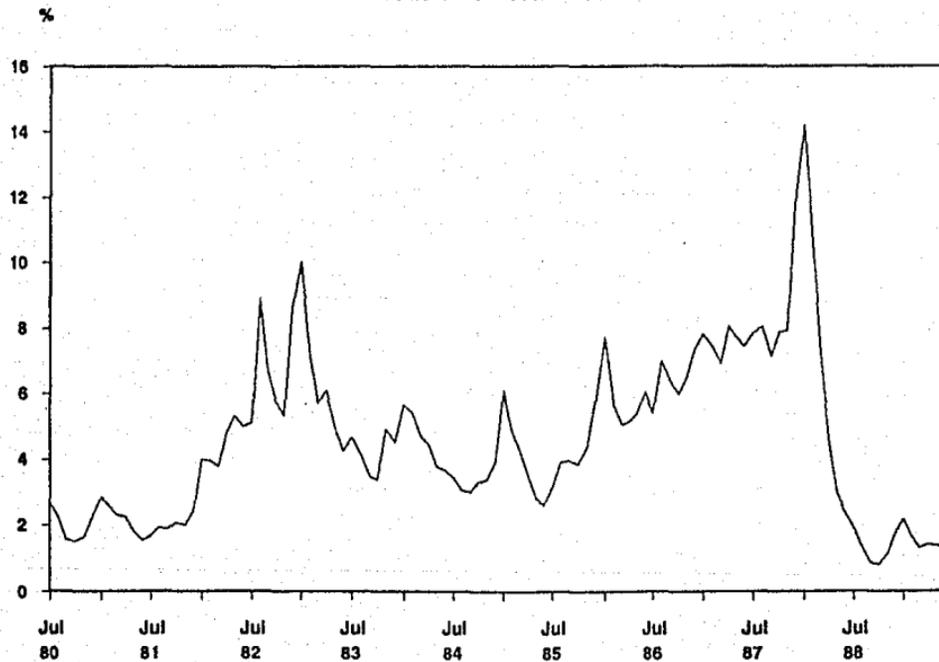


Fig. 4

CONCLUSIONES

Cuando se cuenta con *información* acerca de una variable que nos interesa, resulta atractivo tratar de entender el comportamiento que ha venido mostrando ésta; un problema que resulta interesante, es tratar de averiguar el comportamiento de esa variable en otros periodos de tiempo sobre los cuales no tenemos ningún tipo de información. Nos interesa predecir, posiblemente, porque algunas de las decisiones que vayamos a tomar o que tomarán las personas para las cuales hacemos el estudio, estarán fuertemente relacionadas con aquellas condiciones que tratamos de predecir en tales periodos.

Dentro de los métodos estadísticos que se han propuesto solucionar este problema, encontramos a los *Modelos Lineales Dinámicos*, los cuales se han desarrollado dentro del Enfoque Bayesiano. Cómo hemos visto, estos modelos utilizan las relaciones recursivas de Kalman, las cuales les permiten actualizar probabilidades.

Desde el punto de vista de Sullivan y Claycombe, que compartimos, ninguna predicción debe aceptarse como final sino que debe revisarse constantemente ya que, cada nueva pieza de información contribuirá a la confiabilidad o desconfianza de las hipótesis usadas. En este sentido, los modelos lineales dinámicos proveen la estructura necesaria para llevar a cabo esa revisión de predicciones.

Una ventaja que, a nuestro juicio, también presenta este método es que las observaciones van incorporándose *una a una*. Consideramos que esta situación es positiva por varias causas. En primer lugar, existen ventajas computacionales pues no se manejan todos los datos a la vez, al realizar los cálculos, permitiendo que algunos métodos sean fácilmente implementados en estos términos.

Otra posibilidad que abren estos métodos son, desde luego, aquéllos casos en los cuales se tiene *poca* información de índole numérica. En estas situaciones el interesado podrá actualizar sus predicciones e ir conociendo al fenómeno a medida que nuevas piezas de información vaya obteniendo. Su conocimiento sobre fenómenos semejantes jugará un papel importante en la determinación de las probabilidades iniciales, así como también, el conocimiento que tenga sobre factores que influyan en el comportamiento del fenómeno que le interesa. Esto último también es aplicable a aquellas situaciones en las cuales se tiene *más* información.

Respecto a la condición de estacionaridad que varios modelos requieren como supuesto inicial, éste método no hace hincapié en esa situación; haciendo flexible el uso del método para series que no tienen esa característica. Aún más, el método es tal, que

no proporciona un método estándar para modelar a las series, ya que como hemos visto, varios modelos estadísticos pueden llegar a formularse como modelos lineales dinámicos.

El estadístico tratará de modelar el conocimiento que el interesado tiene sobre el fenómeno, eligiendo el método que a su juicio, sea el más apropiado para reflejarlo. Los modelos lineales dinámicos ofrecen mayores posibilidades de modelaje, situación que los pone en un plano competitivo respecto de los demás modelos.

Como pudimos observar en el Capítulo V, varios modelos estadísticos pueden establecerse como Modelos Lineales Dinámicos. En el caso de los Modelos $AR(p)$ y $MA(q)$, Harrison y Stevens explican cómo pueden expresarse éstos en términos de los modelos lineales dinámicos; sin embargo, debe observarse, tal como lo hace Priestley¹, que los modelos $AR(p)$ tienen problemas en cuanto a la expresión para la distribución de las predicciones, debido a que se debe considerar la dependencia estocástica de la matriz F_{t+k} y la observación y_{t+k} ; de modo que la propuesta que hacen Harrison y Stevens de utilizar una estimación de la matriz F_t , debe ser motivo de una más amplia investigación.

En nuestra opinión, una restricción del método es determinar cuáles serán las matrices V_{t+k} y W_{t+k} . El modelo lineal dinámico que Harrison y Stevens plantean no deja vislumbrar de qué manera pueden construirse esas matrices, ni cómo estimarlas, salvo que se construyan a través del conocimiento del interesado sobre el fenómeno, situación que parece bastante compleja e inverosímil.

¹Véase la discusión del artículo de Harrison y Stevens (1978).

APENDICE

A.1 El Teorema de Bayes¹

Sea $X = (X_1, X_2, \dots, X_m)^T$ y $Y = (Y_1, Y_2, \dots, Y_n)^T$ dos vectores aleatorios. Sean g_1 y g_2 las f.d.p.g.² marginales de X y Y respectivamente. Sean $h_1(\cdot | y)$ la f.d.p.g. condicional de X dado el valor de $Y = y$ y $h_2(\cdot | x)$ la f.d.p.g. condicional de Y dado el valor de $X = x$. Entonces para cualesquier punto $x \in R^m$ tal que $g_1(x) > 0$ y cualquier $y \in R^n$

$$h_2(y | x) = \frac{h_1(x | y)g_2(y)}{\int_{R^n} h_1(x | t)g_2(t) d\mu(t)}.$$

En términos de eventos el Teorema de Bayes se establece como sigue:

Sean A_1, A_2, \dots una sucesión de eventos disjuntos para los cuales $\bigcup_{i=1}^{\infty} A_i = S$ donde S es el espacio muestral, i.e. el conjunto de todos los posibles resultados de un experimento y $\Pr(A_i) > 0$ para $i = 1, 2, \dots$. Además, sea B otro evento tal que $\Pr(B) > 0$. Entonces:

$$\Pr(A_i | B) = \frac{\Pr(B | A_i) \Pr(A_i)}{\sum_{j=1}^{\infty} \Pr(B | A_j) \Pr(A_j)} \quad i = 1, 2, 3, \dots$$

¹ Véase DeGroot (1970; 11-12).

² Funciones de densidad de probabilidad generalizadas.

A.2 Definiciones y Teoremas relevantes sobre probabilidades y esperanzas condicionales

En esta sección se muestran algunos teoremas y definiciones que pueden ser de interés para personas poco familiarizadas con distribuciones y esperanzas condicionales³.

Definición 1. Dados dos eventos A y B , definimos a la *función de probabilidad condicional* $\Pr(A | B)$, del evento A dado el evento B , por:

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)}; \quad \Pr(B) > 0$$

la $\Pr(A | B)$ no está definida cuando $\Pr(B) = 0$. A *grosso modo* se puede decir que $\Pr(A | B)$ es la probabilidad del evento A habiendo observado la ocurrencia del evento B , por esta razón es poco importante que la $\Pr(A | B)$ no esté definida cuando la $\Pr(B) = 0$.

Definición 2. La *función de densidad condicional* $p_{X|Y}(x | y)$, de x dado $Y = y$ para todo x y y tal que $p_Y(y) > 0$, está dada por:

$$p_{X|Y}(x | y) := \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\int_{\mathbb{R}} p_{X,Y}(s, y) ds}$$

Para el caso multivariado tenemos lo siguiente:

$$\begin{aligned} & p_{X_1, X_2, \dots, X_n | Y_1, Y_2, \dots, Y_m}(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_m) \\ & := \frac{p_{X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m}(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)}{p_{Y_1, Y_2, \dots, Y_m}(y_1, y_2, \dots, y_m)} \\ & = \frac{p_{X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m}(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)}{\int \dots \int_{\mathbb{R}^k} p_{X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m}(s_1, s_2, \dots, s_n, y_1, y_2, \dots, y_m) ds_1, ds_2, \dots, ds_n} \end{aligned}$$

Definición 3. La *esperanza condicional*⁴ de la variable aleatoria X dada la variable aleatoria $Y = y$, se define como:

$$E_{X|Y}[X | y] := \int x p_{X|Y}(x | y) dx$$

cuando X es un vector aleatorio de dimensión n :

$$E_{X|Y}[X | y] := \{E_{X_1|Y}[X_1 | y], \dots, E_{X_n|Y}[X_n | y]\}^T.$$

³ Véase Hogg y Craig (1959; 90-94)

⁴ A veces llamada también media condicional.

Teorema 1. (Función de Densidad Condicional). Sean X y Y variables aleatorias. La función de densidad condicional $p_{X|Y}(x | y)$, satisface lo siguiente:

$$\begin{aligned}
 & p_{X|Y}(x | y) \geq 0 \\
 & \int p_{X|Y}(x | y) dx = 1 \\
 & p_{X|Y}(x | y) = p_X(x) \quad \text{si } X \text{ y } Y \text{ son independientes.}
 \end{aligned}$$

Teorema 2. (Esperanzas condicionales). Sean X , Y y Z variables aleatorias; c y d constantes y $g(\cdot)$ una función de valor real. Supongamos que las esperanzas no condicionales de X , Y y Z existen. Entonces

$$\begin{aligned}
 E_{X|Y}[X | y] &= E_X[X] \quad \text{si } X \text{ y } Y \text{ son independientes} \\
 E_X[X] &= E_Y\{E_{X|Y}[X | y]\} \\
 E_{X|Y}[g(y)X | y] &= g(y)E_{X|Y}[X | y] \\
 E_X[g(y)X] &= E_Y\{g(y)E_{X|Y}[X | y]\} \\
 E_{X|Y}[c | y] &= c \\
 E_{X|Y}[g(y) | y] &= g(y) \\
 E_{X,Z|Y}[cX + dZ | y] &= c E_{X|Y}[X | y] + d E_{Z|Y}[Z | y].
 \end{aligned}$$

A.3 Tres Teoremas Importantes

En este apéndice se presentan los tres teoremas que nos permiten obtener los resultados sobre Filtros de Kalman que aparecen en el Capítulo V. La demostración de estos resultados puede verse en DeGroot(1970; 51-56).

Definición. Sean X_1, X_2, \dots, X_n n variables aleatorias. X_1, X_2, \dots, X_n se distribuyen conjuntamente como una *normal multivariada* si su función de distribución de probabilidad tiene la siguiente forma:

$$p_X(x) = (2\pi)^{n/2} |\Sigma_X|^{-1/2} \exp\{-\frac{1}{2}(x - E[X])^T \Sigma_X^{-1}(x - E[X])\}$$

$|\Sigma_X| = \det \Sigma_X$ y el supraíndice -1 denota a la matriz inversa.

Teorema I. Sea X una variable *normal* de dimensión n con media μ y covarianza Σ y sea $Y = BX + \nu$, donde B es una matriz de dimensión $m \times n$ y ν es un vector columna de dimensión m , entonces Y es una variable aleatoria *normal* de dimensión $m \times 1$ con media $BX + \nu$ y covarianza $B\Sigma B^T$.

Teorema II. Sea X una variable aleatoria *normal* de dimensión n con media μ y covarianza Σ . Si X , μ y Σ se particionan de la siguiente manera:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

entonces X_1 es una variable aleatoria normal con media μ_1 y covarianza Σ_{11} .

Teorema III. Sea X una variable *normal* de dimensión n con media μ y covarianza Σ . Si X , μ y Σ se particionan como sigue:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

entonces la distribución condicional de X_1 dada $X_2 = x_2$ es *normal* con media $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ y covarianza $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

A.4 El Filtro de Kalman

El esquema de Jazwinski. Demostración.

En el Capítulo 5 hacemos un estudio de los Filtro de Kalman; en este apartado demostramos la versión que da Jazwinski⁵ para presentar al Filtro de Kalman.

El sistema se describe mediante las siguientes expresiones:

$$x_{k+1} = \Phi(t_{k+1}, t_k)x_k + \Gamma(t_k)w_{k+1}, \quad k = 0, 1, \dots$$

donde x_k es el estado del sistema al tiempo t_k , de dimensión $n \times 1$, Φ es la matriz de transición del estado de dimensión $n \times n$ no-singular, Γ es una matriz de $n \times r$, y $\{w_k, k = 1, 2, \dots\}$ es una sucesión de ruido, cada w_k es un vector de dimensión r tal que $w_k \sim \mathcal{N}(0, Q_k)$. Las observaciones están dadas por:

$$y_k = M(t_k)x_k + v_k$$

x_0 , $\{w_k\}$ y $\{v_k\}$ se suponen independientes. Supongamos que $v_k \sim \mathcal{N}(0, R_k)$ ⁶
Denotemos por:

$$Y_r := \{y_1, \dots, y_r\}$$

$$\hat{x}_s^r := E[x_s | Y_r]$$

y

$$P_s^r := E\{[(x_s - \hat{x}_s^r)(x_s - \hat{x}_s^r)^T] | Y_r\}$$

entonces, el filtro lineal óptimo (de varianza mínima) tiene como media condicional y matriz de covarianzas a:

$$\begin{aligned}\hat{x}_{k+1}^k &= \Phi(t_{k+1}, t_k)\hat{x}_k^k \\ P_{k+1}^k &= E\{[(\Phi(x_k - \hat{x}_k^k) + \Gamma(t_k)w_{k+1})(\Phi(x_k - \hat{x}_k^k) + \Gamma(t_k)w_{k+1})^T] | Y_k\} \\ &= \Phi(t_{k+1}, t_k)P_k^k\Phi^T(t_{k+1}, t_k) + \Gamma(t_k)Q_{k+1}\Gamma^T(t_k).\end{aligned}$$

Además:

$$\begin{aligned}\hat{x}_k^k &= \hat{x}_k^{k-1} + K(t_k)(y_k - M(t_k)\hat{x}_k^{k-1}), \\ P_k^k &= P_k^{k-1} - K(t_k)M(t_k)P_k^{k-1},\end{aligned}$$

donde

$$K(t_k) = P_k^{k-1}M^T(t_k)[M(t_k)P_k^{k-1}M^T(t_k) + R_k]^{-1}$$

⁵Jazwinski(1970) págs. 200-201.

⁶Jazwinski no hace mención de este supuesto, sin embargo lo utiliza al obtener las expresiones para la media y la varianza del estado siguiente.

es la ganancia de Kalman. El esquema de predicción se complementa con las condiciones iniciales (\hat{x}_k^k, P_k^k) .

Demostración. Supondremos que la distribución condicional de x_k dadas $y_{k-1}, y_{k-2}, \dots, y_1$ es normal con media \hat{x}_k^{k-1} y matriz de varianzas y covarianzas P_k^{k-1} , para determinar a través de ella, la distribución condicional de x_{k+1} dadas y_k, y_{k-1}, \dots, y_1 .

Observemos que la distribución conjunta para x_{k+1} y y_k dadas las observaciones y_{k-1}, \dots, y_1 es normal con media

$$\begin{pmatrix} \Phi(t_{k+1}, t_k) \hat{x}_k^{k-1} \\ M(t_k) \hat{x}_k^{k-1} \end{pmatrix}$$

y matriz de varianzas y covarianzas

$$\left(\begin{array}{c|c} \Phi P_k^{k-1} \Phi^T + \Gamma Q_{k+1} \Gamma^T & \Phi P_k^{k-1} M^T \\ \hline \text{---} & \text{---} \\ \text{---} & M P_k^{k-1} M^T + R_k \end{array} \right).$$

Por lo tanto si consideramos la distribución condicional de x_{k+1} dadas y_k, y_{k-1}, \dots, y_1 obtenemos una distribución normal con media \hat{x}_{k+1}^k y matriz de varianzas y covarianzas P_{k+1}^k , donde

$$\begin{aligned} \hat{x}_{k+1}^k &= \Phi \hat{x}_k^{k-1} + \Phi P_k^{k-1} M^T (M P_k^{k-1} M^T + R_k)^{-1} (y_k - M \hat{x}_k^{k-1}) \\ &= \Phi \{ \hat{x}_k^{k-1} + P_k^{k-1} M^T (M P_k^{k-1} M^T + R_k)^{-1} (y_k - M \hat{x}_k^{k-1}) \} \end{aligned}$$

por otra parte

$$\begin{aligned} P_{k+1}^k &= \Phi P_k^{k-1} \Phi^T + \Gamma Q_{k+1} \Gamma^T - \Phi P_k^{k-1} M^T (M P_k^{k-1} M^T + R_k)^{-1} M P_k^{k-1} \Phi^T \\ &= \Phi \{ P_k^{k-1} - P_k^{k-1} M^T (M P_k^{k-1} M^T + R_k)^{-1} M P_k^{k-1} \} \Phi^T + \Gamma Q_{k+1} \Gamma^T \end{aligned}$$

Haciendo uso de las definiciones de \hat{x}_k^k y P_k^k , podemos notar que:

$$\begin{aligned} \hat{x}_{k+1}^k &= E\{\Phi(t_{k+1}, t_k) x_k + \Gamma(t_k) w_{k+1} \mid Y_k\} \\ &= \Phi(t_{k+1}, t_k) \hat{x}_k^k. \end{aligned}$$

Asimismo podemos observar que

$$\begin{aligned} P_{k+1}^k &= E\{[(\Phi(x_k - \hat{x}_k^k) + \Gamma w_{k+1})(\Phi(x_k - \hat{x}_k^k) + \Gamma w_{k+1})^T] \mid Y_k\} \\ &= \Phi(t_{k+1}, t_k) P_k^k \Phi^T(t_{k+1}, t_k) + \Gamma(t_k) Q_{k+1} \Gamma^T(t_k). \end{aligned}$$

Por lo tanto obtenemos que

$$\begin{aligned} \hat{x}_k^k &= \hat{x}_k^{k-1} + K(t_k) [y_k - M \hat{x}_k^{k-1}] \\ P_k^k &= P_k^{k-1} - K(t_k) M P_k^{k-1} \end{aligned}$$

y

$$K(t_k) = P_k^{k-1} M^T (M P_k^{k-1} M^T + R_k)^{-1}$$

Con lo cual queda probado el resultado. •

Apéndice A.5 El Modelo Lineal Dinámico

Demstración de las relaciones de Kalman en la versión de Harrison y Stevens.

Considérese nuevamente, el modelo presentado en la sección 5.2, el cual se expresa en los siguientes términos:

$$y_t = F_t \theta_t + v_t \quad (1)$$

$$\theta_t = G \theta_{t-1} + w_t \quad (2)$$

con

$$v_t \sim \mathcal{N}(0, V_t), w_t \sim \mathcal{N}(0, W_t) \text{ y } t = 1, 2, 3, \dots$$

Supongamos que la información concerniente al parámetro θ_{t_0} puede modelarse a través de una distribución normal con media m_0 y matriz de covarianzas C_0 ; supongamos también que la primera observación de que disponemos es y_{t_1} .

Bajo las hipótesis anteriores, la distribución posterior para θ_t es:

$$\theta_t \mid y_t, y_{t-1}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1} \sim \mathcal{N}(m_t, C_t)$$

Los valores m_t y C_t se obtienen de la siguiente manera:

$$m_t = G m_{t-1} + A e$$

$$C_t = R - A \hat{Y}_t A^T$$

donde

$$A = R F_t^T \hat{Y}_t^{-1}$$

$$\hat{Y}_t = F_t R F_t^T + V_t$$

$$R = G C_{t-1} G^T + W_t$$

$$e = y_t - \hat{y}_t$$

$$\hat{y}_t = F_t G m_{t-1}$$

Demstración. Supongamos que el resultado es válido para θ_{t-1} , es decir,

$$\theta_{t-1} \mid y_{t-1}, y_{t-2}, \dots, y_{t_1}, F_{t-1}, F_{t-2}, \dots, F_{t_1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$$

Ahora, trataremos de conocer cuál es la distribución condicional de $\theta_t \mid y_{t-1}, y_{t-2}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1}$, y en base a ella calcularemos posteriormente, la distribución de $\theta_t \mid y_t, y_{t-1}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1}$. Observemos que,

$$\theta_{t-1} \mid y_{t-1}, y_{t-2}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$$

y que además, al sustituir la ecuación (2) en la ecuación (1):

$$y_t = F_t [G \theta_{t-1} + w_t] + v_t \quad (3)$$

De aquí en adelante seguiremos la siguiente notación:

$$\begin{aligned} y_{T-1} &:= \{y_{t-1}, y_{t-2}, \dots, y_{t_1}\} \\ F_T &:= \{F_t, F_{t-1}, \dots, F_{t_1}\} \\ \hat{y} &:= E\{y_t \mid y_{T-1}, F_T\} \\ \hat{Y} &:= E\{(y_t - \hat{y}) \mid y_{T-1}, F_T\} \end{aligned}$$

En base a lo anterior, usando la ecuación (3), obtenemos que

$$\begin{aligned} \hat{y} &= E\{F_t[G\theta_{t-1} + w_t] + v_t \mid y_{T-1}, F_T\} \\ &= F_t G m_{t-1} \end{aligned}$$

y en consecuencia

$$\begin{aligned} \hat{Y} &= E\{[F_t(G\theta_{t-1} + w_t) + v_t - F_t G m_{t-1}] [F_t(G\theta_{t-1} + w_t) + v_t - F_t G m_{t-1}]^T \mid y_{T-1}, F_T\} \\ &= E\{[F_t G(\theta_{t-1} - m_{t-1}) + F_t w_t + v_t] [F_t G(\theta_{t-1} - m_{t-1}) + F_t w_t + v_t]^T \mid y_{T-1}, F_T\} \\ &= F_t G C_{t-1} G^T F_t^T + F_t W_t F_t^T + V_t \\ &= F_t [G C_{t-1} G^T + W_t] F_t^T + V_t. \end{aligned}$$

Si $R := G C_{t-1} G^T + W_t$, entonces $\hat{Y} = F_t R F_t^T + V_t$. De este modo, hemos obtenido la distribución para $y_t \mid y_{T-1}, F_T$.

Calculemos ahora la distribución conjunta⁷ de $\theta_t \mid y_{T-1}, F_T$ y $y_t \mid y_{T-1}, F_T$; para ello, necesitamos obtener el valor de la matriz de varianzas y covarianzas de θ_t y y_t dadas y_{T-1} y F_T . Denotemos por $\bar{\theta}$ a la esperanza de θ_t condicionada sobre y_{T-1} y F_T . De ahí que:

$$\begin{aligned} E\{(y_t - \hat{y})[\theta - \bar{\theta}]^T \mid y_{T-1}, F_T\} &= E\{[F_t G(\theta_{t-1} - m_{t-1}) + F_t w_t + v_t][G\theta_{t-1} + w_t - G m_{t-1}]^T \mid y_{T-1}, F_T\} \\ &= E\{[F_t G(\theta_{t-1} - m_{t-1}) + F_t w_t + v_t][(\theta_{t-1} - m_{t-1})^T G^T + w_t^T \mid y_{T-1}, F_T\} \\ &= F_t G C_{t-1} G^T + F_t W_t. \end{aligned}$$

Por lo tanto, la matriz de covarianzas entre θ_t y y_t dadas y_{T-1} y F_T es:

$$\left(\begin{array}{c|c} G C_{t-1} G^T + W_t & G C_{t-1} G^T F_t^T + W_t F_t^T \\ \hline F_t G C_{t-1} G^T + F_t W_t & F_t R F_t^T + V_t \end{array} \right)$$

Utilizando los resultados que hemos obtenido hasta ahora, podemos calcular la distribución condicional⁸ de $\theta_t \mid y_T, F_T$.

$$\theta_t \mid y_t, y_{t-1}, \dots, y_{t_1}, F_t, F_{t-1}, \dots, F_{t_1} \sim \mathcal{N}(m_t, C_t)$$

⁷Véanse los Teoremas I y II en el apéndice A.3.

⁸Usando el Teorema III del apéndice A.3.

Siendo

$$\begin{aligned}m_t &= Gm_{t-1} + \{GC_{t-1}G^T + W_t\}F_t^T[F_tRF_t^T + V_t]^{-1}(y_t - F_tGm_{t-1}) \\ &= Gm_{t-1} + RF_t^T\hat{Y}^{-1}[y_t - \hat{y}] \\ &= Gm_{t-1} + Ae\end{aligned}$$

donde

$$e = y_t - \hat{y}$$

y

$$A = RF_t^T\hat{Y}^{-1}.$$

Finalmente

$$\begin{aligned}C_t &= GC_{t-1}G^T + W_t - \{GC_{t-1}G^T + W_t\}F_t^T[F_tRF_t^T + V_t]^{-1}\{F_tGC_{t-1}G^T + F_tW_t\} \\ &= R - RF_t^T\hat{Y}^{-1}F_tR \\ &= R - RF_t^T\hat{Y}^{-1}\hat{Y}\hat{Y}^{-1}(RF_t)^T \\ &= R - A\hat{Y}A^T.\end{aligned}$$

Por lo tanto

$$\theta_t | y_t, y_{t-1}, \dots, y_1, F_t, F_{t-1}, \dots, F_1 \sim N(m_t, C_t)$$

con

$$\begin{aligned}m_t &= Gm_{t-1} + Ae \\ C_t &= R - A\hat{Y}_tA^T.\end{aligned}$$

De esta manera, hemos probado las relaciones de Kalman que proporcionan Harrison y Stevens para el Modelo Lineal Dinámico.

BIBLIOGRAFIA

INTRODUCCION A SERIES DE TIEMPO DESDE UN PUNTO DE VISTA BAYESIANO

Alegría, Alejandro
Teoría - Matemático, 1988.
UNAM.

TIME SERIES AND FORECASTING: AN APPLIED APPROACH

Bowerman, Bruce L.
O'Connell, Richard T.
Duxbury Press. North Scituate, 1979.

TIME SERIES ANALYSIS: FORECASTING AND CONTROL

Box, George E. P.
Jenkins, Gwilym M.
Holden-day. San Francisco, 1970.

BAYESIAN ANALYSIS OF LINEAR MODELS

Bromeling, Lyle D.
Marcel Dekker. New York, 1985.
Statistics: Textbooks and Monographs Vol. 60.

INTRODUCTION TO STATISTICAL TIME SERIES

Fuller, W. A.
John Wiley. New York 1976.

OPTIMAL STATISTICAL DECISIONS

DeGroot, Morris H.
McGraw-Hill. New York, 1970.

ADAPTATIVE FILTERING, PREDICTION AND CONTROL

Goodwin, Graham Clifford
Sin, Kwai Sang
Prentice-Hall, Englewood Cliffs, 1984.

BAYESIAN FORECASTING

Harrison, P. J.
Stevens, C. F.
Journal of Royal Statistical Society, B, Vol. 38. pp. 205-247.

INTRODUCTION TO MATHEMATICAL STATISTICS

Hogg, Robert V.
Craig, Allen T.
Macmillan, New York, 1959.

STOCHASTIC PROCESSES AND FILTERING THEORY

Jazwinak, Andrew H.
Academic Press, New York, 1970.
Mathematics in Science and Engineering Vol. 64.

APPLIED TIME SERIES ANALYSIS FOR MANAGERIAL FORECASTING

Nelson, Charles R.
Holden-day, San Francisco, 1973.

FUNDAMENTALS OF FORECASTING

Sullivan, William G.
Claycombe, W. Wayne
Reston Publishing, Reston, 1977.

AN INTRODUCTION TO BAYESIAN INFERENCE IN ECONOMETRICS

Zellner, Arnold
John Wiley, New York, 1971.