

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



**MODELOS LINEALES DINÁMICOS
PARA EL PRONÓSTICO DE LA
DEMANDA DE DINERO EN MÉXICO**

TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ACTUARÍA

PRESENTA

JUAN PABLO LÓPEZ ESCAMILLA

ASESOR: DR. JORGE FRANCISCO DE LA VEGA GÓNGORA

CIUDAD DE MÉXICO

2022

“Con fundamento en el artículo 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada ”NOMBRE COMPLETO DE LA TESIS O TESINA”, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr. autorización para que fijen la obra en cualquier medio, incluido el electrónico y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por la divulgación una contraprestación”.

AUTOR

FECHA

FIRMA

DEDICATORIA

Agradecimientos

¡Muchas gracias a todos!

Índice general

1. Introducción	1
1.1. Historia de los billetes y monedas en México	1
1.2. Uso del dinero en el presente	3
1.3. Revisión de literatura	5
1.4. Objetivos de la tesis	8
2. Marco teórico	9
2.1. Notación	9
2.2. Modelo lineal dinámico	10
2.3. Modelo lineal dinámico univariado	12
2.3.1. Actualización de ecuaciones	12
2.3.2. Distribuciones de pronóstico	21
2.3.3. Varianzas desconocidas	28
2.3.4. Análisis retrospectivo	40
2.3.5. Intervenciones	48
2.4. Modelo lineal dinámico multivariado	56
2.4.1. Modelos lineales dinámicos matriz normal	57
2.4.2. Distribución de los errores de evolución	60
2.4.3. Distribución Wishart inversa	60
2.4.4. Distribución matriz normal-Wishart inversa	61
2.4.5. Distribución matriz T	62
3. Pronóstico de la demanda agregada de dinero	63
3.1. Implementación en R	63
3.2. Verificación del modelo	63
3.3. Análisis de intervención durante la pandemia	63

3.4. Comparación con Error Correction Model	63
4. Pronóstico de la demanda de dinero por denominación	65
4.1. Implementación en R	65
4.2. Verificación del modelo	65
4.3. Análisis de intervención durante la pandemia	65
4.4. Comparación con Error Correction Model	65
5. Conclusión	67
5.1. Resumen de los resultados	67
5.2. Limitaciones y mejoras futuras	67
A. Resultados de probabilidad y estadística	69
A.1. Distribuciones marginales y condicionales de la normal mul- tivariable	69
A.2. Distribución normal condicional y regresión lineal	70
A.3. Distribución normal-gamma	72
 Bibliografía	 75

Capítulo 1

Introducción

En este capítulo se presenta la relevancia del dinero en la economía de México; la larga historia que tiene en el país y su importancia en el presente, reforzada durante la pandemia. Esto nos lleva a la necesidad de pronosticar correctamente la demanda de billetes y monedas para satisfacer el uso de efectivo de la población. Se comentan los modelos que históricamente se han empleado en el pronóstico de la demanda de dinero. Finalmente, se definen los objetivos de la tesis.



1.1. Historia de los billetes y monedas en México

El uso de billetes y monedas cuenta con una historia de varios siglos en el territorio mexicano. En la primera etapa de la Conquista, los españoles usaban las pocas monedas que cargaban con ellos junto con los productos que usaban los indígenas para intercambiar, como cacao o mantas de algodón. Con el crecimiento del comercio en la Nueva España, se hizo evidente la necesidad de establecer una moneda para realizar transacciones económicas. En 1535 se fundó la Casa de Moneda de México, la primera en América. Las primeras monedas acuñadas en el país fueron hechas a mano, pero de grosor parejo y redondas, de forma parecida a las monedas que actualmente tenemos en ~~México~~. Por su parte, los indígenas rechazaron estas monedas debido a que ellos utilizaban otros medios de cambio, como el cacao, cuentas de jade o cuentas de algodón. Con el paso del tiem-

po, debido a la importancia del **Imperio Español**, y a que muchos países no tenían plata ni casas de moneda, la moneda mexicana dominó los mercados internacionales durante tres siglos. Con el paso de los años, los materiales y las técnicas de acuñación de monedas evolucionaron. El valor nominal de las monedas era equivalente al valor del metal que contenían.

Durante la Guerra de Independencia, los capitalistas enviaron su dinero a Europa y muchas minas fueron tomadas por los combatientes, lo que causó una gran escasez de dinero. Para enfrentar esta situación, las fuerzas insurgentes fabricaron sus propias monedas. Morelos ordenó la acuñación de monedas de cobre, con su valor expresado en la cara de la moneda. Esta es la primera vez que circuló una moneda fiduciaria¹ en el país, estas monedas valían ocho, dos, uno, y medio real. Fue en este periodo, debido a la escasez de dinero, que surgieron unas piezas de cartón, escritas a mano con tinta negra, con valor de medio real. Estas piezas son reconocidas por especialistas como el predecesor más antiguo del billete mexicano.

Una vez terminada la Guerra de Independencia, la economía de México estaba destruida. Iturbide intentó reactivar la economía aplicando varias medidas; entre ellas la emisión de papel moneda. Esta fue la primera vez que se emitieron billetes mexicanos oficiales. Sin embargo, la población rechazó su uso debido a que la población estaba acostumbrada a usar monedas de plata y a que la aceptación de los billetes era obligatoria. Tras el colapso del Primer Imperio mexicano, se estableció la moneda republicana en 1823. Continuaron los esfuerzos para que la población aceptara los nuevos billetes, incluso se imprimieron sobre documentos religiosos para facilitar su aceptación, pero el nuevo billete republicano también fue rechazado.

Décadas después, con la llegada de Maximiliano de Habsburgo, se acuñaron las primeras monedas de un peso, bajo el sistema decimal. Fue en este periodo cuando por fin se logró introducir exitosamente el billete: eran emitidos por un banco privado y de aceptación voluntaria. Cuando se restauró la república, la acuñación de monedas con el peso como unidad continuó. Durante el Porfiriato, gracias a un sistema bancario más avanzado y funcional,

¹Una moneda fiduciaria no está vinculada al valor del metal que se empleó en su fabricación, su valor está basado en la confianza pública que se tiene en ella.

el billete fue finalmente aceptado como medio de pago de forma generalizada.

Durante la Revolución se alteró la fabricación y distribución de monedas, además resurgió el rechazo por los billetes. Huerta ordenó a bancos privado emitir grandes cantidades de billetes, en consecuencia, los billetes se depreciaron. Debido a la falta de dinero, las fuerzas revolucionarias crearon sus propias monedas y billetes, lo que complicó el problema monetario del país. Después de la Revolución, comenzó el proceso para sustituir las monedas de oro y plata por monedas de metales industriales.

En 1925 surgió el Banco de México. Entre sus facultades está el monopolio de la emisión de billetes y monedas. Uno de los objetivos del nuevo banco central era reconciliar a la población con el uso de billetes. Hasta la década de 1930 la aceptación de los billetes fue generalizada de nuevo [1].

Las monedas han sido utilizadas en territorio mexicano por casi 500 años, mientras que los billetes, a pesar de ser rechazadas en varios periodos del siglo XIX y durante la Revolución, son aceptados desde hace 90 años. La historia nos enseña que los billetes y monedas son tan importantes que, en épocas de escasez de efectivo (como en la Guerra de Independencia y en la Revolución), diversos grupos han emitido sus propios billetes y monedas de necesidad para enfrentar la escasez. Sin embargo, la emisión descontrolada de dinero también ha generado complicaciones. Recordemos que Victoriano Huerta ordenó la emisión de un enorme número de billetes, lo que provocó que se depreciaran. Tanto la escasez como el exceso de dinero han traído grandes problemas económicos al país en el pasado, lo que muestra la importancia satisfacer la demanda de efectivo con precisión. Para esto, es esencial pronosticar correctamente la demanda futura de dinero.



1.2. Uso del dinero en el presente

Los billetes y monedas conservan su gran importancia en el presente de México. De acuerdo con una encuesta con representatividad nacional realizada por el Banco de México en el 2020 [2, p. 13], dirigida a personas de entre 18 y 79 años de edad que viven en localidades de 50 mil habitantes

o más, el 97 % de la población objetivo acostumbra usar efectivo: de estas personas, el 41 % mencionaron que la razón por la que usan efectivo es porque es más práctico, fácil o rápido; el 29 % únicamente tiene esta forma de pago, y el 8 % porque es más seguro. La encuesta también señala que el 35 % de la población objetivo guarda efectivo en algún lugar distinto a una institución financiera: de este grupo de personas el 42 % lo hace para alguna eventualidad; el 22 % para ahorrar, y el 8 % para tener tranquilidad. La baja inclusión financiera que se tiene en México impide que un sector importante de la población tenga acceso a otros medios de pago además del efectivo. En 2021 el 22 % de la población adulta nunca había tenido un producto financiero² [3, p. 15].

Esto muestra que a pesar de las innovaciones tecnológicas dentro del sector financiero, los billetes y las monedas aún son un componente esencial de la economía de la mayoría de los mexicanos por diversas razones.

Dado que el efectivo es ampliamente utilizado por los mexicanos, es fundamental proveer a la economía con suficientes billetes y monedas para cubrir la demanda que se tiene de estos activos. Sin embargo, la fabricación y distribución de los billetes y monedas no es inmediata, por lo que es necesario realizar pronósticos de la demanda de dinero para determinar la cantidad que se debe producir en el presente con el objetivo de cubrir el efectivo requerido en el futuro. Adicionalmente, la cantidad de dinero está estrechamente vinculada con el gasto y, en consecuencia, con el nivel de los precios. Por lo tanto, la demanda de dinero es de interés para la política monetaria del país [4, p. 7].

El uso de billetes y monedas se amplificó durante la pandemia de COVID-19 en el 2020. Esto fue un fenómeno mundial, no ocurrió únicamente en México. En la Figura 1.1 es fácil notar que durante el 2020 los billetes y monedas en circulación tuvieron un incremento considerablemente mayor que en los demás años. Antes de 2020 el comportamiento del efectivo era claro: la tendencia general es creciente, con una estacionalidad en la que el dinero incrementa a finales de cada año y posteriormente disminuye. En enero y febrero de 2020 el dinero en circulación disminuyó, que es lo

²Hay cuatro tipos básicos de productos financieros. Productos de captación (cuentas bancarias), crédito formal, seguros y cuentas de ahorro para el retiro.

que normalmente pasaba antes de la pandemia, pero a partir de marzo comenzó a aumentar rápidamente. Esto ocurrió en un momento en el que la actividad económica disminuyó, lo que en primera instancia puede parecer contradictorio, pues entre mayor actividad económica se espera mayor uso de efectivo.

En el informe trimestral del Banco de México de abril-junio 2021 se explica este fenómeno [5, pp. 52-55]. Las variables que normalmente son utilizadas para pronosticar la demanda de dinero, como el PIB o las tasas de interés, no son suficientes para explicar lo sucedido después de marzo 2020. Esto significa que el incremento en la demanda de dinero debe estar relacionado con otro tipo de factores que normalmente no eran importantes antes de la pandemia. Una explicación es que ante la incertidumbre provocada por la pandemia, las personas buscaron contar con una mayor liquidez para enfrentar posibles gastos relacionados con salud o pérdida de empleo. Además, la circulación de billetes de 500 pesos fue la que más aumentó, lo que refuerza la idea del aumento de la demanda por motivos precautorios. Aún no es claro si esta situación es temporal o permanente, pero es evidente que en el inicio de 2022 la demanda de dinero no ha recuperado su comportamiento pre-pandémico.

Lo sucedido en el 2020 con la demanda de dinero debido a factores que anteriormente no eran considerados nos muestra la necesidad de utilizar modelos en los que se pueda incluir información externa con facilidad. En este contexto los modelos lineales dinámicos son de especial utilidad.

1.3. Revisión de literatura

El análisis y pronóstico de la demanda de dinero se realiza frecuentemente por los bancos centrales, instituciones e investigadores de distintos países desde hace décadas. Gracias a esto, se ha desarrollado una rica variedad de estudios teóricos y empíricos.

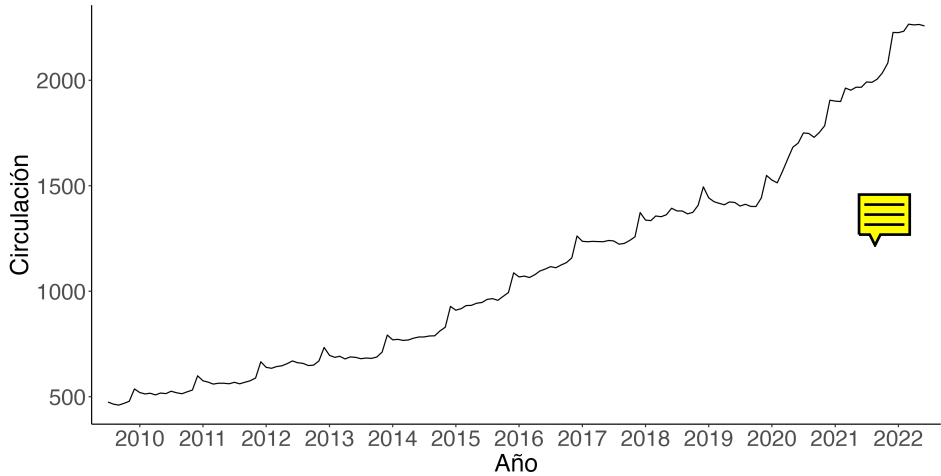


Figura 1.1: Billetes y monedas en poder del público (miles de millones de pesos).

Fuente: Banco de México [6]

Aunque hay una gran variedad de marcos teóricos que analizan la demanda de dinero desde distintas perspectivas, en general coinciden en que esta depende de una variable de escala (medida de actividad económica) y de variables de costo de oportunidad (que indiquen ganancias de utilizar activos distintos al efectivo). Algunos de estos marcos teóricos fueron formulados en la década de 1920, lo que señala la larga historia que tiene la teoría de este tema. Por ejemplo, en la teoría de Keynes se postuló la siguiente función de demanda de dinero: $m^d = f(y, i)$, en donde m^d es el dinero real, y es el ingreso real, mientras que i es la tasa de interés. Otros modelos anteriores ya habían propuesto explicar la demanda de dinero mediante variables de escala y de costo de oportunidad, pero la contribución importante de Keynes fue usar la tasa de interés. A partir de esta teoría, surgieron otras que consideraban estas variables. A pesar de las grandes diferencias que estas teorías tienen entre sí, la conclusión generalmente es que relación de la demanda de dinero es positiva con el ingreso y negativa con i . Esta conclusión es el punto de inicio de los modelos empíricos.

La variable de interés que se estudia en los modelos empíricos varía. Por lo general se han utilizado distintos agregados monetarios: en algunos análisis se utilizan agregados monetarios estrechos, como M0 o M1, mientras que otros se han concentrado en agregados más amplios, como M2, M3 o M4 ³. Además, la definición de los agregados monetarios puede cambiar en cada país, por lo que la variable de interés en este tipo de estudios no es homogénea.

Sobre la **variable de escala**, entre las más empleadas se encuentran el producto nacional bruto (PNB), producto interno bruto (PIB) o producto nacional neto (PNN), aunque la mayoría de los estudios utilizan en PNB. Respecto a la variable de costo de oportunidad, se pueden utilizar una o varias tasas de corto plazo, de largo plazo, o incluso de ambos tipos. También se han utilizado los spreads de las tasas de interés o la inflación. De hecho, para estimar la demanda de dinero en países que no tienen una industria financiera desarrollada, solo se utiliza la inflación para representar el costo de oportunidad de poseer efectivo debido a que la sustitución de dinero por otros activos es limitada y es posible que no existan datos de las tasas de interés. Otro factor que puede ser considerado son las influencias extranjeras, en este caso se considera el tipo de cambio o las tasas de retorno de valores extranjeros.

Los modelos de corrección de error (ECM por sus siglas en inglés) se convirtieron en los más empleados en el análisis de la demanda de dinero en la década de 1990. Una característica importante de estos modelos es que la teoría describe el equilibrio en el largo plazo, pero los datos definen el comportamiento en el corto plazo. Las relaciones a largo plazo se obtienen mediante cointegración, y si ocurre un desequilibrio, el mecanismo de corrección de error ajustará el modelo en el corto plazo para volver al equilibrio. Esto representa una ventaja importante frente a los modelos

³Los agregados monetarios en México se definen de la siguiente manera:

M0 = Billetes y monedas en poder del público.

M1 = M0 + depósitos de disponibilidad inmediata en bancos, entidades de ahorro y crédito popular.

M2 = M1 + instrumentos monetarios a plazo en poder de residentes.

M3 = M2 + valores públicos en poder de residentes.

M4 = M3 + instrumentos monetarios en poder de no residentes. [6]

utilizados anteriormente, los cuales se enfocaban en la teoría económica sin examinar los datos a profundidad. [7]

En la actualidad los ECM conservan su importancia en el análisis de la demanda de dinero. Ibarra [8] empleó un ECM para estimar una función de demanda de dinero en México para el periodo 1980-2005. Este estudio se concentra en el agregado M1. Primero se estimó la función de demanda de dinero en el largo plazo. Posteriormente, se empleó un ECM para corregir los desequilibrios en el corto plazo. Finalmente, se concluye que el modelo estimado es estable para el periodo estudiado.

Noriega *et al* [9] realizaron un estudio parecido para el periodo 2006-2010.

1.4. Objetivos de la tesis

La principal meta de este trabajo es el pronóstico de la demanda de dinero, aunque también se analizan las relaciones con otras variables. Se emplean principalmente los modelos lineales dinámicos, pero para determinar su utilidad se compara su desempeño contra un modelo que es frecuentemente aplicado en el estudio de la demanda de dinero: el error correction model (ECM). La variable de interés es el agregado M0, es decir, billetes y monedas en poder del público. Se ilustran las ventajas que tienen los modelos lineales dinámicos ante cambios estructurales tomando como ejemplo lo sucedido durante la pandemia. Posteriormente, se aplican los modelos lineales dinámicos para pronosticar la demanda de dinero de las distintas denominaciones de billetes y monedas.

Capítulo 2

Marco teórico

2.1. Notación

La teoría y los procedimientos de modelos lineales dinámicos presentados en esta tesis provienen del libro de West y Harrison [10]. Por lo tanto, la notación empleada es la misma que la de **este** libro.

La serie de tiempo es representada por Y_t , en donde el índice t denota el valor de la serie en la t -ésima observación. En este trabajo, Y_t es la observación en el tiempo t . En la notación no se distingue si Y_t es variable aleatoria o un valor observado.

Las funciones de densidad se denotan como $f(\cdot)$. Las varianzas y las covarianzas están representadas como $V[\cdot]$ y $C[\cdot, \cdot]$ respectivamente.

Los caracteres romanos, tanto mayúsculas como minúsculas, se utilizan para cantidades que son conocidas, o que son inciertas pero potencialmente observables (por ejemplo, observaciones futuras). Por otro lado, las cantidades que son inciertas y no observables, es decir, los parámetros desconocidos, se denotan con letras griegas. Los vectores y matrices se representan en negritas. Las matrices están siempre en mayúsculas, mientras que en el caso de los vectores depende del contexto.

2.2. Modelo lineal dinámico

Se considera que la información disponible en el tiempo $t-1$ es D_{t-1} y se tiene como objetivo modelar \mathbf{Y}_t . El primer paso es definir una parametrización $\boldsymbol{\theta}_{t-1}$ que contenga toda la información relevante para predecir futuras observaciones. Esta información relevante, la cual también puede incluir creencias subjetivas de expertos, se representa a través de la distribución de probabilidad $(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$. Es importante notar que la parametrización está indexada, por lo tanto, es dinámica, cambia a lo largo del tiempo. Con esta formulación probabilística se pueden asimilar fácilmente cambios no anticipados y sus incertidumbres asociadas. Una razón por la cual puede ser necesario realizar cambios en el modelo es si se recibe información relevante de una fuente externa que no fue considerada inicialmente, por lo que es necesario combinarla con la información probabilística existente. Otra razón importante surge cuando el modelo tiene un desempeño inadecuado, en este caso es necesario modificar el modelo para corregirlo. En el modelo lineal dinámico se emplea el paradigma bayesiano, aunque también son utilizados en el análisis frecuentista, en donde se conocen como **modelos estructurales**.

Definición 2.2.0.1 (Modelo Lineal Dinámico). La forma de un Modelo Lineal Dinámico (DLM por sus siglas en inglés) está caracterizada por un conjunto de cuádruplas:

$$\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t = \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$$

donde

- \mathbf{F}_t es una matriz $n \times r$ conocida.
- \mathbf{G}_t es una matriz $n \times n$ conocida.
- \mathbf{V}_t es una matriz de covarianza $r \times r$ conocida.
- \mathbf{W}_t es una matriz de covarianza $n \times n$ conocida.

Sea \mathbf{Y}_t un vector $r \times 1$, entonces \mathbf{Y}_t se relaciona con el vector $n \times 1$ de parámetros $\boldsymbol{\theta}_t$ en el tiempo t , y con la secuencia $\boldsymbol{\theta}_t$ a lo largo del tiempo, de la siguiente manera:

$$\text{Ecuación de observación:} \quad \mathbf{Y}_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim \mathbf{N}[\mathbf{0}, \mathbf{V}_t]$$

$$\text{Ecuación de sistema:} \quad \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathbf{N}[\mathbf{0}, \mathbf{W}_t]$$

Las secuencias de errores $\boldsymbol{\omega}_t$ y $\boldsymbol{\nu}_t$ son internamente y mutuamente independientes. O alternativamente:

$$(\mathbf{Y}_t \mid \boldsymbol{\theta}_t) \sim \mathbf{N}[\mathbf{F}_t^\top \boldsymbol{\theta}_t, \mathbf{V}_t] \quad (2.1)$$

y

$$(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}) \sim \mathbf{N}[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t] \quad (2.2)$$

Las ecuaciones 2.1 y 2.2 están implícitamente condicionadas en el conjunto de información anterior al tiempo t , D_{t-1} . Esto incluye, entre otros elementos, los valores de las varianzas \mathbf{V}_t y \mathbf{W}_t y las observaciones pasadas $\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots$, así como el conjunto de información inicial D_0 . No se incluye D_{t-1} en las ecuaciones por simplicidad notacional.

Para el tiempo t , se definen los siguientes términos:

- \mathbf{F}_t es la matriz de diseño.
- $\boldsymbol{\theta}_t$ es el vector de estados.
- $\boldsymbol{\mu}_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t$ es la respuesta media, o nivel.
- $\boldsymbol{\nu}_t$ es el error de observación.
- \mathbf{G}_t es la matriz de sistema.
- $\boldsymbol{\omega}_t$ es el error de sistema.

El modelo es dinámico porque los valores de los parámetros cambian a lo largo del tiempo, pero están estocásticamente relacionados a través de la ecuación de sistema. Esto es apropiado cuando se desea modelar fenómenos

en los cuales las circunstancias cambian. La ecuación de sistema describe la evolución de los parámetros con el paso del tiempo, es decir, la dinámica del modelo. La matriz de sistema \mathbf{G}_t define una relación funcional determinista de los parámetros en cierto periodo con sus valores en el periodo inmediatamente interior. El error de sistema $\boldsymbol{\omega}_t$ permite propagación imperfecta, lo que es importante para modelar el incremento de incertidumbre, o equivalentemente la pérdida de valor de la información entre $t - 1$ y t . Esto quiere decir que la información actual es más importante que la información pasada.

2.3. Modelo lineal dinámico univariado

El modelo lineal dinámico univariado es un caso particular de la definición 2.2.0.1 En el caso univariado $r = 1$, por lo tanto, el modelo está caracterizado por la cuádrupla

$$\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$$

Definición 2.3.0.1 (Modelo Lineal Dinámico univariado). Para cada t , el modelo lineal dinámico univariado se define así:

$$\text{Ecuación de observación:} \quad Y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim \mathbf{N}[0, V_t]$$

$$\text{Ecuación de sistema:} \quad \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathbf{N}[\mathbf{0}, \mathbf{W}_t]$$

$$\text{Información inicial:} \quad (\boldsymbol{\theta}_0 \mid D_0) \sim \mathbf{N}[\mathbf{m}_0, \mathbf{C}_0],$$

donde \mathbf{m}_0 y \mathbf{C}_0 son momentos a priori, las secuencias de errores $\boldsymbol{\omega}_t$ y ν_t son internamente y mutuamente independientes, y son independientes de $(\boldsymbol{\theta}_0 \mid D_0)$

2.3.1. Actualización de ecuaciones

Si se considera que el modelo es cerrado a información externa en $t \geq 1$ (posteriormente se mostrará la manera de incluir información externa),

entonces dada la información inicial a priori D_0 en $t = 0$, en cualquier t futura el conjunto de información es

$$D_t = \{Y_t, D_{t-1}\}$$

El modelo asume que en cualquier tiempo la información existente está representada y es resumida de manera suficiente por la distribución posterior del vector de estados. Los resultados se presentan en el siguiente teorema:

Teorema 2.3.1.1. *Para el Modelo lineal dinámico univariado de la definición 2.3.0.1, los pronósticos a un periodo posterior y las distribuciones posteriores para cada t se obtienen de la siguiente manera:*

A. *Posterior en $t - 1$:*

Dadas la media \mathbf{m}_{t-1} y la matriz de covarianza \mathbf{C}_{t-1} ,

$$(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) \sim \mathbf{N}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$$

B. *Priori en t :*

$$(\boldsymbol{\theta}_t \mid D_{t-1}) \sim \mathbf{N}[\mathbf{a}_t, \mathbf{R}_t],$$

donde

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{y} \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}_t$$

C. *Pronóstico para t :*

$$(Y_t \mid D_{t-1}) \sim \mathbf{N}[f_t, Q_t],$$

donde

$$f_t = \mathbf{F}_t^\top \mathbf{a}_t \quad \text{y} \quad Q_t = \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t + V_t$$

D. *Posterior en t :*

$$(\boldsymbol{\theta}_t \mid D_t) \sim \mathbf{N}[\mathbf{m}_t, \mathbf{C}_t]$$

donde

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t \quad \text{y} \quad \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t^\top$$

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1} \quad \text{y} \quad e_t = Y_t - f_t$$

Demostración. [10, pp. 104-105]

La prueba es por inducción. En la demostración, \mathbf{Y}_t es un vector debido a que la prueba es válida para el caso univariado y para el multivariado.

Base de inducción. Se demuestra la validez para $t = 1$.

Por definición del modelo lineal dinámico, se tiene el siguiente modelo:

$$\mathbf{Y}_1 = \mathbf{F}_1^\top \boldsymbol{\theta}_1 + \nu_1, \quad \nu_1 \sim \mathbf{N}[\mathbf{0}, \mathbf{V}_1]$$

$$\boldsymbol{\theta}_1 = \mathbf{G}_1 \boldsymbol{\theta}_0 + \omega_1, \quad \omega_1 \sim \mathbf{N}[\mathbf{0}, \mathbf{W}_1]$$

$$(\boldsymbol{\theta}_0 \mid D_0) \sim \mathbf{N}[\mathbf{m}_0, \mathbf{C}_0].$$

Entonces,

- $E[\boldsymbol{\theta}_1 \mid D_0] = E[\mathbf{G}_1 \boldsymbol{\theta}_0 + \omega_1 \mid D_0] = \mathbf{G}_1 E[\boldsymbol{\theta}_0 \mid D_0] + E[\omega_1 \mid D_0] = \mathbf{G}_1 \mathbf{m}_0;$
- $V[\boldsymbol{\theta}_1 \mid D_0] = V[\mathbf{G}_1 \boldsymbol{\theta}_0 + \omega_1 \mid D_0] = \mathbf{G}_1 V[\boldsymbol{\theta}_0 \mid D_0] \mathbf{G}_1^\top + V[\omega_1 \mid D_0] = \mathbf{G}_1 \mathbf{C}_0 \mathbf{G}_1^\top + \mathbf{W}_1.$

Sean $\mathbf{a}_1 = \mathbf{G}_1 \mathbf{m}_0$ y $\mathbf{R}_1 = \mathbf{G}_1 \mathbf{C}_0 \mathbf{G}_1^\top + \mathbf{W}_1$, entonces

$$(\boldsymbol{\theta}_1 \mid D_0) \sim \mathbf{N}[\mathbf{a}_1, \mathbf{R}_1].$$

Al utilizar los resultados del Apéndice A.2 se obtiene la siguiente distribución conjunta entre $\boldsymbol{\theta}_1$ y \mathbf{Y}_1 :

$$\begin{pmatrix} \theta_1 \\ Y_1 \end{pmatrix} \Big| D_0 \sim N \left[\begin{pmatrix} \mathbf{a}_1 \\ \mathbf{F}_1^\top \mathbf{a}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_1 \mathbf{F}_1 \\ \mathbf{F}_1^\top \mathbf{R}_1 & \mathbf{F}_1^\top \mathbf{R}_1 \mathbf{F}_1 + \mathbf{V}_1 \end{pmatrix} \right]$$

Sean $\mathbf{Q}_1 = \mathbf{F}_1^\top \mathbf{R}_1 \mathbf{F}_1 + \mathbf{V}_1$ y $\mathbf{f}_1 = \mathbf{F}_1^\top \mathbf{a}_1$, entonces

$$(\mathbf{Y}_1 \mid D_0) \sim N[\mathbf{f}_1, \mathbf{Q}_1],$$

Como \mathbf{Q}_1 es una matriz de covarianza, entonces es simétrica.

Sea $\mathbf{A}_1 = \mathbf{R}_1 \mathbf{F}_1 \mathbf{Q}_1^{-1}$, entonces $C[\theta_1, \mathbf{Y}_1] = \mathbf{R}_1 \mathbf{F}_1 = \mathbf{R}_1 \mathbf{F}_1 \mathbf{Q}_1^{-1} \mathbf{Q}_1 = \mathbf{A}_1 \mathbf{Q}_1$ y $C[\mathbf{Y}_1 \theta_1] = C[\theta_1, \mathbf{Y}_1]^\top = \mathbf{Q}_1^\top \mathbf{A}_1^\top = \mathbf{Q}_1 \mathbf{A}_1^\top$.

Por lo tanto, la distribución conjunta entre θ_1 y \mathbf{Y}_1 también se puede escribir como

$$\begin{pmatrix} \theta_1 \\ \mathbf{Y}_1 \end{pmatrix} \Big| D_0 \sim N \left[\begin{pmatrix} \mathbf{a}_1 \\ \mathbf{f}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{R}_1 & \mathbf{A}_1 \mathbf{Q}_1 \\ \mathbf{Q}_1 \mathbf{A}_1^\top & \mathbf{Q}_1 \end{pmatrix} \right]$$

Por último, la teoría presentada en A.2 indica que si $\mathbf{e}_1 = \mathbf{Y}_1 - \mathbf{f}_1$, entonces

$$(\theta_1 \mid D_0, \mathbf{Y}_1) \sim N[\mathbf{a}_1 + \mathbf{A}_1 \mathbf{e}_1, \mathbf{R}_1 - \mathbf{A}_1 \mathbf{Q}_1 \mathbf{A}_1^\top],$$

como $D_1 = \{D_0, \mathbf{Y}_1\}$, entonces

$$(\theta_1 \mid D_1) \sim N[\mathbf{m}_1, \mathbf{C}_1],$$

con $\mathbf{m}_1 = \mathbf{a}_1 + \mathbf{A}_1 \mathbf{e}_1$ y $\mathbf{C}_1 = \mathbf{R}_1 - \mathbf{A}_1 \mathbf{Q}_1 \mathbf{A}_1^\top$

Se supone que el teorema es válido para t . Se probará la validez para $t + 1$

Se tiene el siguiente modelo:

$$\mathbf{Y}_{t+1} = \mathbf{F}_{t+1}^\top \theta_{t+1} + \nu_{t+1}, \quad \nu_{t+1} \sim N[0, V_{t+1}]$$

$$\boldsymbol{\theta}_{t+1} = \mathbf{G}_{t+1}\boldsymbol{\theta}_t + \boldsymbol{\omega}_{t+1}, \quad \boldsymbol{\omega}_{t+1} \sim \mathbf{N}[\mathbf{0}, \mathbf{W}_{t+1}]$$

Se supone por hipótesis de inducción que

$$(\boldsymbol{\theta}_t \mid D_t) \sim \mathbf{N}[\mathbf{m}_t, \mathbf{C}_t].$$

Entonces,

- $E[\boldsymbol{\theta}_{t+1} \mid D_t] = E[\mathbf{G}_{t+1}\boldsymbol{\theta}_t + \boldsymbol{\omega}_{t+1} \mid D_t] = \mathbf{G}_{t+1}E[\boldsymbol{\theta}_t \mid D_t] + E[\boldsymbol{\omega}_{t+1} \mid D_t] = \mathbf{G}_{t+1}\mathbf{m}_t;$
- $V[\boldsymbol{\theta}_{t+1} \mid D_t] = V[\mathbf{G}_{t+1}\boldsymbol{\theta}_t + \boldsymbol{\omega}_{t+1} \mid D_t] = \mathbf{G}_{t+1}V[\boldsymbol{\theta}_t \mid D_t]\mathbf{G}_{t+1}^\top + V[\boldsymbol{\omega}_{t+1} \mid D_t] = \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}_{t+1}^\top + \mathbf{W}_{t+1}.$

Sean $\mathbf{a}_{t+1} = \mathbf{G}_{t+1}\mathbf{m}_t$ y $\mathbf{R}_{t+1} = \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}_{t+1}^\top + \mathbf{W}_{t+1}$, entonces

$$(\boldsymbol{\theta}_{t+1} \mid D_t) \sim \mathbf{N}[\mathbf{a}_{t+1}, \mathbf{R}_{t+1}].$$

Al utilizar los resultados del Apéndice A.2 se obtiene la siguiente distribución conjunta entre $\boldsymbol{\theta}_{t+1}$ y \mathbf{Y}_{t+1} :

$$\left(\begin{matrix} \boldsymbol{\theta}_{t+1} \\ \mathbf{Y}_{t+1} \end{matrix} \middle| D_t \right) \sim \mathbf{N} \left[\begin{pmatrix} \mathbf{a}_{t+1} \\ \mathbf{F}_{t+1}^\top \mathbf{a}_{t+1} \end{pmatrix}, \begin{pmatrix} \mathbf{R}_{t+1} & \mathbf{R}_{t+1}\mathbf{F}_{t+1} \\ \mathbf{F}_{t+1}^\top \mathbf{R}_{t+1} & \mathbf{F}_{t+1}^\top \mathbf{R}_{t+1} \mathbf{F}_{t+1} + \mathbf{V}_{t+1} \end{pmatrix} \right]$$

Sean $\mathbf{Q}_{t+1} = \mathbf{F}_{t+1}^\top \mathbf{R}_{t+1} \mathbf{F}_{t+1} + \mathbf{V}_{t+1}$ y $\mathbf{f}_{t+1} = \mathbf{F}_{t+1}^\top \mathbf{a}_{t+1}$, entonces

$$(\mathbf{Y}_{t+1} \mid D_t) \sim \mathbf{N}[\mathbf{f}_{t+1}, \mathbf{Q}_{t+1}],$$

Sea $\mathbf{A}_{t+1} = \mathbf{R}_{t+1} \mathbf{F}_{t+1} \mathbf{Q}_{t+1}^{-1}$, entonces $C[\boldsymbol{\theta}_{t+1}, \mathbf{Y}_{t+1}] = \mathbf{R}_{t+1} \mathbf{F}_{t+1} = \mathbf{R}_{t+1} \mathbf{F}_{t+1} \mathbf{Q}_{t+1}^{-1} \mathbf{Q}_{t+1} = \mathbf{A}_{t+1} \mathbf{Q}_{t+1}$ y $C[\mathbf{Y}_{t+1} \boldsymbol{\theta}_{t+1}] = C[\boldsymbol{\theta}_{t+1}, \mathbf{Y}_{t+1}]^\top = \mathbf{Q}_{t+1}^\top \mathbf{A}_{t+1}^\top = \mathbf{Q}_{t+1} \mathbf{A}_{t+1}^\top.$

Por lo tanto, la distribución conjunta entre $\boldsymbol{\theta}_{t+1}$ y \mathbf{Y}_{t+1} también se puede escribir como

$$\begin{pmatrix} \boldsymbol{\theta}_{t+1} \\ \mathbf{Y}_{t+1} \end{pmatrix} \Big| D_t \sim \mathbf{N} \left[\begin{pmatrix} \mathbf{a}_{t+1} \\ \mathbf{f}_{t+1} \end{pmatrix}, \begin{pmatrix} \mathbf{R}_{t+1} & \mathbf{A}_{t+1} \mathbf{Q}_{t+1} \\ \mathbf{Q}_{t+1} \mathbf{A}_{t+1}^\top & \mathbf{Q}_{t+1} \end{pmatrix} \right]$$

Por último, la teoría presentada en A.2 indica que si $\mathbf{e}_{t+1} = \mathbf{Y}_{t+1} - \mathbf{f}_{t+1}$, entonces

$$(\boldsymbol{\theta}_{t+1} \mid D_t, \mathbf{Y}_{t+1}) \sim \mathbf{N}[\mathbf{a}_{t+1} + \mathbf{A}_{t+1} \mathbf{e}_{t+1}, \mathbf{R}_{t+1} - \mathbf{A}_{t+1} \mathbf{Q}_{t+1} \mathbf{A}_{t+1}^\top],$$

como $D_{t+1} = \{D_t, \mathbf{Y}_{t+1}\}$, entonces

$$(\boldsymbol{\theta}_{t+1} \mid D_{t+1}) \sim \mathbf{N}[\mathbf{m}_{t+1}, \mathbf{C}_{t+1}],$$

con $\mathbf{m}_{t+1} = \mathbf{a}_{t+1} + \mathbf{A}_{t+1} \mathbf{e}_{t+1}$ y $\mathbf{C}_{t+1} = \mathbf{R}_{t+1} - \mathbf{A}_{t+1} \mathbf{Q}_{t+1} \mathbf{A}_{t+1}^\top$

□

Es importante analizar lo que ocurre cuando la esperanza de la distribución de pronóstico es muy diferente al valor observado de Y_t . Si esto sucede, e_t tendrá un valor grande, por lo que la diferencia entre \mathbf{a}_t y \mathbf{m}_t será considerable. Entonces, cuando el error es grande, se tendrá un fuerte ajuste en la distribución posterior respecto a la priori. Así es como la distribución posterior se ajusta con base en la Y_t observada.

Ejemplo 2.3.1.1. En este ejemplo se utilizará la base de datos freeny disponible en el paquete datasets de R para ilustrar el Teorema 2.3.1.1. En este conjunto de datos se tiene una variable y que representa la serie de tiempo de las ganancias cuatrimestrales desde 1962 C2 hasta 1971 C4. Se tienen como variables explicativas el índice de precio, el nivel de ingreso y el potencial de mercado. Únicamente con el objetivo de facilitar el ejemplo, se inventará un contexto ficticio para la base de datos. Este contexto ficticio es que la variable y representa las ganancias de una empresa de refrescos en millones de pesos. Como variables explicativas se consideran el índice de precios, que se refiere al precio de los refrescos, y el nivel de ingreso, que equivale al ingreso de los consumidores.

Entonces se tiene

$$Y_t = \text{Ganancias en } t, \quad \theta_t = \begin{pmatrix} \text{Intercepto en } t \\ \text{Ingreso en } t \\ \text{Precio en } t \end{pmatrix}$$

En este contexto ficticio, estamos ubicados en el cuarto cuatrimestre de 1966. La serie empieza en 1962 C2, entonces 1966 C4 equivale a $t = 19$. Las ganancias en 1966 C4 ya fueron observadas, mientras que de 1967 C1 en adelante no se han observado, por lo que solo se tiene la información D_{19} , que incluye las observaciones de 1966 C4 y anteriores.

Se inicia con la distribución posterior de los parámetros en $t = 19$. Supongamos que ya fue obtenida anteriormente y es de la siguiente forma:

$$(\theta_{19} | D_{19}) \sim \mathbf{N} \left[\begin{pmatrix} 8 \\ 0.35 \\ -0.27 \end{pmatrix}, \begin{pmatrix} 0.00002 & 0.00001 & -0.00002 \\ 0.00001 & 0.00004 & -0.00001 \\ -0.00002 & -0.00001 & 0.00005 \end{pmatrix} \right]$$

Ahora se debe obtener la distribución a priori de los parámetros en $t = 20$. Es importante recalcar que en este momento aún no se ha observado Y_{20} . Además, la matriz de sistema \mathbf{G}_{20} y la matriz de covarianza \mathbf{W}_{20} son conocidas. A continuación se muestra la implementación en R:

```
#A. Posterior en t=19
m19 <- c(8, 0.35, -0.27)
C19 <- matrix(c(0.00002, 0.00001, -0.00002, 0.00001, 0.00004,
               -0.00001, -0.00002, -0.00001, 0.00005),
              ncol = 3)

#Valores conocidos de G20 y W20
G20 <- matrix(c(1.05, 0, 0, 0, 1.02, 0, 0, 0, 0.99), ncol=3)
W20 <- matrix(c(0.00001, 0, 0, 0, 0.0001, -0.00001, 0,
               -0.00001, 0.00005), ncol=3)

#B. Priori de parámetros en t=20
a20 <- G20 %*% m19
```

```
R20 <- G20 %*% C19 %*% t(G20) + W20
```

```
a20
```

```
##           [,1]
## [1,]  8.4000
## [2,]  0.3570
## [3,] -0.2673
```

```
R20
```

```
##           [,1]           [,2]           [,3]
## [1,]  3.205e-05  1.071e-05 -2.079e-05
## [2,]  1.071e-05  1.416e-04 -2.010e-05
## [3,] -2.079e-05 -2.010e-05  9.901e-05
```

Es decir,

$$(\theta_{20} \mid D_{19}) \sim \mathbf{N} \left[\begin{pmatrix} 8.40 \\ 0.3570 \\ -0.2673 \end{pmatrix}, \begin{pmatrix} 0.000032 & 0.000011 & -0.000021 \\ 0.000011 & 0.00014 & -0.000020 \\ -0.000021 & -0.000020 & 0.000099 \end{pmatrix} \right]$$

Posteriormente, se pronostica Y_{20} . Para esto, la matriz \mathbf{F}_{20} y la varianza V_{20} son conocidas. En este caso, $\mathbf{F}_{20} = [1, \text{Ingreso en } t=20, \text{Precio en } t=20]^T$

```
F20 <- c(1, 6.06093, 4.51018) #Variables explicativas en t=20.
V20 <- 0.002
```

```
#C. Pronóstico a un periodo.
```

```
f20 <- t(F20) %*% a20
Q20 <- t(F20) %*% R20 %*% F20 + V20
```

```
f20
```

```
Q20
```

```
##          [,1]
## [1,]  9.358
```

```
##          [,1]
## [1,]  0.008092
```

Entonces,

$$(Y_{20} \mid D_{19}) \sim \mathbf{N}[9.358, 0.0081]$$

Con esta distribución, es posible calcular un intervalo de pronóstico del 95 % para Y_{20} :

```
c(qnorm(0.025, mean = f20, sd = sqrt(Q20)),
  qnorm(0.975, mean = f20, sd = sqrt(Q20)))
```

```
## [1]  9.182  9.534
```

Supongamos que, unos meses después, se observan las ganancias en 1967 C1 (es decir, el valor de Y_{20}). Se observa que $Y_{20} = 9.31378$. Por lo tanto, las ventas observadas están dentro del intervalo de pronóstico obtenido anteriormente. Gracias a que ahora se conoce el valor de Y_{20} , ya se tiene $D_{20} = \{D_{19}, Y_{20}\}$, entonces podemos actualizar la distribución de parámetros de $t = 20$.

#Valor observado de Y20:

```
Y20 <- 9.31378
```

#D. Posterior en t=20

```
A20 = R20 %*% F20 %*% solve(Q20)
```

```
e20 = Y20-f20
```

```
m20 = a20 + A20 %*% e20
```

```
C20 = R20 - A20 %*% Q20 %*% t(A20)
```

```
m20
```

```
C20
```

```
##          [,1]
## [1,]  8.4000
## [2,]  0.3527
## [3,] -0.2690

##          [,1]      [,2]      [,3]
## [1,]  3.205e-05  1.040e-05 -2.091e-05
## [2,]  1.040e-05  6.674e-05 -4.933e-05
## [3,] -2.091e-05 -4.933e-05  8.759e-05
```

Entonces, la distribución posterior de los parámetros en $t = 20$ es



$$(\theta_{20} \mid D_{20}) \sim \mathbf{N} \left[\begin{pmatrix} 8.4 \\ 0.3527 \\ -0.2690 \end{pmatrix}, \begin{pmatrix} 0.000032 & 0.000011 & -0.000021 \\ 0.00001 & 0.000067 & -0.000049 \\ -0.000021 & -0.000049 & 0.000088 \end{pmatrix} \right]$$

En este ejemplo el error e_t es pequeño, por eso las distribuciones a priori y posterior de los parámetros son muy parecidas.

2.3.2. Distribuciones de pronóstico

Normalmente se desea realizar pronósticos hacia más de un periodo en el futuro, por lo que es necesario encontrar una expresión para calcular pronósticos para varios periodos del futuro.

Definición 2.3.2.1 (Función de pronóstico). La función de pronóstico $f_t(k)$, en cualquier tiempo t , se define para todos los enteros $k \geq 0$ como

$$f_t(k) = E[\mu_{t+k} \mid D_t] = E[\mathbf{F}_{t+k}^T \boldsymbol{\theta}_{t+k} \mid D_t],$$

donde $\mu_t = \mathbf{F}_t^T \boldsymbol{\theta}_t$ es la función de respuesta media para cualquier $t \geq 0$

Para $k > 0$, la función de pronóstico equivale al valor esperado de las observaciones futuras dada la información actual:

$$f_t(k) = E[Y_{t+k} \mid D_t]$$

Para $k = 0$ la ecuación anterior no es válida debido a que, dada la información actual D_t , ya se conoce el valor real de $Y_t = \mathbf{F}_t^T \boldsymbol{\theta}_t + \nu_t$, el cual

nunca será igual que $\mathbf{F}_t^\top \boldsymbol{\theta}_t$ por la presencia del error ν_t con distribución normal.

La definición de la función de pronóstico esta formulada en términos de μ_{t+k} en lugar de Y_{t+k} con el objetivo de permitir el cálculo de la estimación posterior del nivel actual $f_t(0)$. El siguiente teorema proporciona las distribuciones de pronóstico.

Teorema 2.3.2.2. *Para $0 \leq j < k$, en cada tiempo t la distribución conjunta futura es normal y definida por las siguientes covarianzas y las distribuciones marginales a k periodos futuros:*

A. *Distribución de estados:* $(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathbf{N}[\mathbf{a}_t(k), \mathbf{R}_t(k)],$

B. *Distribución de pronóstico:* $(Y_{t+k} \mid D_t) \sim \mathbf{N}[f_t(k), Q_t(k)],$

C. *Covarianzas de estados:* $C[\boldsymbol{\theta}_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] = \mathbf{C}_t(k, j),$

D. *Covarianzas de observaciones:* $C[Y_{t+k}, Y_{t+j} \mid D_t] = \mathbf{F}_{t+k}^\top \mathbf{C}_t(k, j) \mathbf{F}_{t+j},$

E. *Otras covarianzas:*

$$C[\boldsymbol{\theta}_{t+k}, Y_{t+j} \mid D_t] = \mathbf{C}_t(k, j) \mathbf{F}_{t+j}, \quad C[Y_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] = \mathbf{F}_{t+k}^\top \mathbf{C}_t(k, j),$$

donde $f_t(k) = \mathbf{F}_{t+k}^\top \mathbf{a}_t(k)$ y $Q_t(k) = \mathbf{F}_{t+k}^\top \mathbf{R}_t(k) \mathbf{F}_{t+k} + V_{t+k}$. Lo anterior puede ser calculado recursivamente:

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k} \mathbf{a}_t(k-1),$$

$$\mathbf{R}_t(k) = \mathbf{G}_{t+k} \mathbf{R}_t(k-1) \mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k},$$

$$\mathbf{C}_t(k, j) = \mathbf{G}_{t+k} \mathbf{C}_t(k-1, j), \quad k = j+1, \dots,$$

con $\mathbf{a}_t(0) = \mathbf{m}_t$, $\mathbf{R}_t(0) = \mathbf{C}_t$ y $\mathbf{C}_t(j, j) = \mathbf{R}_t(j)$ como valores iniciales

Demostración. [10, p. 107]

Sea $\mathbf{H}_{t+k}(r) = \mathbf{G}_{t+k} \mathbf{G}_{t+k-1} \dots \mathbf{G}_{t+k-r+1}$ para toda t y para $r \leq k$, con $\mathbf{H}_{t+k}(0) = \mathbf{I}$. Tras aplicar la ecuación de sistema varias veces, se nota el

siguiente comportamiento:

Se aplica la ecuación de sistema una vez:

$$\theta_{t+k} = \mathbf{G}_{t+k}\theta_{t+k-1} + \omega_{t+k} = \mathbf{H}_{t+k}(1)\theta_{t+k-1} + \sum_{r=k}^k \mathbf{H}_{t+k}(k-r)\omega_{t+r}$$

Se aplica la ecuación de sistema dos veces:

$$\begin{aligned} \theta_{t+k} &= \mathbf{G}_{t+k}[\mathbf{G}_{t+k-1}\theta_{t+k-2} + \omega_{t+k-1}] + \omega_{t+k} \\ &= \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}\theta_{t+k-2} + \mathbf{G}_{t+k}\omega_{t+k-1} + \omega_{t+k} \\ &= \mathbf{H}_{t+k}(2)\theta_{t+k-2} + \sum_{r=k-1}^k \mathbf{H}_{t+k}(k-r)\omega_{t+r} \end{aligned}$$

Se aplica la ecuación de sistema tres veces:

$$\begin{aligned} \theta_{t+k} &= \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}[\mathbf{G}_{t+k-2}\theta_{t+k-3} + \omega_{t+k-2}] + \mathbf{G}_{t+k}\omega_{t+k-1} + \omega_{t+k} \\ &= \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}\mathbf{G}_{t+k-2}\theta_{t+k-3} + \mathbf{G}_{t+k}\mathbf{G}_{t+k-1}\omega_{t+k-2} + \mathbf{G}_{t+k}\omega_{t+k-1} + \omega_{t+k} \\ &= \mathbf{H}_{t+k}(3)\theta_{t+k-3} + \sum_{r=k-2}^k \mathbf{H}_{t+k}(k-r)\omega_{t+r} \end{aligned}$$

Entonces, tras aplicar la ecuación de sistema k veces, se obtiene

$$\theta_{t+k} = \mathbf{H}_{t+k}(k)\theta_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)\omega_{t+r}$$

Primero se probará la distribución de estados. Por normalidad, se tiene

$$\begin{aligned} E[\theta_{t+k} \mid D_t] &= E\left[\mathbf{H}_{t+k}(k)\theta_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)\omega_{t+r} \mid D_t\right] \\ &= \mathbf{H}_{t+k}(k)E[\theta_t \mid D_t] + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)E[\omega_{t+r} \mid D_t] \xrightarrow{0} \\ &= \mathbf{H}_{t+k}(k)\mathbf{m}_t \quad (\text{por Teorema 2.3.1.1}). \end{aligned}$$

Entonces,

$$\begin{aligned}\mathbf{a}_t(k) &= E[\boldsymbol{\theta}_{t+k} \mid D_t] = \mathbf{H}_{t+k}(k)\mathbf{m}_t = \mathbf{G}_{t+k}\mathbf{H}_{t+k-1}(k-1)\mathbf{m}_t \\ &= \mathbf{G}_{t+k}\mathbf{a}_t(k-1).\end{aligned}$$

En el caso de la varianza, por independencia se tiene

$$\begin{aligned}V[\boldsymbol{\theta}_{t+k} \mid D_t] &= V\left[\mathbf{H}_{t+k}(k)\boldsymbol{\theta}_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)\boldsymbol{\omega}_{t+r} \mid D_t\right] \\ &= \mathbf{H}_{t+k}(k)V[\boldsymbol{\theta}_t \mid D_t]\mathbf{H}_{t+k}(k)^\top + \\ &\quad \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)V[\boldsymbol{\omega}_{t+r} \mid D_t]\mathbf{H}_{t+k}(k-r)^\top \\ &= \mathbf{H}_{t+k}(k)\mathbf{C}_t\mathbf{H}_{t+k}(k)^\top + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)\mathbf{W}_{t+r}\mathbf{H}_{t+k}(k-r)^\top\end{aligned}$$

Entonces,

$$\begin{aligned}\mathbf{R}_t(k) &= V[\boldsymbol{\theta}_{t+k} \mid D_t] \\ &= \mathbf{H}_{t+k}(k)\mathbf{C}_t\mathbf{H}_{t+k}(k)^\top + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r)\mathbf{W}_{t+r}\mathbf{H}_{t+k}(k-r)^\top \\ &= \mathbf{G}_{t+k}\mathbf{H}_{t+k-1}(k-1)\mathbf{C}_t\mathbf{H}_{t+k-1}(k-1)^\top\mathbf{G}_{t+k}^\top + \\ &\quad \mathbf{G}_{t+k}\left(\sum_{r=1}^{k-1} \mathbf{H}_{t+k-1}(k-1-r)\mathbf{W}_{t+r}\mathbf{H}_{t+k-1}(k-1-r)^\top\right)\mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k} \\ &= \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k}\end{aligned}$$

Por lo tanto,

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathbf{N}[\mathbf{a}_t(k), \mathbf{R}_t(k)]$$

Con $\mathbf{a}_t(0) = \mathbf{m}_t$ y $\mathbf{R}_t(0) = \mathbf{C}_t$, que es lo establecido en el Teorema 2.3.1.1.

El siguiente paso es probar la distribución de pronóstico. Usando los supuestos de normalidad e independencia, se obtiene

$$E[Y_{t+k} \mid D_t] = \mathbf{F}_{t+k}^\top E[\boldsymbol{\theta}_{t+k} \mid D_t] + E[\nu_{t+k} \mid D_t] = \mathbf{F}_{t+k}^\top \mathbf{a}_t(k) = f_t(k)$$

$$\begin{aligned} V[Y_{t+k} \mid D_t] &= \mathbf{F}_{t+k}^\top V[\boldsymbol{\theta}_{t+k} \mid D_t] \mathbf{F}_{t+k} + V[\nu_{t+k} \mid D_t] \\ &= \mathbf{F}_{t+k}^\top \mathbf{R}_t(k) \mathbf{F}_{t+k} + V_{t+k} = Q_t(k) \end{aligned}$$

Entonces,

$$(Y_{t+k} \mid D_t) \sim \mathbf{N}[f_t(k), Q_t(k)]$$

Por último, se calculan las covarianzas.

$$\begin{aligned} C[\boldsymbol{\theta}_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] &= C[\mathbf{G}_{t+k} \boldsymbol{\theta}_{t+k-1} + \boldsymbol{\omega}_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] \\ &= \mathbf{G}_{t+k} C[\boldsymbol{\theta}_{t+k-1}, \boldsymbol{\theta}_{t+j} \mid D_t] \\ &= \mathbf{G}_{t+k} \mathbf{G}_{t+k-1} \dots \mathbf{G}_{t+j+1} C[\boldsymbol{\theta}_{t+j}, \boldsymbol{\theta}_{t+j} \mid D_t] \\ &= \mathbf{H}_{t+k}(k-j) V[\boldsymbol{\theta}_{t+j} \mid D_t] \\ &= \mathbf{H}_{t+k}(k-j) \mathbf{R}_t(j) \end{aligned}$$

Entonces,

$$\begin{aligned} \mathbf{C}_t(k, j) &= C[\boldsymbol{\theta}_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] = \mathbf{H}_{t+k}(k-j) \mathbf{R}_t(j) \\ &= \mathbf{G}_{t+k} \mathbf{H}_{t+k-1}(k-1-j) \mathbf{R}_t(j) = \mathbf{G}_{t+k} \mathbf{C}_t(k-1, j) \end{aligned}$$

$$\text{Con } \mathbf{C}_t(j, j) = \mathbf{H}_{t+j}(j-j) \mathbf{R}_t(j) = \mathbf{R}_t(j)$$

$$\begin{aligned} C[Y_{t+k}, Y_{t+j} \mid D_t] &= C[\mathbf{F}_{t+k}^\top \boldsymbol{\theta}_{t+k} + \nu_{t+k}, \mathbf{F}_{t+j}^\top \boldsymbol{\theta}_{t+j} + \nu_{t+j} \mid D_t] \\ &= \mathbf{F}_{t+k}^\top C[\boldsymbol{\theta}_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] \mathbf{F}_{t+j} = \mathbf{F}_{t+k}^\top \mathbf{C}_t(k, j) \mathbf{F}_{t+j} \end{aligned}$$

$$C[\boldsymbol{\theta}_{t+k}, Y_{t+j} \mid D_t] = C[\boldsymbol{\theta}_{t+k}, \mathbf{F}_{t+j}^\top \boldsymbol{\theta}_{t+j} + \nu_{t+j} \mid D_t]$$

$$= C[\boldsymbol{\theta}_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] \mathbf{F}_{t+j} = \mathbf{C}_t(k, j) \mathbf{F}_{t+j}$$

$$\begin{aligned} C[Y_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] &= C[\mathbf{F}_{t+k}^\top \boldsymbol{\theta}_{t+k} + \nu_{t+k}, \boldsymbol{\theta}_{t+j} \mid D_t] \\ &= \mathbf{F}_{t+k}^\top C[\boldsymbol{\theta}_{t+j}, \boldsymbol{\theta}_{t+k} \mid D_t] = \mathbf{F}_{t+k}^\top \mathbf{C}_t(k, j) \end{aligned}$$

□

Ejemplo 2.3.2.1. Se parte del final del ejemplo pasado. Estamos ubicados en 1967 C1, es decir, $t = 20$. Se desea obtener la distribución de los pronósticos de los ingresos en $t = 21$ y en $t = 22$ dada la información actual D_{20} .

Los valores iniciales son la esperanza y la varianza de $(\boldsymbol{\theta}_{20} \mid D_{20})$, obtenidas en la última parte del ejemplo pasado. Después se debe definir la distribución de $(\boldsymbol{\theta}_{21} \mid D_{20})$ y posteriormente se calcula la distribución de $(Y_{21} \mid D_{20})$.

```
#Valores iniciales.
#m20 y C20 se definieron en el ejemplo pasado
a20_0 <- m20
R20_0 <- C20

#k = 1
#Valores conocidos de F21, G21, V21 y W21
F21 <- c(1, 6.071, 4.504) #Variables explicativas en t=21
G21 <- matrix(c(1.01, 0, 0, 0, 1, 0, 0, 0, 0.98), ncol=3)
V21 <- 0.001
W21 <- matrix(c(0.00001, 0, 0, 0, 0.0001, -0.00001, 0,
                -0.00001, 0.00005), ncol=3)

#Distribución de estados en t=21
a20_1 <- G21 %*% a20_0
R20_1 <- G21 %*% R20_0 %*% t(G21) + W21

#Distribución de pronóstico de Y21
```

```
f20_1 <- t(F21) %*% a20_1
Q20_1 <- t(F21) %*% R20_1 %*% F21 + V21
```

```
f20_1
```

```
##           [,1]
## [1,] 9.438
```

```
Q20_1
```

```
##           [,1]
## [1,] 0.006659
```

Entonces, $(Y_{21} \mid D_{20}) \sim N[9.438, 0.006659]$. Ahora, con los resultados de Y_{21} , se obtiene la distribución de pronóstico de Y_{22} .

```
#k = 2
#Valores conocidos de F22, G22, V22 y W22
F22 <- c(1, 6.08, 4.494) #Variables explicativas en t=22
G22 <- matrix(c(1, 0, 0, 0, 1, 0, 0, 0, 0.99), ncol=3)
V22 <- 0.002
W22 <- matrix(c(0.00002, 0, 0, 0, 0.0001, -0.00001, 0,
                -0.00001, 0.00005), ncol=3)
```

```
#Distribución de estados en t=22
a20_2 <- G22 %*% a20_1
R20_2 <- G22 %*% R20_1 %*% t(G22) + W22
```

```
#Distribución de pronóstico de Y22
f20_2 <- t(F22) %*% a20_1
Q20_2 <- t(F22) %*% R20_1 %*% F22 + V22
```

```
f20_2
```

```
Q20_2
```

```
##          [,1]
## [1,] 9.444
```

```
##          [,1]
## [1,] 0.007668
```

Por lo tanto, $(Y_{22} \mid D_{20}) \sim \mathbf{N}[9.444, 0.007668]$. De esta forma, se pueden obtener las distribuciones de pronóstico para más periodos futuros.

2.3.3. Varianzas desconocidas

La teoría desarrollada anteriormente tiene como supuesto que todos los elementos de la cuádrupla $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$ son conocidos. Normalmente \mathbf{F}_t , \mathbf{G}_t y \mathbf{W}_t son definidas por el modelador, pero es común que la varianza de observación V_t sea desconocida y grande en comparación con \mathbf{W}_t . En esta sección se considera el caso en el que la varianza es una constante desconocida, $V_t = V$ para toda t . Su recíproco, $\phi = V^{-1}$, se conoce como precisión de observación.

Definición 2.3.3.1. Para cada t , si $\phi = V^{-1}$, el modelo lineal dinámico se define como

$$\text{Ecuación de observación:} \quad Y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim \mathbf{N}[0, V]$$

$$\text{Ecuación de sistema:} \quad \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathbf{N}[\mathbf{0}, V \mathbf{W}_t^*]$$

$$\text{Información inicial:} \quad (\boldsymbol{\theta}_0 \mid D_0, \phi) \sim \mathbf{N}[\mathbf{m}_0, V \mathbf{C}_0^*],$$

$$(\phi \mid D_0) \sim \mathbf{G} \left[\frac{n_0}{2}, \frac{n_0 S_0}{2} \right].$$

Los valores iniciales \mathbf{m}_0 , \mathbf{C}_0^* , n_0 y S_0 son conocidos, al igual que las matrices \mathbf{F}_t , \mathbf{G}_t y \mathbf{W}_t^* . Es importante notar que las matrices de covarianza \mathbf{C}_0^* y \mathbf{W}_t^* están multiplicadas por V . Esto no implica pérdida de generalidad, ya que cuando V es conocida esta definición es equivalente a la definición

2.3.0.1, donde V está absorbida dentro de las matrices de covarianza \mathbf{C}_0 y \mathbf{W}_t . En el siguiente teorema se muestran las distribuciones para el caso en el que V es desconocida.

Teorema 2.3.3.2. *Se tienen las siguientes distribuciones para el modelo lineal dinámico de la definición 2.3.3.1:*

A. *Condicional en V :*

$$\begin{aligned}(\boldsymbol{\theta}_{t-1} \mid D_{t-1}, V) &\sim \mathbf{N}[\mathbf{m}_{t-1}, V\mathbf{C}_{t-1}^*] \\(\boldsymbol{\theta}_t \mid D_{t-1}, V) &\sim \mathbf{N}[\mathbf{a}_t, V\mathbf{R}_t^*] \\(Y_t \mid D_{t-1}, V) &\sim \mathbf{N}[f_t, VQ_t^*] \\(\boldsymbol{\theta}_t \mid D_t, V) &\sim \mathbf{N}[\mathbf{m}_t, V\mathbf{C}_t^*]\end{aligned}$$

donde

$$\begin{aligned}\mathbf{a}_t &= \mathbf{G}_t \mathbf{m}_{t-1}, & \mathbf{R}_t^* &= \mathbf{G}_t \mathbf{C}_{t-1}^* \mathbf{G}_t^\top + \mathbf{W}_t^* \\f_t &= \mathbf{F}_t^\top \mathbf{a}_t, & Q_t^* &= 1 + \mathbf{F}_t^\top \mathbf{R}_t^* \mathbf{F}_t, \\e_t &= Y_t - f_t, & \mathbf{A}_t &= \mathbf{R}_t^* \mathbf{F}_t / Q_t^*, \\ \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t, & \mathbf{C}_t^* &= \mathbf{R}_t^* - \mathbf{A}_t \mathbf{A}_t^\top Q_t^*.\end{aligned}$$

B. *Distribución de la precisión $\phi = V^{-1}$:*

$$\begin{aligned}(\phi \mid D_{t-1}) &\sim \mathbf{G} \left[\frac{n_{t-1}}{2}, \frac{n_{t-1} S_{t-1}}{2} \right], \\(\phi \mid D_t) &\sim \mathbf{G} \left[\frac{n_t}{2}, \frac{n_t S_t}{2} \right],\end{aligned}$$

$$\text{con } n_t = n_{t-1} + 1, \quad n_t S_t = n_{t-1} S_{t-1} + e_t^2 / Q_t^*.$$

C. *No condicionado en V :*

$$\begin{aligned}(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) &\sim \mathbf{T}_{n_{t-1}}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}], \\(\boldsymbol{\theta}_t \mid D_{t-1}) &\sim \mathbf{T}_{n_{t-1}}[\mathbf{a}_t, \mathbf{R}_t], \\(Y_t \mid D_{t-1}) &\sim \mathbf{T}_{n_{t-1}}[f_t, Q_t], \\(\boldsymbol{\theta}_t \mid D_t) &\sim \mathbf{T}_{n_t}[\mathbf{m}_t, \mathbf{C}_t],\end{aligned}$$

$$\text{con } \mathbf{R}_t = S_{t-1} \mathbf{R}_t^*, \quad Q_t = S_{t-1} Q_t^*, \quad \mathbf{C}_t = S_t \mathbf{C}_t^*.$$

D. Ecuaciones de actualización.

Con $Q_t = \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t + S_{t-1}$ y $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / Q_t$ se tiene

$$\begin{aligned} n_t &= n_{t-1} + 1, & S_t &= S_{t-1} + \frac{S_{t-1}}{n_t} \left(\frac{e_t^2}{Q_t} - 1 \right), \\ \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t & \mathbf{C}_t &= \frac{S_t}{S_{t-1}} (\mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t^\top Q_t). \\ \mathbf{R}_t &= \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}_t \end{aligned}$$

Demostración. El punto A. del teorema es el mismo caso del teorema 2.3.1.1. La única diferencia es la forma en la que están escritas las varianzas, pero son equivalentes. Los puntos B. y C. se probarán por inducción.

Base de inducción. Se demuestra la validez para $t = 1$.

Sea $d_0 = n_0 S_0$. Entonces, por la definición 2.3.3.1 tenemos

$$(\phi \mid D_0) \sim \mathbf{G} \left[\frac{n_0}{2}, \frac{d_0}{2} \right]$$

Entonces,

$$f(\phi \mid D_0) \propto \phi^{\frac{n_0}{2}-1} \exp \left(-\frac{d_0 \phi}{2} \right)$$

En el punto A. se tiene que $(Y_1 \mid D_0, \phi) \sim \mathbf{N}[f_1, \phi^{-1} Q_1^*]$. Entonces,

$$f(Y_1 \mid D_0, \phi) \propto \phi^{\frac{1}{2}} \exp \left(-\frac{e_1^2 \phi}{2 Q_1^*} \right)$$

Por el teorema de Bayes, la distribución posterior de ϕ es

$$\begin{aligned} f(\phi \mid D_1) &\propto f(\phi \mid D_0) f(Y_1 \mid D_0, \phi) \propto \phi^{\frac{n_0}{2}-\frac{1}{2}} \exp \left(-\frac{d_0 \phi}{2} - \frac{e_1^2 \phi}{2 Q_1^*} \right) \\ &= \phi^{\frac{n_1}{2}-1} \exp(-\phi[d_0 + e_1^2/Q_1^*]/2) = \phi^{\frac{n_1}{2}-1} \exp(-d_1 \phi/2) \\ &= \phi^{\frac{n_1}{2}-1} \exp(-n_1 S_1 \phi/2), \end{aligned}$$

donde $n_1 = n_0 + 1$, $d_1 = d_0 + e_1^2/Q_1^*$. Por lo tanto,

$$(\phi \mid D_1) \sim \mathbf{G} \left[\frac{n_1}{2}, \frac{n_1 S_1}{2} \right]$$

Notemos que $n_1 S_1 = d_1 = n_0 S_0 + e_1^2 / Q_1^*$.

Para el punto C., se necesitan los resultados del apéndice A.3. Para obtener las distribuciones con ϕ desconocida, se deben marginalizar las distribuciones de el punto A. utilizando la distribución apropiada de ϕ , como se muestra en el apéndice. Las distribuciones sin condicionar en ϕ son t de Student.

Por la definición 2.3.3.1 se tiene

$$(\phi \mid D_0) \sim \mathbf{G} \left[\frac{n_0}{2}, \frac{n_0 S_0}{2} \right].$$

Como $(\theta_0 \mid D_0, \phi) \sim \mathbf{N}[\mathbf{m}_0, \phi^{-1} \mathbf{C}_0^*]$, con los resultados del apéndice A.3 se obtiene

$$(\theta_0 \mid D_0) \sim \mathbf{T}_{n_0}[\mathbf{m}_0, \mathbf{C}_0],$$

con $\mathbf{C}_0 = S_0 \mathbf{C}_0^*$.

Para la distribución a priori, $(\theta_1 \mid D_0, \phi) \sim \mathbf{N}[\mathbf{a}_1, \phi^{-1} \mathbf{R}_1^*]$, entonces

$$(\theta_1 \mid D_0) \sim \mathbf{T}_{n_0}[\mathbf{a}_1, \mathbf{R}_1],$$

con $\mathbf{R}_1 = S_0 \mathbf{R}_1^*$

Para el pronóstico a un periodo, $(Y_1 \mid D_0, \phi) \sim \mathbf{N}[f_1, \phi^{-1} Q_1^*]$, entonces

$$(Y_1 \mid D_0) \sim \mathbf{T}_{n_0}[f_1, Q_1],$$

con $Q_1 = S_0 Q_1^*$.

Finalmente, una vez que se observa Y_1 , entonces se obtiene la distribución posterior. Como ya se cuenta con D_1 , entonces la distribución de ϕ es

$$(\phi \mid D_1) \sim \mathbf{G} \left[\frac{n_1}{2}, \frac{n_1 S_1}{2} \right].$$

Como $(\theta_1 \mid D_1, \phi) \sim \mathbf{N}[\mathbf{m}_1, \phi^{-1} \mathbf{C}_1^*]$ entonces,

$$(\theta_1 \mid D_1) \sim \mathbf{T}_{n_1}[\mathbf{m}_1, \mathbf{C}_1],$$

con $\mathbf{C}_1 = S_1 \mathbf{C}_1^*$.

Para el punto D. basta con desarrollar las ecuaciones obtenidas anteriormente. Los valores iniciales \mathbf{m}_0 , \mathbf{C}_0^* , n_0 , S_0 son conocidos.

Recordemos que $n_1 S_1 = n_0 S_0 + e_1^2 / Q_1^*$, entonces

$$S_1 = \frac{n_0 S_0}{n_1} + \frac{e_1^2}{Q_1^* n_1} = \frac{(n_1 - 1) S_0 + \frac{e_1^2 S_0}{Q_1}}{n_1} = S_0 + \frac{S_0}{n_1} \left(\frac{e_1^2}{Q_1} - 1 \right)$$

Por el punto A.,

$$\mathbf{A}_1 = \mathbf{R}_1^* \mathbf{F}_1 / Q_1^* = \frac{\frac{\mathbf{R}_1}{S_0} \mathbf{F}_1}{\frac{Q_1}{S_0}} = \mathbf{R}_1 \mathbf{F}_1 / Q_1$$

Por el punto C.,

$$Q_1 = S_0 Q_1^* = S_0 [1 + \mathbf{F}_1^\top \mathbf{R}_1^* \mathbf{F}_1] = S_0 + \mathbf{F}_1^\top \mathbf{R}_1 \mathbf{F}_1$$

$$\mathbf{C}_1 = S_1 \mathbf{C}_1^* = S_1 [\mathbf{R}_1^* - \mathbf{A}_1 \mathbf{A}_1^\top Q_1^*] = \frac{S_1}{S_0} [\mathbf{R}_1 - \mathbf{A}_1 \mathbf{A}_1^\top Q_1]$$

Sea $\mathbf{W}_1 = S_0 \mathbf{W}_1^*$, entonces

$$\begin{aligned} \mathbf{R}_1 &= S_0 \mathbf{R}_1^* = S_0 [\mathbf{G}_1 \mathbf{C}_0^* \mathbf{G}_1^\top + \mathbf{W}_1^*] = S_0 \left[\mathbf{G}_1 \frac{\mathbf{C}_0}{S_0} \mathbf{G}_1^\top + \frac{\mathbf{W}_1}{S_0} \right] \\ &= \mathbf{G}_1 \mathbf{C}_0 \mathbf{G}_1^\top + \mathbf{W}_1 \end{aligned}$$

Se supone que el teorema es válido para t . Se probará la validez para $t + 1$

El punto A. es equivalente al teorema 2.3.1.1.

Sea $d_t = n_t S_t$. Entonces, por hipótesis de inducción tenemos

$$(\phi \mid D_t) \sim \mathbf{G} \left[\frac{n_t}{2}, \frac{d_t}{2} \right]$$

Entonces,

$$f(\phi \mid D_t) \propto \phi^{\frac{n_t}{2}-1} \exp\left(-\frac{d_t \phi}{2}\right)$$

En el punto A. se tiene que $(Y_{t+1} \mid D_t, \phi) \sim \mathbf{N}[f_{t+1}, \phi^{-1}Q_{t+1}^*]$. Entonces,

$$f(Y_{t+1} \mid D_t, \phi) \propto \phi^{\frac{1}{2}} \exp\left(-\frac{e_{t+1}^2 \phi}{2Q_{t+1}^*}\right)$$

Por el teorema de Bayes, la distribución posterior de ϕ es

$$\begin{aligned} f(\phi \mid D_{t+1}) &\propto f(\phi \mid D_t) f(Y_{t+1} \mid D_t, \phi) \propto \phi^{\frac{n_t}{2}-\frac{1}{2}} \exp\left(-\frac{d_t \phi}{2} - \frac{e_{t+1}^2 \phi}{2Q_{t+1}^*}\right) \\ &= \phi^{\frac{n_{t+1}}{2}-1} \exp(-\phi[d_t + e_{t+1}^2/Q_{t+1}^*]/2) = \phi^{\frac{n_{t+1}}{2}-1} \exp(-d_{t+1} \phi/2) \\ &= \phi^{\frac{n_{t+1}}{2}-1} \exp(-n_{t+1} S_{t+1} \phi/2), \end{aligned}$$

donde $n_{t+1} = n_t + 1$, $d_{t+1} = d_t + e_{t+1}^2/Q_{t+1}^*$. Por lo tanto,

$$(\phi \mid D_{t+1}) \sim \mathbf{G}\left[\frac{n_{t+1}}{2}, \frac{n_{t+1} S_{t+1}}{2}\right]$$

Notemos que $n_{t+1} S_{t+1} = d_{t+1} = n_t S_t + e_{t+1}^2/Q_{t+1}^*$.

Para el punto C., se necesitan los resultados del apéndice A.3.

Como $(\theta_t \mid D_t, \phi) \sim \mathbf{N}[\mathbf{m}_t, \phi^{-1}\mathbf{C}_t^*]$, con los resultados del apéndice A.3 se obtiene

$$(\theta_t \mid D_t) \sim \mathbf{T}_{n_t}[\mathbf{m}_t, \mathbf{C}_t],$$

con $\mathbf{C}_t = S_t \mathbf{C}_t^*$.

Para la distribución a priori, $(\theta_{t+1} \mid D_t, \phi) \sim \mathbf{N}[\mathbf{a}_{t+1}, \phi^{-1}\mathbf{R}_{t+1}^*]$, entonces

$$(\theta_{t+1} \mid D_t) \sim \mathbf{T}_{n_t}[\mathbf{a}_{t+1}, \mathbf{R}_{t+1}],$$

con $\mathbf{R}_{t+1} = S_t \mathbf{R}_{t+1}^*$

Para el pronóstico a un periodo, $(Y_{t+1} \mid D_t, \phi) \sim \mathbf{N}[f_{t+1}, \phi^{-1}Q_{t+1}^*]$, entonces

$$(Y_{t+1} \mid D_t) \sim \mathbf{T}_{n_t}[f_{t+1}, Q_{t+1}],$$

con $Q_{t+1} = S_t Q_{t+1}^*$.

Finalmente, una vez que se observa Y_{t+1} , entonces se obtiene la distribución posterior. Como ya se cuenta con D_{t+1} , entonces la distribución de ϕ es

$$(\phi \mid D_{t+1}) \sim \mathbf{G}\left[\frac{n_{t+1}}{2}, \frac{n_{t+1}S_{t+1}}{2}\right].$$

Como $(\boldsymbol{\theta}_{t+1} \mid D_{t+1}, \phi) \sim \mathbf{N}[\mathbf{m}_{t+1}, \phi^{-1}\mathbf{C}_{t+1}^*]$ entonces,

$$(\boldsymbol{\theta}_{t+1} \mid D_{t+1}) \sim \mathbf{T}_{n_{t+1}}[\mathbf{m}_{t+1}, \mathbf{C}_{t+1}],$$

con $\mathbf{C}_{t+1} = S_{t+1}\mathbf{C}_{t+1}^*$.

Para el punto D. basta con desarrollar las ecuaciones obtenidas anteriormente.

Recordemos que $n_t S_t = n_{t-1} S_{t-1} + e_t^2 / Q_t^*$, entonces

$$S_t = \frac{n_{t-1} S_{t-1}}{n_t} + \frac{e_t^2}{Q_t^* n_t} = \frac{(n_t - 1) S_{t-1} + \frac{e_t^2 S_{t-1}}{Q_t^*}}{n_t} = S_{t-1} + \frac{S_{t-1}}{n_t} \left(\frac{e_t^2}{Q_t^*} - 1 \right)$$

Por el punto A.,

$$\mathbf{A}_t = \mathbf{R}_t^* \mathbf{F}_t / Q_t^* = \frac{\frac{\mathbf{R}_t}{S_{t-1}} \mathbf{F}_t}{\frac{Q_t}{S_{t-1}}} = \mathbf{R}_t \mathbf{F}_t / Q_t$$

Por el punto C.,

$$Q_t = S_{t-1} Q_t^* = S_{t-1} [1 + \mathbf{F}_t^\top \mathbf{R}_t^* \mathbf{F}_t] = S_{t-1} + \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t$$

$$\mathbf{C}_t = S_t \mathbf{C}_t^* = S_t [\mathbf{R}_t^* - \mathbf{A}_t \mathbf{A}_t^\top Q_t^*] = \frac{S_t}{S_{t-1}} [\mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t^\top Q_t]$$

Sea $\mathbf{W}_{t+1} = S_t \mathbf{W}_{t+1}^*$, entonces

$$\begin{aligned} \mathbf{R}_{t+1} &= S_t \mathbf{R}_{t+1}^* = S_t [\mathbf{G}_{t+1} \mathbf{C}_t^* \mathbf{G}_{t+1}^\top + \mathbf{W}_{t+1}^*] = S_t \left[\mathbf{G}_{t+1} \frac{\mathbf{C}_t}{S_t} \mathbf{G}_{t+1}^\top + \frac{\mathbf{W}_{t+1}}{S_t} \right] \\ &= \mathbf{G}_{t+1} \mathbf{C}_t \mathbf{G}_{t+1}^\top + \mathbf{W}_{t+1} \end{aligned}$$

□

En la práctica no es necesario calcular \mathbf{R}_t^* , Q_t^* , y \mathbf{C}_t^* ; solo se muestran en el teorema para ilustrar que la diferencia entre el caso con varianza conocida y el de varianza desconocida. Cuando V es conocida, entonces $\mathbf{R}_t = V \mathbf{R}_t^*$, $Q_t = V Q_t^*$, y $\mathbf{C}_t = V \mathbf{C}_t^*$. Si V es desconocida, $\mathbf{R}_t = S_{t-1} \mathbf{R}_t^*$, $Q_t = S_{t-1} Q_t^*$, y $\mathbf{C}_t = S_t \mathbf{C}_t^*$. S_{t-1} es una estimación de V con la información D_{t-1} , mientras que S_t es la estimación de V actualizada una vez que se observa Y_t . La otra diferencia es que la actualización de \mathbf{C}_t está escalada por $\frac{S_t}{S_{t-1}}$ para corregir con la estimación actualizada de V .

Ejemplo 2.3.3.1. Para comparar los resultados cuando V es conocida y cuando es desconocida, supongamos que estamos en $t = 19$ y se tienen los mismos valores y el mismo contexto que en el ejemplo 2.3.1.1.

Si $n_{19} = 19.5$, la distribución posterior de los parámetros en $t = 19$ es de la siguiente forma:

$$(\boldsymbol{\theta}_{19} \mid D_{19}) \sim \mathbf{T}_{19.5} \left[\begin{pmatrix} 8 \\ 0.35 \\ -0.27 \end{pmatrix}, \begin{pmatrix} 0.00002 & 0.00001 & -0.00002 \\ 0.00001 & 0.00004 & -0.00001 \\ -0.00002 & -0.00001 & 0.00005 \end{pmatrix} \right]$$

Esta distribución tiene los mismos valores que la del ejemplo pasado, la diferencia es que como en este caso asumimos que V es desconocida, entonces la distribución es t de Student. Los valores de \mathbf{G}_{20} y \mathbf{W}_{20} son conocidos e iguales a los utilizados en el ejemplo anterior. Ahora se debe obtener la distribución a priori de los parámetros en $t = 20$. A continuación se muestra la implementación en R:

```
#A. Posterior en t=19
m19 <- c(8, 0.35, -0.27)
```

```
C19 <- matrix(c(0.00002, 0.00001, -0.00002, 0.00001, 0.00004,
               -0.00001, -0.00002, -0.00001, 0.00005),
              ncol = 3)
```

```
#Valores conocidos de G20 y W20
```

```
G20 <- matrix(c(1.05, 0, 0, 0, 1.02, 0, 0, 0, 0.99), ncol=3)
```

```
W20 <- matrix(c(0.00001, 0, 0, 0, 0.0001, -0.00001, 0,
               -0.00001, 0.00005), ncol=3)
```

```
#B. Priori de parámetros en t=20
```

```
a20 <- G20 %*% m19
```

```
R20 <- G20 %*% C19 %*% t(G20) + W20
```

```
a20
```

```
##           [,1]
## [1,]  8.4000
## [2,]  0.3570
## [3,] -0.2673
```

```
R20
```

```
##           [,1]      [,2]      [,3]
## [1,]  3.205e-05  1.071e-05 -2.079e-05
## [2,]  1.071e-05  1.416e-04 -2.010e-05
## [3,] -2.079e-05 -2.010e-05  9.901e-05
```

Es decir,

$$(\boldsymbol{\theta}_{20} \mid D_{19}) \sim \mathbf{T}_{19.5} \left[\begin{pmatrix} 8.40 \\ 0.3570 \\ -0.2673 \end{pmatrix}, \begin{pmatrix} 0.000032 & 0.000011 & -0.000021 \\ 0.000011 & 0.00014 & -0.000020 \\ -0.000021 & -0.000020 & 0.000099 \end{pmatrix} \right],$$

que también tiene los mismos valores que en el ejemplo anterior, pero con distribución t de Student.

El siguiente paso es obtener la distribución de Y_{20} condicionada en D_{19} . En el ejemplo anterior se necesitaba el valor de V para calcular la varianza de la distribución. En este caso en lugar de V se utiliza su estimación en $t = 19$, S_{19} . En el ejemplo anterior $V = 0.002$, aquí supongamos que S_{19} fue obtenida anteriormente y su valor es $S_{19} = 0.002$ con el fin de ilustrar lo parecidos que son ambos casos.

```
F20 <- c(1, 6.06093, 4.51018) # Variables explicativas en t=20.
S19 <- 0.002 # Estimación de V en T=19
n19 <- 19.5 # Grados de libertad
#C. Pronóstico a un periodo.
f20 <- as.numeric(t(F20) %*% a20)
Q20 <- as.numeric(t(F20) %*% R20 %*% F20 + S19)
f20
```

```
## [1] 9.358
```

```
Q20
```

```
## [1] 0.008092
```

Entonces,

$$(Y_{20} | D_{19}) \sim \mathbf{T}_{19.5}[9.358, 0.0081]$$

Ahora es posible calcular el intervalo de pronóstico del 95 % para Y_{20} :

```
c(qst(0.025, nu = n19, mu = f20, sigma = sqrt(Q20)),
  qst(0.975, nu = n19, mu = f20, sigma = sqrt(Q20)))
```

```
## [1] 9.170 9.546
```

Este intervalo es ligeramente más amplio que el del ejemplo anterior debido a que la distribución en este caso es una t de Student. Entre más

periodos pasen, los grados de libertad aumentarán y la t de Student convergerá a una distribución normal. Sin embargo, el centro del intervalo es el mismo.

Al igual que en el ejemplo anterior, unos meses después se observa que $Y_{20} = 9.31378$. Ahora que se conoce el valor de Y_{20} , ya se tiene $D_{20} = \{D_{19}, Y_{20}\}$, entonces podemos actualizar la distribución de parámetros de $t = 20$. Para esto, primero se tiene que actualizar la estimación de V en $t = 20$, es decir, se debe calcular S_{20} .

```
#Valor observado de Y20:
Y20 <- 9.31378
#D. Posterior en t=20
A20 <- R20 %*% F20 / Q20
e20 <- Y20-f20
m20 <- a20 + A20 %*% e20
n20 <- n19 + 1
S20 <- S19 + (S19/n20)*(e20^2/Q20-1) # Actualiza la estimación
                                     # de V
C20 = (S20/S19)*(R20-A20 %*% t(A20) * Q20)

m20

##           [,1]
## [1,]  8.4000
## [2,]  0.3527
## [3,] -0.2690

C20

##           [,1]           [,2]           [,3]
## [1,]  3.087e-05  1.002e-05 -2.014e-05
## [2,]  1.002e-05  6.428e-05 -4.751e-05
## [3,] -2.014e-05 -4.751e-05  8.436e-05
```

Entonces, la distribución posterior de los parámetros en $t = 20$ es

$$(\boldsymbol{\theta}_{20} \mid D_{20}) \sim \mathbf{T}_{20.5} \left[\begin{pmatrix} 8.4 \\ 0.3527 \\ -0.2690 \end{pmatrix}, \begin{pmatrix} 0.000031 & 0.00001 & -0.00002 \\ 0.00001 & 0.000064 & -0.000048 \\ -0.00002 & -0.000048 & 0.000084 \end{pmatrix} \right]$$

En este caso, el parámetro locacional es igual a la esperanza del ejemplo anterior, pero el parámetro de escala es ligeramente distinto comparado con la varianza del ejemplo pasado. Esto es porque esta vez se utiliza la estimación actualizada en los cálculos y no el valor verdadero de V .

Para realizar pronósticos a más de un periodo en el futuro cuando la varianza V es desconocida el procedimiento es muy parecido al caso de V conocida, la diferencia es que las distribuciones son t de Student y que en el cálculo de $Q_t(k)$ se utiliza S_t en vez de V_{t+k} . Recordemos que se considera que la varianza es constante a lo largo del tiempo, por eso se usa la estimación S_t para todas las k futuras.

Teorema 2.3.3.3. *Con el modelo de la definición 2.3.3.1 se tienen las siguientes distribuciones para $k > 0$:*

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathbf{T}_{n_t}[\mathbf{a}_t(k), \mathbf{R}_t(k)]$$

$$(Y_{t+k} \mid D_t) \sim \mathbf{T}_{n_t}[f_t(k), Q_t(k)],$$

con $f_t(k) = \mathbf{F}_{t+k}^\top \mathbf{a}_t(k)$ y $Q_t(k) = \mathbf{F}_{t+k}^\top \mathbf{R}_t(k) \mathbf{F}_{t+k} + S_t$. Esto se calcula recursivamente:

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k} \mathbf{a}_t(k-1),$$

$$\mathbf{R}_t(k) = \mathbf{G}_{t+k} \mathbf{R}_t(k-1) \mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k},$$

Demostración. Por el teorema 2.3.3.2, la distribución actualizada de ϕ en t es

$$(\phi \mid D_t) \sim \mathbf{G} \left[\frac{n_t}{2}, \frac{n_t S_t}{2} \right].$$

Al igual que en el teorema 2.3.3.2, factorizamos V de las distribuciones cuando esta es conocida (las del teorema 2.3.2.2).

Sea $\mathbf{R}_t^*(k) = \mathbf{G}_{t+k}\mathbf{R}_t^*(k-1)\mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k}^*$, con $\mathbf{W}_{t+k} = S_t\mathbf{W}_{t+k}^*$. Entonces,

$$(\boldsymbol{\theta}_{t+k} \mid D_t, \phi) \sim \mathbf{N}[\mathbf{a}_t(k), \phi^{-1}\mathbf{R}_t^*(k)].$$

Por el apéndice A.3 se obtiene la distribución cuando ϕ es desconocida:

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathbf{T}_{n_t}[\mathbf{a}_t(k), \mathbf{R}_t(k)],$$

$$\text{con } \mathbf{R}_t(k) = \mathbf{R}_t^*(k)S_t = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k}.$$

Sea $Q^*(k)_t = \mathbf{F}_t^\top \mathbf{R}^*(k)_t \mathbf{F}_t + 1$. Entonces para los pronósticos tenemos

$$(Y_{t+k} \mid D_t, \phi) \sim \mathbf{N}[f_t(k), \phi^{-1}Q_t^*(k)].$$

Entonces, la distribución cuando ϕ es desconocida es

$$(Y_{t+k} \mid D_t) \sim \mathbf{T}_{n_t}[f_t(k), Q_t(k)],$$

$$\text{con } Q_t(k) = Q_t^*(k)S_t = \mathbf{F}_t^\top \mathbf{R}(k)_t \mathbf{F}_t + S_t. \quad \square$$

2.3.4. Análisis retrospectivo

Normalmente es de interés analizar lo ocurrido en el pasado para conocer mejor el fenómeno que se desea modelar, lo que permite tener un mejor entendimiento de lo que puede ocurrir en el futuro. Para esto se hacen inferencias sobre los vectores de estados pasados $\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}, \dots$. Entonces, el objetivo es obtener las distribuciones marginales retrospectivas $(\boldsymbol{\theta}_{t-k} \mid D_t)$. A la realización de inferencias sobre valores anteriores del vector de estados mediante el empleo de información actual se le conoce como filtración. A $(\boldsymbol{\theta}_{t-k} \mid D_t)$, con $k \geq 1$ se le conoce como distribución filtrada k pasos del vector de estados en el tiempo t . Las distribuciones filtradas pueden ser obtenidas recursivamente y, como se muestra en el siguiente teorema, los resultados son parecidos a lo obtenido en las distribuciones de pronóstico.

Teorema 2.3.4.1. *En el modelo lineal dinámico univariado $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$, para toda t sea*

$$\mathbf{B}_t = \mathbf{C}_t \mathbf{G}_{t+1}^\top \mathbf{R}_{t+1}^{-1}.$$

Entonces para toda k , $1 \leq k < t$, la distribución filtrada k pasos del vector de estados en el tiempo t es

$$(\boldsymbol{\theta}_{t-k} \mid D_t) \sim \mathbf{N}[\mathbf{a}_t(-k), \mathbf{R}_t(-k)],$$

donde

$$\mathbf{a}_t(-k) = \mathbf{m}_{t-k} + \mathbf{B}_{t-k}[\mathbf{a}_t(-k+1) - \mathbf{a}_{t-k+1}]$$

$$\mathbf{R}_t(-k) = \mathbf{C}_{t-k} + \mathbf{B}_{t-k}[\mathbf{R}_t(-k+1) - \mathbf{R}_{t-k+1}]\mathbf{B}_{t-k}^\top$$

con valores iniciales

$$\mathbf{a}_t(0) = \mathbf{m}_t, \quad \mathbf{R}_t(0) = \mathbf{C}_t$$

Demostración. La demostración es por inducción.

Base de inducción. Se demuestra la validez para $k = 1$.

Por definición de probabilidad condicional,

$$f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t) = \frac{f(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t \mid D_t)}{f(\boldsymbol{\theta}_t \mid D_t)},$$

entonces,

$$f(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t \mid D_t) = f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t)f(\boldsymbol{\theta}_t \mid D_t).$$

Con esto se obtiene que

$$f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t) = \int f(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t \mid D_t) d\boldsymbol{\theta}_t = \int f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t) f(\boldsymbol{\theta}_t \mid D_t) d\boldsymbol{\theta}_t$$

Entonces, para obtener $f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t)$ primero hay que determinar $f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t)$ y $f(\boldsymbol{\theta}_t \mid D_t)$. En el teorema 2.3.1.1 se vio que

$$(\boldsymbol{\theta}_t \mid D_t) \sim \mathbf{N}[\mathbf{m}_t, \mathbf{C}_t].$$

En el caso de $f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t)$, notemos que por teorema de Bayes

$$\begin{aligned} f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t) &= f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, Y_t, D_{t-1}) \\ &= \frac{f(Y_t \mid \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t, D_{t-1})f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_{t-1})}{f(Y_t \mid \boldsymbol{\theta}_t, D_{t-1})} \\ &= f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_{t-1}) \end{aligned}$$

debido a que $f(Y_t \mid \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t, D_{t-1}) = f(Y_t \mid \boldsymbol{\theta}_t, D_{t-1})$ porque, dado $\boldsymbol{\theta}_t$, Y_t es independiente de $\boldsymbol{\theta}_{t-1}$.

Por teorema de Bayes,

$$f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_{t-1}) \propto f(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, D_{t-1})f(\boldsymbol{\theta}_{t-1} \mid D_{t-1})$$

Por definición del modelo lineal dinámico y por 2.3.1.1 se sabe que

$$\begin{aligned} (\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, D_{t-1}) &\sim \mathbf{N}[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t] \\ (\boldsymbol{\theta}_{t-1} \mid D_{t-1}) &\sim \mathbf{N}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]. \end{aligned}$$

Con estas distribuciones y utilizando los resultados del apéndice A.2 se obtiene la siguiente distribución conjunta:

$$\begin{pmatrix} \boldsymbol{\theta}_t \\ \boldsymbol{\theta}_{t-1} \end{pmatrix} \Big| D_{t-1} \sim \mathbf{N} \left[\begin{pmatrix} \mathbf{G}_t \mathbf{m}_{t-1} \\ \mathbf{m}_{t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{G}_t^\top \mathbf{C}_{t-1} \mathbf{G}_t + \mathbf{W}_t & \mathbf{G}_t \mathbf{C}_{t-1} \\ \mathbf{C}_{t-1} \mathbf{G}_t^\top & \mathbf{C}_{t-1} \end{pmatrix} \right].$$

Entonces se tiene la siguiente distribución condicional,

$$(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_{t-1}) \sim \mathbf{N}[\mathbf{h}_t(1), \mathbf{H}_t(1)],$$

con

$$\begin{aligned} \mathbf{h}_t(1) &= \mathbf{m}_{t-1} + \mathbf{C}_{t-1} \mathbf{G}_t^\top [\mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}_t]^{-1} [\boldsymbol{\theta}_t - \mathbf{G}_t \mathbf{m}_{t-1}] \\ &= \mathbf{m}_{t-1} + \mathbf{C}_{t-1} \mathbf{G}_t^\top \mathbf{R}_t^{-1} [\boldsymbol{\theta}_t - \mathbf{a}_t] \end{aligned}$$

$$\mathbf{H}_t(1) = \mathbf{C}_{t-1} - \mathbf{C}_{t-1} \mathbf{G}_t^\top [\mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}_t]^{-1} \mathbf{G}_t \mathbf{C}_{t-1}$$

$$= \mathbf{C}_{t-1} + \mathbf{C}_{t-1} \mathbf{G}_t^\top \mathbf{R}_t^{-1} \mathbf{G}_t \mathbf{C}_{t-1}$$

Sea $\mathbf{B}_{t-1} = \mathbf{C}_{t-1} \mathbf{G}_t^\top \mathbf{R}_t^{-1}$, entonces

$$\mathbf{h}_t(1) = \mathbf{m}_{t-1} + \mathbf{B}_{t-1}[\boldsymbol{\theta}_t - \mathbf{a}_t]$$

$$\mathbf{H}_t(1) = \mathbf{C}_{t-1} - \mathbf{B}_{t-1} \mathbf{R}_t \mathbf{B}_{t-1}^\top$$

Como vimos anteriormente, $f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t) = f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_{t-1})$, entonces

$$(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t) \sim \mathbf{N}[\mathbf{h}_t(1), \mathbf{H}_t(1)].$$

Como $(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t)$ y $(\boldsymbol{\theta}_t \mid D_t)$ se distribuyen normal, entonces $(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t)$ también se distribuye normal. Por lo tanto, ahora es necesario encontrar la esperanza y la varianza.

$$\begin{aligned} \mathbf{a}_t(-1) &= E[\boldsymbol{\theta}_{t-1} \mid D_t] = E[E[\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t] \mid D_t] = E[\mathbf{h}_t(1) \mid D_t] \\ &= \mathbf{m}_{t-1} + \mathbf{B}_{t-1}[E[\boldsymbol{\theta}_t \mid D_t] - \mathbf{a}_t] = \mathbf{m}_{t-1} + \mathbf{B}_{t-1}[\mathbf{a}_t(0) - \mathbf{a}_t] \end{aligned}$$

$$\begin{aligned} \mathbf{R}_t(-1) &= V[\boldsymbol{\theta}_{t-1} \mid D_t] = E[V[\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t] \mid D_t] + V[E[\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_t, D_t] \mid D_t] \\ &= E[\mathbf{H}_t(1) \mid D_t] + V[\mathbf{h}_t(1) \mid D_t] \\ &= \mathbf{C}_{t-1} - \mathbf{B}_{t-1} \mathbf{R}_t \mathbf{B}_{t-1}^\top + \mathbf{B}_{t-1} V[\boldsymbol{\theta}_t \mid D_t] \mathbf{B}_{t-1}^\top \\ &= \mathbf{C}_{t-1} + \mathbf{B}_{t-1}[\mathbf{R}_t(0) - \mathbf{R}_t] \mathbf{B}_{t-1}^\top. \end{aligned}$$

Entonces,

$$(\boldsymbol{\theta}_{t-1} \mid D_t) \sim \mathbf{N}[\mathbf{a}_t(-1), \mathbf{R}_t(-1)]$$

Se supone que el teorema es válido para k . Se probará la validez para $k+1$

Por definición de probabilidad condicional,

$$f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t) = \frac{f(\boldsymbol{\theta}_{t-k-1}, \boldsymbol{\theta}_{t-k} \mid D_t)}{f(\boldsymbol{\theta}_{t-k} \mid D_t)},$$

entonces,

$$f(\boldsymbol{\theta}_{t-k-1}, \boldsymbol{\theta}_{t-k} \mid D_t) = f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t) f(\boldsymbol{\theta}_{t-k} \mid D_t).$$

Por lo tanto,

$$\begin{aligned} f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t) &= \int f(\boldsymbol{\theta}_{t-k-1}, \boldsymbol{\theta}_{t-k} \mid D_t) d\boldsymbol{\theta}_{t-k} \\ &= \int f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t) f(\boldsymbol{\theta}_{t-k} \mid D_t) d\boldsymbol{\theta}_{t-k} \end{aligned}$$

Entonces, para obtener $f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t)$ primero hay que determinar $f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t)$ y $f(\boldsymbol{\theta}_{t-k} \mid D_t)$. Por hipótesis de inducción,

$$(\boldsymbol{\theta}_{t-k} \mid D_t) \sim \mathbf{N}[\mathbf{a}_t(-k), \mathbf{R}_t(-k)].$$

En el caso de $f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t)$, sea $\mathbf{Y} = \{Y_t, \dots, Y_{t-k}\}$, entonces por teorema de Bayes se tiene

$$\begin{aligned} f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t) &= f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, \mathbf{Y}, D_{t-k-1}) \\ &= \frac{f(\mathbf{Y} \mid \boldsymbol{\theta}_{t-k-1}, \boldsymbol{\theta}_{t-k}, D_{t-k-1}) f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_{t-k-1})}{f(\mathbf{Y} \mid \boldsymbol{\theta}_{t-k}, D_{t-k-1})} \\ &= f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_{t-k-1}) \end{aligned}$$

debido a que $f(\mathbf{Y} \mid \boldsymbol{\theta}_{t-k-1}, \boldsymbol{\theta}_{t-k}, D_{t-k-1}) = f(\mathbf{Y} \mid \boldsymbol{\theta}_{t-k}, D_{t-k-1})$ porque, dado $\boldsymbol{\theta}_{t-k}$, \mathbf{Y} es independiente de $\boldsymbol{\theta}_{t-k-1}$.

Por teorema de Bayes,

$$f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_{t-k-1}) \propto f(\boldsymbol{\theta}_{t-k} \mid \boldsymbol{\theta}_{t-k-1}, D_{t-k-1}) f(\boldsymbol{\theta}_{t-k-1} \mid D_{t-k-1})$$

Por definición del modelo lineal dinámico y por 2.3.1.1 se sabe que

$$\begin{aligned} (\boldsymbol{\theta}_{t-k} \mid \boldsymbol{\theta}_{t-k-1}, D_{t-k-1}) &\sim \mathbf{N}[\mathbf{G}_{t-k} \boldsymbol{\theta}_{t-k-1}, \mathbf{W}_{t-k}] \\ (\boldsymbol{\theta}_{t-k-1} \mid D_{t-k-1}) &\sim \mathbf{N}[\mathbf{m}_{t-k-1}, \mathbf{C}_{t-k-1}]. \end{aligned}$$

Con estas distribuciones y utilizando los resultados del apéndice A.2 se obtiene la siguiente distribución conjunta:

$$\begin{aligned} & \left(\boldsymbol{\theta}_{t-k} \middle| D_{t-k-1} \right) \\ & \sim \mathbf{N} \left[\begin{pmatrix} \mathbf{G}_{t-k} \mathbf{m}_{t-k-1} \\ \mathbf{m}_{t-k-1} \end{pmatrix}, \begin{pmatrix} \mathbf{G}_{t-k}^\top \mathbf{C}_{t-k-1} \mathbf{G}_{t-k} + \mathbf{W}_{t-k} & \mathbf{G}_{t-k} \mathbf{C}_{t-k-1} \\ \mathbf{C}_{t-k-1} \mathbf{G}_{t-k}^\top & \mathbf{C}_{t-k-1} \end{pmatrix} \right]. \end{aligned}$$

Entonces se tiene la siguiente distribución condicional,

$$(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_{t-k-1}) \sim \mathbf{N}[\mathbf{h}_t(k+1), \mathbf{H}_t(k+1)],$$

con

$$\mathbf{h}_t(k+1) = \mathbf{m}_{t-k-1} + \mathbf{C}_{t-k-1} \mathbf{G}_{t-k}^\top \mathbf{R}_{t-k}^{-1} [\boldsymbol{\theta}_{t-k} - \mathbf{a}_{t-k}]$$

$$\mathbf{H}_t(k+1) = \mathbf{C}_{t-k-1} - \mathbf{C}_{t-k-1} \mathbf{G}_{t-k}^\top \mathbf{R}_{t-k}^{-1} \mathbf{G}_{t-k} \mathbf{C}_{t-k-1}.$$

Sea $\mathbf{B}_{t-k-1} = \mathbf{C}_{t-k-1} \mathbf{G}_{t-k}^\top \mathbf{R}_{t-k}^{-1}$, entonces

$$\mathbf{h}_t(k+1) = \mathbf{m}_{t-k-1} + \mathbf{B}_{t-k-1} [\boldsymbol{\theta}_{t-k} - \mathbf{a}_{t-k}]$$

$$\mathbf{H}_t(k+1) = \mathbf{C}_{t-k-1} - \mathbf{B}_{t-k-1} \mathbf{R}_{t-k} \mathbf{B}_{t-k-1}^\top$$

Como vimos anteriormente, $f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t) = f(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_{t-k-1})$, entonces

$$(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t) \sim \mathbf{N}[\mathbf{h}_t(k+1), \mathbf{H}_t(k+1)].$$

Como $(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t)$ y $(\boldsymbol{\theta}_{t-k} \mid D_t)$ se distribuyen normal, entonces $(\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t)$ también se distribuye normal. Por lo tanto, ahora es necesario encontrar la esperanza y la varianza.

$$\begin{aligned} \mathbf{a}_t(-k-1) &= E[\boldsymbol{\theta}_{t-k-1} \mid D_t] = E[E[\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t] \mid D_t] = E[\mathbf{h}_t(k+1) \mid D_t] \\ &= \mathbf{m}_{t-k-1} + \mathbf{B}_{t-k-1} [E[\boldsymbol{\theta}_{t-k} \mid D_t] - \mathbf{a}_{t-k}] \end{aligned}$$

$$= \mathbf{m}_{t-k-1} + \mathbf{B}_{t-k-1}[\mathbf{a}_t(-k) - \mathbf{a}_{t-k}]$$

$$\begin{aligned} \mathbf{R}_t(-k-1) &= V[\boldsymbol{\theta}_{t-k-1} \mid D_t] \\ &= E[V[\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t] \mid D_t] + V[E[\boldsymbol{\theta}_{t-k-1} \mid \boldsymbol{\theta}_{t-k}, D_t] \mid D_t] \\ &= E[\mathbf{H}_t(k+1) \mid D_t] + V[\mathbf{h}_t(k+1) \mid D_t] \\ &= \mathbf{C}_{t-k-1} - \mathbf{B}_{t-k-1}\mathbf{R}_{t-k}\mathbf{B}_{t-k-1}^\top + \mathbf{B}_{t-k-1}V[\boldsymbol{\theta}_{t-k} \mid D_t]\mathbf{B}_{t-k-1}^\top \\ &= \mathbf{C}_{t-k-1} + \mathbf{B}_{t-k-1}[\mathbf{R}_t(-k) - \mathbf{R}_{t-k}]\mathbf{B}_{t-k-1}^\top. \end{aligned}$$

Por lo tanto,

$$(\boldsymbol{\theta}_{t-k-1} \mid D_t) \sim \mathbf{N}[\mathbf{a}_t(-k-1), \mathbf{R}_t(-k-1)].$$

□

Es interesante analizar lo que ocurre en el cálculo del valor esperado de $(\boldsymbol{\theta}_{t-k} \mid D_t)$, $\mathbf{a}_t(-k)$. Si $\mathbf{a}_t(-k+1)$ y \mathbf{a}_{t-k+1} tienen valores parecidos, o equivalentemente, si las esperanzas de $(\boldsymbol{\theta}_{t-k+1} \mid D_t)$ y de $(\boldsymbol{\theta}_{t-k+1} \mid D_{t-k})$ son similares, entonces el valor de $\mathbf{a}_t(-k)$ será cercano a la esperanza de $(\boldsymbol{\theta}_{t-k} \mid D_{t-k})$, \mathbf{m}_{t-k} . En otras palabras, utilizando toda la información D_t se corrige la distribución de parámetros que originalmente se obtuvo con la información D_{t-k} .

En el caso en el que la varianza de observación V_t es constante y desconocida, las distribuciones se vuelven t de Student y se tiene que considerar la estimación de la varianza.

Corolario 2.3.4.1. *Si la varianza de observación es una constante desconocida $V_t = V = \phi^{-1}$, entonces*

$$(\boldsymbol{\theta}_{t-k} \mid D_t) \sim \mathbf{T}_{n_t}[\mathbf{a}_t(-k), (S_t/S_{t-k})\mathbf{R}_t(-k)].$$

Otro concepto importante es el suavizamiento de la serie. A la estimación retrospectiva de la respuesta media $\mu_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t$ mediante las distribuciones filtradas $(\mu_{t-k} \mid D_t)$ se le conoce como suavizamiento.

Corolario 2.3.4.2. *Las distribuciones suavizadas para la respuesta media de la serie cuando V es desconocida son*

$$(\mu_{t-k} \mid D_t) \sim \mathbf{T}_{n_t}[f_t(-k), (S_t/S_{t-k})\mathbf{F}_{t-k}^\top \mathbf{R}_t(-k)\mathbf{F}_{t-k}],$$

con $f_t(-k) = \mathbf{F}_{t-k}^\top \mathbf{a}_t(-k)$.

Ejemplo 2.3.4.1. En el ejemplo 2.3.3.1 se obtuvo la distribución de $(\theta_{20} | D_{20})$ cuando V es desconocida. Ahora que se cuenta con D_{20} , es posible calcular la distribución filtrada $(\theta_{19} | D_{20})$. Con los valores calculados anteriormente, es rápido obtener la distribución filtrada.

```
B19 <- C19 %*% t(G20) %*% solve(R20)
a20_menos_1 <- m19 + B19 %*% (m20 - a20)
R20_menos_1 <- C19 - B19 %*% (C20 - R20) %*% solve(B19)
```

```
a20_menos_1
```

```
##           [,1]
## [1,]  8.0000
## [2,]  0.3488
## [3,] -0.2708
```

```
S20/S19 * R20_menos_1
```

```
##           [,1]      [,2]      [,3]
## [1,]  1.936e-05  8.966e-06 -2.144e-05
## [2,] -1.216e-05  1.147e-04  2.533e-06
## [3,] -2.960e-05  3.510e-05  6.162e-05
```

Entonces,

$$(\theta_{19} | D_{20}) \sim \mathbf{T}_{20.5} \left[\begin{pmatrix} 8.0 \\ 0.3488 \\ -0.2708 \end{pmatrix}, \begin{pmatrix} 0.000019 & 0.000009 & -0.000021 \\ -0.000012 & 0.000011 & 0.000003 \\ -0.00003 & -0.000035 & 0.000062 \end{pmatrix} \right].$$

Recordemos que en el ejemplo anterior se tenía que

$$(\theta_{19} | D_{19}) \sim \mathbf{T}_{19.5} \left[\begin{pmatrix} 8 \\ 0.35 \\ -0.27 \end{pmatrix}, \begin{pmatrix} 0.00002 & 0.00001 & -0.00002 \\ 0.00001 & 0.00004 & -0.00001 \\ -0.00002 & -0.00001 & 0.00005 \end{pmatrix} \right].$$

Hubo un ligero ajuste en los parámetros de locación y escala en la distribución ahora que se cuenta con D_{20} . Los grados de libertad son otra

diferencia importante, todas las distribuciones filtradas de los parámetros tendrán los grados de libertad asociados a D_{20} , que en este caso es de 20.5.

Ahora es posible obtener fácilmente la distribución suavizada para la respuesta media, solo es necesario utilizar el valor de las variables explicativas en $t = 19$.

```
F19 <- c(1, 6.05563, 4.53957) #Variables explicativas en t=19.

#Parámetro locacional
t(F19) %*% a20_menos_1

##           [,1]
## [1,] 8.883

#Parámetro de escala
S20/S19 * t(F19) %*% R20_menos_1 %*% F19

##           [,1]
## [1,] 0.006278
```

Por lo tanto, la distribución es

$$(\mu_{t-1} \mid D_{20}) \sim \mathbf{T}_{20.5}[8.883, 0.006278].$$

2.3.5. Intervenciones

Es posible que el conocimiento de nueva información externa al modelo provoque que se desee hacer cambios, por lo que es necesario tener un mecanismo formal para incorporar esa información al modelo. En la práctica pueden ocurrir muchos eventos que originalmente no estaban contemplados, por esta razón los modelos no deben ser cerrados, se debe tener la capacidad de realizar los ajustes necesarios. Las intervenciones pueden ser clasificadas hacia adelante o hacia atrás. Las intervenciones hacia adelante son anticipatorias, se interviene el modelo por eventos que pueden afectar



el futuro. Por otro lado, las intervenciones hacia atrás son correctivas, son necesarias cuando el desempeño de los pronósticos ha disminuido, entonces se utiliza retroactivamente información que se debió incluir anteriormente para ajustar el modelo. Idealmente el modelo se debe utilizar para realizar predicciones e inferencias hasta que algún suceso extraordinario ocurra. Los sucesos extraordinarios son los mencionados anteriormente, el surgimiento de información externa o un mal desempeño del modelo.

Se considera el caso en el que se desea intervenir el modelo 2.3.0.1 en el tiempo t . Las varianzas V_t son conocidas, pero los resultados aplican para el caso en el que V es una constante desconocida, la diferencia es que las distribuciones son t de Student y no normales. Se representa la intervención en el tiempo t como I_t , la cual es un conjunto de información que identifica a t como un periodo en el que se realizó una intervención e incluye la información empleada en la intervención. Entonces, la información disponible antes de observar a Y_t es $\{D_{t-1}, I_{t-1}\}$.

La forma de intervención que se presentará es conocida como Intervención subjetiva arbitraria. Recordemos que en el teorema 2.3.1.1 se obtuvo que la distribución a priori de los parámetros en t es

$$(\theta_t \mid D_{t-1}) \sim \mathbf{N}[\mathbf{a}_t, \mathbf{R}_t].$$

La intervención consiste en cambiar los momentos de la distribución a priori de θ_t por los nuevos valores que anticipen los cambios en la serie. Entonces, la información de intervención es

$$I_t = \{\mathbf{a}_t^*, \mathbf{R}_t^*\},$$

donde los valores \mathbf{a}_t^* y \mathbf{R}_t^* son dados por el modelador. Entonces, la distribución a priori de los parámetros post-intervención es

$$(\theta_t \mid I_t, D_{t-1}) \sim \mathbf{N}[\mathbf{a}_t^*, \mathbf{R}_t^*].$$

Para la realización de pronósticos esto no implica ningún problema, pues los momentos \mathbf{a}_t y \mathbf{R}_t simplemente son reemplazados por \mathbf{a}_t^* y \mathbf{R}_t^* . El pronóstico de Y_t , la actualización de la distribución de parámetros una vez

observada Y_t y los pronósticos a más de un periodo en el futuro se realizan con el procedimiento usual, solo que utilizando los nuevos momentos de intervención. El problema es que las distribuciones filtradas requieren la distribución conjunta de θ_t y θ_{t-1} condicionadas en D_{t-1} y en I_t , pero las modificaciones arbitrarias en los momentos de θ_t provocan que la distribución conjunta no sea coherente. Por lo tanto, es necesario incluir la intervención de una forma que sea consistente con los procedimientos mostrados anteriormente. El objetivo es modificar la ecuación de evolución de tal forma que se puedan obtener los momentos deseados en la modificación. Para esto, primero se necesitan los resultados del siguiente lema.

Lema 2.3.5.1. *Sea K_t una matriz cuadrada, triangular inferior, y h_t cualquier vector. Se define*

$$\theta_t^* = K_t \theta_t + h_t,$$

donde $E[\theta_t] = a_t$ y $V[\theta_t] = R_t$. Entonces θ_t^ tiene momentos a_t^* y R_t^* si K_t y h_t se escogen de la siguiente manera:*

$$\begin{aligned} K_t &= U_t Z_t^{-1} \\ h_t &= a_t^* - K_t a_t, \end{aligned}$$

donde U_t y Z_t son las factorizaciones de Cholesky de R_t^ y R_t , entonces $R_t^* = U_t U_t^\top$ y $R_t = Z_t Z_t^\top$. Las matrices U_t y Z_t son triangular inferior, únicas e invertibles.*

Demostración. R_t y R_t^* son matrices de covarianza, entonces son simétricas y positivas definidas. Por lo tanto, su descomposición de Cholesky es $R_t^* = U_t U_t^\top$ y $R_t = Z_t Z_t^\top$, donde U_t y Z_t son triangular inferior, únicas e invertibles.

Por definición de θ_t^* , como

$$E[\theta_t^*] = a_t^* = K_t E[\theta_t] + h_t = K_t a_t + h_t.$$

Entonces,

$$h_t = a_t^* - K_t a_t$$

para cualquier K_t dada.

Por definición de $\boldsymbol{\theta}_t^*$,

$$V[\boldsymbol{\theta}_t^*] = \mathbf{R}_t^* = V[\mathbf{K}_t \boldsymbol{\theta}_t + \mathbf{h}_t] = \mathbf{K}_t \mathbf{R}_t \mathbf{K}_t^\top.$$

Entonces,

$$\mathbf{R}_t^* = \mathbf{U}_t \mathbf{U}_t^\top = (\mathbf{K}_t \mathbf{Z}_t)(\mathbf{K}_t \mathbf{Z}_t)^\top.$$

De aquí se obtiene que $\mathbf{U}_t = \mathbf{K}_t \mathbf{Z}_t$, por lo tanto, $\mathbf{K}_t = \mathbf{U}_t \mathbf{Z}_t^{-1}$. □

El lema muestra cómo se pueden obtener los momentos deseados a partir de $\boldsymbol{\theta}_t$. En el siguiente teorema se introducen los resultados del lema en el modelo lineal dinámico.

Teorema 2.3.5.1. *Se supone que los momentos de intervención son \mathbf{a}_t^* y \mathbf{R}_t^* , y se definen \mathbf{K}_t y \mathbf{h}_t como en el lema 2.3.5.1. Entonces, la distribución a priori de los parámetros es*

$$(\boldsymbol{\theta}_t \mid I_t, D_{t-1}) \sim \mathbf{N}[\mathbf{a}_t^*, \mathbf{R}_t^*],$$

con ecuación de evolución

$$\boldsymbol{\theta}_t = \mathbf{G}_t^* \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^* \quad \boldsymbol{\omega}_t^* \sim \mathbf{N}[\mathbf{h}_t, \mathbf{W}_t^*],$$

donde

$$\begin{aligned} \mathbf{G}_t^* &= \mathbf{K}_t \mathbf{G}_t, \\ \boldsymbol{\omega}_t^* &= \mathbf{K}_t \boldsymbol{\omega}_t + \mathbf{h}_t, \\ \mathbf{W}_t^* &= \mathbf{K}_t \mathbf{W}_t \mathbf{K}_t^\top. \end{aligned}$$

Demostración. En el lema 2.3.5.1 se muestra que para realizar la intervención se requiere calcular $\boldsymbol{\theta}_t^* = \mathbf{K}_t \boldsymbol{\theta}_t + \mathbf{h}_t$. Equivalentemente, lo que se necesita es intervenir $\boldsymbol{\theta}_t$ de la siguiente manera:

$$\boldsymbol{\theta}_t = \mathbf{K}_t[\mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t] + \mathbf{h}_t = \mathbf{K}_t \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{K}_t \boldsymbol{\omega}_t + \mathbf{h}_t = \mathbf{G}_t^* \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t^*$$

De aquí se obtiene que $\mathbf{G}_t^* = \mathbf{K}_t \mathbf{G}_t$, $\boldsymbol{\omega}_t^* = \mathbf{K}_t \boldsymbol{\omega}_t + \mathbf{h}_t$ y que $\mathbf{W}_t^* = \mathbf{K}_t \mathbf{W}_t \mathbf{K}_t^\top$. Además, como

$$\begin{aligned} E[\boldsymbol{\omega}_t^*] &= E[\mathbf{K}_t \boldsymbol{\omega}_t + \mathbf{h}_t] = \mathbf{h}_t \\ V[\boldsymbol{\omega}_t^*] &= V[\mathbf{K}_t \boldsymbol{\omega}_t + \mathbf{h}_t] = \mathbf{K}_t \mathbf{W}_t \mathbf{K}_t^\top = \mathbf{W}_t^*, \end{aligned}$$

entonces la distribución de ω_t^* es

$$\omega_t^* \sim \mathbf{N}[\mathbf{h}_t, \mathbf{W}_t^*].$$

Como θ_{t-1} y ω_t^* se distribuyen normal, entonces $\mathbf{G}_t^* \theta_{t-1} + \omega_t^*$ también se distribuye normal. Entonces, solo es necesario calcular la esperanza y la varianza de la distribución.

$$\begin{aligned} E[\theta_t \mid D_{t-1}, I_t] &= E[\mathbf{G}_t^* \theta_{t-1} + \omega_t^* \mid D_{t-1}, I_t] = \mathbf{G}_t^* \mathbf{m}_{t-1} + \mathbf{h}_t \\ &= \mathbf{K}_t \mathbf{G}_t \mathbf{m}_{t-1} + \mathbf{h}_t = \mathbf{K}_t \mathbf{a}_t + \mathbf{h}_t = E[\mathbf{K}_t \theta_t + \mathbf{h}_t \mid D_{t-1}, I_t] \\ &= \mathbf{a}_t^* \quad (\text{por el lema}). \end{aligned}$$

Por independencia de θ_{t-1} y ω_t^* ,

$$\begin{aligned} V[\theta_t \mid D_{t-1}, I_t] &= V[\mathbf{G}_t^* \theta_{t-1} + \omega_t^* \mid D_{t-1}, I_t] = \mathbf{G}_t^* \mathbf{C}_{t-1} \mathbf{G}_t^{*\top} + \mathbf{K}_t \mathbf{W}_t \mathbf{K}_t^\top \\ &= \mathbf{K}_t \mathbf{G}_t \mathbf{C}_{t-1} (\mathbf{K}_t \mathbf{G}_t)^\top + \mathbf{K}_t \mathbf{W}_t \mathbf{K}_t^\top \\ &= \mathbf{K}_t [\mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}_t] \mathbf{K}_t^\top = \mathbf{K}_t \mathbf{R}_t \mathbf{K}_t^\top \\ &= V[\mathbf{K}_t \theta_t + \mathbf{h}_t \mid D_{t-1}, I_t] = \mathbf{R}_t^* \quad (\text{por el lema}). \end{aligned}$$

Por lo tanto,

$$(\theta_t \mid I_t, D_{t-1}) \sim \mathbf{N}[\mathbf{a}_t^*, \mathbf{R}_t^*].$$

□

Ejemplo 2.3.5.1. Para ilustrar que en el caso con varianza de observación constante y desconocida las intervenciones se aplican de la misma manera que en el teorema 2.3.5.1, se realizará una intervención a lo realizado en el ejemplo 2.3.3.1. Recordemos que se están modelando las ventas de una refresquera. Se cuenta con D_{19} , y la distribución de θ_{20} es

$$(\theta_{20} \mid D_{19}) \sim \mathbf{T}_{19.5} \left[\begin{pmatrix} 8.40 \\ 0.3570 \\ -0.2673 \end{pmatrix}, \begin{pmatrix} 0.000032 & 0.000011 & -0.000021 \\ 0.000011 & 0.00014 & -0.000020 \\ -0.000021 & -0.000020 & 0.000099 \end{pmatrix} \right],$$

Ahora, supongamos durante el último mes la refresquera realizó una campaña de publicidad muy exitosa. Contrataron al cantante más famoso del momento para que los promocionara, por lo que ahora se sabe que las personas están dispuestas a gastar una cantidad mayor de su ingreso en los

refrescos de la empresa. Por lo tanto se intervendrá el modelo para aumentar el parámetro asociado al ingreso: se decide incrementar su valor de 0.357 a 0.7. Sin embargo, aunque se sabe que el parámetro del ingreso debe aumentar, se tiene incertidumbre sobre este incremento, por lo que también se aumentará la varianza del ingreso, de 0.00014 a 0.001. Es decir,

$$\mathbf{a}_{20} = \begin{pmatrix} 8.40 \\ 0.3570 \\ -0.2673 \end{pmatrix}, \quad \mathbf{R}_{20} = \begin{pmatrix} 0.000032 & 0.000011 & -0.000021 \\ 0.000011 & 0.00014 & -0.000020 \\ -0.000021 & -0.000020 & 0.000099 \end{pmatrix}.$$

$$\mathbf{a}_{20}^* = \begin{pmatrix} 8.40 \\ 0.7 \\ -0.2673 \end{pmatrix}, \quad \mathbf{R}_{20}^* = \begin{pmatrix} 0.000032 & 0.000011 & -0.000021 \\ 0.000011 & 0.001 & -0.000020 \\ -0.000021 & -0.000020 & 0.000099 \end{pmatrix}.$$

Por lo tanto,

$$(\boldsymbol{\theta}_{20} \mid D_{19}, I_{20}) \sim \mathbf{T}_{19.5} \left[\begin{pmatrix} 8.40 \\ 0.7 \\ -0.2673 \end{pmatrix}, \begin{pmatrix} 0.000032 & 0.000011 & -0.000021 \\ 0.000011 & 0.001 & -0.000020 \\ -0.000021 & -0.000020 & 0.000099 \end{pmatrix} \right].$$

Con esta intervención, se calculará la distribución del pronóstico de Y_{20} con el procedimiento usual, solo que ahora se utilizan \mathbf{a}_{20}^* y \mathbf{R}_{20}^* en lugar de \mathbf{a}_{20} y \mathbf{R}_{20} .

```
a20_int <- c(8.4000, 0.7, -0.2673)
R20_int <- matrix(c(3.205e-05, 0.000010710, -2.0790e-05,
                    1.071e-05, 0.001, -2.0098e-05,
                    -2.079e-05, -0.000020098, 9.9005e-05),
                  ncol = 3)

F20 <- c(1, 6.06093, 4.51018) #Variables explicativas en t=20.
S19 <- 0.002 # Estimación de V en T=19
n19 <- 19.5 # Grados de libertad

#C. Pronóstico a un periodo.
f20 <- as.numeric(t(F20) %*% a20_int)
Q20 <- as.numeric(t(F20) %*% R20_int %*% F20 + S19)

#Parámetro de ubicación
f20
```

```
## [1] 11.44
```

```
#Parámetro de escala
```

```
Q20
```

```
## [1] 0.03962
```

Entonces, la distribución del pronóstico de Y_{20} es

$$(Y_{20} \mid D_{20}, I_{20}) \sim \mathbf{T}_{19.5}[11.44, 0.3962].$$

Con esta distribución se calcula el siguiente intervalo de pronóstico del 95 % para Y_{20} :

```
c(qst(0.025, nu = n19, mu = f20, sigma = sqrt(Q20)),
  qst(0.975, nu = n19, mu = f20, sigma = sqrt(Q20)))
```

```
## [1] 11.02 11.85
```

Aquí se aprecia el resultado de la intervención. Sin intervención, en el ejemplo 2.3.3.1 este intervalo era [9.17, 9.546]. Con la intervención, el intervalo toma valores más grandes y su longitud es mayor. Este es el efecto que se deseaba modelar, un incremento en las ventas y en la incertidumbre.

Unas semanas después, se conoce el valor real de Y_{20} . El valor observado fue de 11.05, apenas cubierto por el límite inferior del intervalo. Entonces, ahora que se cuenta con la información D_{20} , hay que actualizar la distribución de θ_{20} .

```
#Valor observado de Y20:
```

```
Y20 <- 11.05
```

```
#D. Posterior en t=20
```

```
A20 <- R20_int %*% F20 / Q20
```

```
e20 <- Y20-f20
```

```
m20 <- a20_int + A20 %*% e20
```

```
n20 <- n19 + 1
```

```
S20 <- S19 + (S19/n20)*(e20^2/Q20-1)
C20 <- (S20/S19)*(R20_int-A20 %*% t(A20) * Q20)
```

```
m20
```

```
##           [,1]
## [1,]  8.4000
## [2,]  0.6416
## [3,] -0.2703
```

```
C20
```

```
##           [,1]           [,2]           [,3]
## [1,]  3.640e-05  1.162e-05 -2.364e-05
## [2,]  1.162e-05  1.104e-04 -7.492e-05
## [3,] -2.364e-05 -7.492e-05  1.098e-04
```

Entonces,

$$(\theta_{20} \mid D_{20}) \sim \mathbf{T}_{20.5} \left[\begin{pmatrix} 8.4 \\ 0.6416 \\ -0.2703 \end{pmatrix}, \begin{pmatrix} 0.000036 & 0.000012 & -0.000024 \\ 0.000012 & 0.00011 & -0.000075 \\ -0.000024 & -0.000075 & 0.00011 \end{pmatrix} \right].$$

Como Y_{20} tuvo un valor observado bajo en comparación con la distribución que se obtuvo para su pronóstico, la distribución de los parámetros se ajustó para reducir el efecto del ingreso a 0.6416.

Hasta ahora no se han utilizado los resultados del lema 2.3.5.1 y del teorema 2.3.5.1 ya que para las distribuciones de pronóstico y para las actualizaciones no son necesarios. Sin embargo, para obtener las distribuciones filtradas se deben aplicar estos resultados. Primero hay que calcular las descomposiciones de Cholesky de \mathbf{R}_{20} y de \mathbf{R}_{20} , con esto es posible calcular \mathbf{K}_{20} . Posteriormente, se calculan \mathbf{G}_{20}^* y \mathbf{W}_{20}^* .

```
U20 <- chol(R20_int)
Z20 <- chol(R20)
```

```

K20 <- U20 %*% solve(Z20)
G20_int <- K20 %*% G20
W20_int <- K20 %*% W20 %*% t(K20)

```

Con esto ya se tienen los elementos necesarios para calcular las distribuciones filtradas, solo es cuestión de usar \mathbf{G}_{20}^* y \mathbf{W}_{20}^* en lugar de \mathbf{G}_{20} y \mathbf{W}_{20} .

```

B19 <- C19 %*% t(G20_int) %*% solve(R20_int)
a20_menos_1 <- m19 + B19 %*% (m20 - a20_int)
R20_menos_1 <- C19 - B19 %*% (C20 - R20_int) %*% solve(B19)

```

```
a20_menos_1
```

```

##           [,1]
## [1,]  7.9994
## [2,]  0.3440
## [3,] -0.2712

```

```
S20/S19 * R20_menos_1
```

```

##           [,1]           [,2]           [,3]
## [1,] -2.193e-05  0.0001042 -1.693e-05
## [2,] -4.918e-04  0.0011079 -2.590e-05
## [3,] -1.569e-04  0.0002566  3.208e-05

```

Por lo tanto,

$$(\boldsymbol{\theta}_{19} \mid D_{20}) \sim \mathbf{T}_{20.5} \left[\begin{pmatrix} 7.9994 \\ 0.3440 \\ -0.2712 \end{pmatrix}, \begin{pmatrix} -0.000022 & 0.000104 & -0.000017 \\ -0.000492 & 0.001108 & -0.000026 \\ -0.000157 & 0.000257 & 0.000032 \end{pmatrix} \right].$$

2.4. Modelo lineal dinámico multivariado

El modelo lineal dinámico multivariado para un vector de series de tiempo \mathbf{Y}_t se definió en 2.2.0.1. Se supone que \mathbf{Y}_t es un vector de r observaciones, la cuádrupla que define el modelo es

$$\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t = \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$$

con

- F_t es una matriz $n \times r$ conocida.
- G_t es una matriz $n \times n$ conocida.
- V_t es una matriz de covarianza $r \times r$ conocida.
- W_t es una matriz de covarianza $n \times n$ conocida.

Las ecuaciones del modelo son

$$\text{Ecuación de observación:} \quad Y_t = F_t^\top \theta_t + \nu_t, \quad \nu_t \sim \mathbf{N}[\mathbf{0}, V_t]$$

$$\text{Ecuación de sistema:} \quad \theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim \mathbf{N}[\mathbf{0}, W_t]$$

Al igual que en el modelo univariado, el vector de parámetros θ_t es de dimensión n . Cuando todos los componentes del modelo son conocidos la actualización de distribuciones, los pronósticos y las distribuciones filtradas son muy parecidos al caso univariado, la única diferencia es que algunas cantidades son vectores o matrices en vez de números, pero las ecuaciones son las mismas. Cuando $r = 1$, los resultados son idénticos a los del caso univariado. Sin embargo, asumir que V_t es conocida es una limitación importante debido a que en la práctica esta matriz de covarianza será desconocida. El problema es que no hay un análisis conjugado para el modelo lineal dinámico multivariado cuando la matriz de covarianza de observación es desconocida. Sin embargo, hay modelos que son extensiones del modelo lineal dinámico multivariado que permiten el análisis conjugado. En estos modelos, los parámetros se presentan en una matriz Θ_t en lugar de un vector θ_t . Entonces, la distribución de los parámetros dada la matriz de covarianza de observación es matriz normal, que es una generalización de la normal multivariada. Estos modelos se presentan a continuación.

2.4.1. Modelos lineales dinámicos matriz normal

Estos modelos son una extensión del modelo lineal dinámico multivariado que permiten realizar el análisis conjugado cuando la matriz de covarianza de observación es desconocida si se puede asumir que las series individuales siguen un modelo lineal dinámico univariado con matrices F_t y

\mathbf{G}_t comunes. Esto es razonable cuando se tienen series que son componentes de otra serie. Un ejemplo de esto es la demanda de dinero por denominación.

Se supone que se tienen q series univariadas Y_{tj} que siguen un modelo lineal dinámico univariado para cada $j = 1, \dots, q$ con la siguiente cuádrupla:

$$\{\mathbf{F}_t, \mathbf{G}_t, V_t \sigma_j^2, \mathbf{W}_t \sigma_j^2\}.$$

El modelo es de dimensión n y todos los elementos de la cuádrupla son conocidos con la excepción de los factores de escala σ_j^2 . Entonces, las series univariadas tienen las siguientes ecuaciones de observación y evolución:

$$\text{Ecuación de observación:} \quad Y_{tj} = \mathbf{F}_t^\top \boldsymbol{\theta}_{tj} + \nu_{tj}, \quad \nu_{tj} \sim \mathbf{N}[0, V_t \sigma_j^2]$$

$$\text{Ecuación de sistema:} \quad \boldsymbol{\theta}_{tj} = \mathbf{G}_t \boldsymbol{\theta}_{t-1,j} + \boldsymbol{\omega}_{tj}, \quad \boldsymbol{\omega}_{tj} \sim \mathbf{N}[\mathbf{0}, \mathbf{W}_t \sigma_j^2].$$

Las matrices \mathbf{F}_t , \mathbf{G}_t y \mathbf{W}_t son las mismas para todas las q series. Además, se tiene un factor de escala V_t conocido común para todas las series.

La estructura transversal conjunta de todas las series en el tiempo t está dada por las covarianzas de los errores de observación entre las q series, mediante una matriz de covarianza $q \times q$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,q} \\ \sigma_{2,1} & \sigma_2^2 & \dots & \sigma_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q,1} & \sigma_{q,2} & \dots & \sigma_q^2 \end{pmatrix},$$

para todas i y j , $i \neq j$, σ_{ij} determina la covarianza entre Y_{ti} y Y_{tj} : Condicionando en $\boldsymbol{\Sigma}$ se tiene que

$$C[\nu_{ti}, \nu_{tj}] = V_t \sigma_{i,j} \quad (2.3)$$

y

$$C[\boldsymbol{\omega}_{ti}, \boldsymbol{\omega}_{tj}] = \mathbf{W}_t \sigma_{i,j}. \quad (2.4)$$

Las series individuales Y_{tj} siguen un modelo lineal dinámico univariado, los parámetros $\boldsymbol{\theta}_{tj}$ tienen valores diferentes para cada serie. La estructura de correlación establecida por $\boldsymbol{\Sigma}$ afecta los errores de observación mediante

2.3, y los errores de evolución a través de 2.4. Si σ_{ij} es grande y positiva, entonces las series i y j tendrán patrones similares en sus parámetros y en las variaciones de sus observaciones.

El modelo puede representarse en notación matricial de la siguiente manera:

- $\mathbf{Y}_t = (Y_{t1}, \dots, Y_{tq})^\top$, el vector de dimensión q de las observaciones en el tiempo t .
- $\boldsymbol{\nu}_t = (\nu_{t1}, \dots, \nu_{tq})^\top$, el vector de dimensión q de errores de observación en el tiempo t .
- $\boldsymbol{\Theta}_t = [\boldsymbol{\theta}_{t1}, \dots, \boldsymbol{\theta}_{tq}]$, la matriz $n \times q$ cuyas columnas son los vectores de parámetros de los modelos lineales dinámicos individuales.
- $\boldsymbol{\Omega}_t = [\boldsymbol{\omega}_{t1}, \dots, \boldsymbol{\omega}_{tq}]$, la matriz $n \times q$ cuyas columnas son los errores de evolución de los modelos lineales dinámicos.

Entonces, de esta forma el modelo se define así:

$$\begin{aligned}\mathbf{Y}_t^\top &= \mathbf{F}_t^\top \boldsymbol{\Theta}_t + \boldsymbol{\nu}_t^\top, \\ \boldsymbol{\Theta}_t &= \mathbf{G}_t \boldsymbol{\Theta}_{t-1} + \boldsymbol{\Omega}_t.\end{aligned}$$

Cuando $q = 1$, este modelo es equivalente al modelo lineal dinámico univariado, con factor de escala observacional $\boldsymbol{\Sigma} = \sigma_1^2$ desconocido. Para cualquier q , si $\boldsymbol{\Sigma}$ es diagonal, entonces las q series no están relacionadas.

Un punto importante es que \mathbf{G}_t y \mathbf{F}_t son comunes para todas las q series. \mathbf{F}_t es un vector de dimensión n , al igual que en caso univariado. Condicionando en $\boldsymbol{\Sigma}$, los errores de distribución tienen una distribución normal multivariada,

$$\boldsymbol{\nu}_t \sim \mathbf{N}[\mathbf{0}, V_t \boldsymbol{\Sigma}].$$

La matriz de errores de evolución tiene una distribución matriz normal, descrita a continuación.

2.4.2. Distribución de los errores de evolución

La matriz aleatoria $\mathbf{\Omega}_t$ tiene una distribución matriz normal con media una matriz $\mathbf{0}$, matriz de covarianza izquierda \mathbf{W}_t y matriz de covarianza derecha $\mathbf{\Sigma}$. La función de densidad es

$$f(\mathbf{\Omega}_t) = k(\mathbf{W}_t, \mathbf{\Sigma}) \exp\left\{-\frac{1}{2} \text{tr}[\mathbf{\Omega}_t^\top \mathbf{W}_t^{-1} \mathbf{\Omega}_t \mathbf{\Sigma}^{-1}]\right\},$$

donde

$$k(\mathbf{W}_t, \mathbf{\Sigma}) = (2\pi)^{-qn/2} |\mathbf{W}_t|^{-q/2} |\mathbf{\Sigma}|^{-n/2}.$$

Todas las distribuciones marginales y condicionales de elementos de $\mathbf{\Omega}_t$, y sus funciones lineales, tienen distribución normal univariada, multivariada o matriz normal. La definición de la distribución es válida cuando alguna de las matrices de covarianza es semidefinida positiva. Si cualquiera de las matrices de covarianza es una matriz cero, entonces $\mathbf{\Omega}_t$ es cero con probabilidad 1. Esta distribución se denota como

$$\mathbf{\Omega}_t \sim \mathbf{N}[\mathbf{0}, \mathbf{W}_t, \mathbf{\Sigma}].$$

Esta distribución es la base de las distribuciones a priori y posterior de la matriz de parámetros $\mathbf{\Theta}_t$. Condicionando en $\mathbf{\Sigma}$, la distribución de $\mathbf{\Theta}_t$ es matriz normal. Sin embargo, cuando $\mathbf{\Sigma}$ es desconocida se necesitan considerar otras distribuciones para poder aplicar el modelo.

2.4.3. Distribución Wishart inversa

Sea $\mathbf{\Sigma}$ una matriz $q \times q$ positiva definida, entonces $\mathbf{\Sigma}$ tiene una distribución Wishart inversa si su función de densidad es de la siguiente forma:

$$f(\mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-(q+n/2)} \exp\left\{-\frac{1}{2} \text{tr}[n\mathbf{S}\mathbf{\Sigma}^{-1}]\right\},$$

donde $n > 0$ es el parámetro conocido de los grados de libertad, y \mathbf{S} es una matriz $q \times q$ positiva definida conocida. La notación es

$$\mathbf{\Sigma} \sim W_n^{-1}[\mathbf{S}].$$

Sea $\mathbf{\Phi} = \mathbf{\Sigma}^{-1}$, entonces $\mathbf{\Phi}$ tiene una distribución Wishart con n grados de libertad y media $E[\mathbf{\Phi}] = \mathbf{S}^{-1}$. Entonces, \mathbf{S} es una estimación de $\mathbf{\Sigma}$. Si \mathbf{S}

tiene la forma

$$\mathbf{S} = \begin{pmatrix} S_1 & S_{1,2} & \dots & S_{1,q} \\ S_{2,1} & S_2 & \dots & S_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ S_{q,1} & S_{q,2} & \dots & S_q \end{pmatrix},$$

entonces S_j es una estimación de la varianza σ_j^2 , y S_{jk} es una estimación de la covarianza σ_{jk} . Si $q = 1$, entonces $\mathbf{\Sigma}$, $\mathbf{\Phi} = \phi$ y $\mathbf{S} = S$ son escalares. En este caso, $\phi \sim G[n/2, nS/2]$, con $E[\phi] = 1/S$. Es claro que estos resultados son idénticos a los del modelo lineal dinámico univariado cuando la varianza de observación es desconocida. Las distribuciones marginales de las varianzas σ_j^2 en la diagonal de $\mathbf{\Sigma}$ son gamma inversa, mientras que la distribución de $\phi_j = \sigma_j^{-2}$ es una gamma, $\phi_j \sim G[n/2, nS_j/2]$.

Ahora que se definió la distribución de la matriz de covarianza $\mathbf{\Sigma}$, y se sabe que la distribución de $\mathbf{\Theta}_t$ condicionada en $\mathbf{\Sigma}$ es una matriz normal, es necesario obtener la distribución conjunta de ambas matrices.

2.4.4. Distribución matriz normal-Wishart inversa

Se supone que $\mathbf{\Sigma}$ tiene una distribución normal-Wishart. Si $\mathbf{\Theta}$ es una matriz $p \times q$, su distribución condicionada en $\mathbf{\Sigma}$ es una matriz normal con matriz de media $p \times q$ \mathbf{m} , matriz de covarianza izquierda $p \times p$ \mathbf{C} y matriz de covarianza derecha $\mathbf{\Sigma}$. Además, \mathbf{m} y \mathbf{C} son conocidas. Entonces,

$$(\mathbf{\Theta}|\mathbf{\Sigma}) \sim \mathbf{N}[\mathbf{m}, \mathbf{C}, \mathbf{\Sigma}].$$

Con esta distribución y la de $\mathbf{\Sigma}$ se define la distribución conjunta de $\mathbf{\Theta}$ y $\mathbf{\Sigma}$, conocida como matriz normal-Wishart inversa. Cuando $q = 1$ y $\mathbf{\Sigma}$ es un escalar, esta distribución es equivalente a la normal-gamma presentada en A.3. La notación de la distribución conjunta es

$$(\mathbf{\Theta}, \mathbf{\Sigma}) \sim NW_n^{-1}[\mathbf{m}, \mathbf{C}, \mathbf{S}].$$

Por último, es necesario conocer la distribución marginal de la matriz $\mathbf{\Theta}$.

2.4.5. Distribución matriz T

En el caso univariado, cuando la varianza de observación es desconocida, la distribución marginal de los parámetros es una t de Student. En el caso de la matriz Θ , su distribución es la generalización matricial de una t de Student. Esta distribución es la matriz T. Su notación es

$$\Theta \sim \mathbf{T}[\mathbf{m}, \mathbf{C}, \mathbf{S}].$$

Las columnas individuales de Θ tienen una distribución t de Student multivariada de dimensión p con n grados de libertad. Si se tiene que $\Theta = [\theta_1, \dots, \theta_q]$ y $\mathbf{m} = [\mathbf{m}_1, \dots, \mathbf{m}_q]$, entonces la distribución marginal de cualquiera de las columnas es

$$\theta_j \sim \mathbf{T}_n[\mathbf{m}_j, \mathbf{C}\mathbf{S}_j].$$



Capítulo 3

Pronóstico de la demanda agregada de dinero

- 3.1. Implementación en R
- 3.2. Verificación del modelo
- 3.3. Análisis de intervención durante la pandemia
- 3.4. Comparación con Error Correction Model

Capítulo 4

Pronóstico de la demanda de dinero por denominación

- 4.1. Implementación en R
- 4.2. Verificación del modelo
- 4.3. Análisis de intervención durante la pandemia
- 4.4. Comparación con Error Correction Model

Capítulo 5

Conclusión

5.1. Resumen de los resultados

5.2. Limitaciones y mejoras futuras

Apéndice A

Resultados de probabilidad y estadística

A.1. Distribuciones marginales y condicionales de la normal multivariada

Sea \mathbf{X} una variable aleatoria con distribución normal multivariada con $E[\mathbf{X}] = \mathbf{m}$ y $V[\mathbf{X}] = \mathbf{V}$, entonces se usa la notación $\mathbf{X} \sim \mathbf{N}[\mathbf{m}, \mathbf{V}]$

Si se tienen las siguientes particiones

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \quad \mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix},$$

entonces $\mathbf{X}_i \sim \mathbf{N}[\mathbf{m}_i, \mathbf{V}_{ii}]$, $i = 1, 2$.

Para estas particiones, se obtienen la siguientes distribución condicional para $(\mathbf{X}_1 \mid \mathbf{X}_2)$ sin pérdida de generalidad:

$$(\mathbf{X}_1 \mid \mathbf{X}_2) \sim \mathbf{N}[\mathbf{m}_1(\mathbf{X}_2), \mathbf{V}_{11}(\mathbf{X}_2)],$$

donde

$$\mathbf{m}_1(\mathbf{X}_2) = \mathbf{m}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{X}_2 - \mathbf{m}_2)$$

y

$$\mathbf{V}_{11}(\mathbf{X}_2) = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$$

A la matriz $\mathbf{A}_1 = \mathbf{V}_{12}\mathbf{V}_{22}^{-1}$ se le conoce como la matriz de regresión de \mathbf{X}_1 en \mathbf{X}_2 . Los momentos condicionales pueden expresarse en términos de \mathbf{A}_1 como

$$\mathbf{m}_1(\mathbf{X}_2) = \mathbf{m}_1 + \mathbf{A}_1(\mathbf{X}_2 - \mathbf{m}_2)$$

y

$$\mathbf{V}_{11}(\mathbf{X}_2) = \mathbf{V}_{11} - \mathbf{A}_1\mathbf{V}_{22}^{-1}\mathbf{A}_1^\top.$$

A.2. Distribución normal condicional y regresión lineal

Sean \mathbf{Y} un vector $p \times 1$ y $\boldsymbol{\theta}$ un vector $n \times 1$ relacionados a través de la siguiente distribución condicional:

$$(\mathbf{Y} \mid \boldsymbol{\theta}) \sim \mathbf{N}[\mathbf{F}^\top \boldsymbol{\theta}, \mathbf{V}],$$

donde \mathbf{F} es una matriz $n \times p$ constante, y \mathbf{V} es una matriz $p \times p$ constante, simétrica y positiva definida. Equivalentemente, se tiene que

$$\mathbf{Y} = \mathbf{F}^\top \boldsymbol{\theta} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim \mathbf{N}[\mathbf{0}, \mathbf{V}]$$

La distribución marginal de $\boldsymbol{\theta}$ es de la siguiente forma:

$$\boldsymbol{\theta} \sim \mathbf{N}[\mathbf{a}, \mathbf{R}],$$

con \mathbf{a} y \mathbf{R} constantes, y $\boldsymbol{\theta}$ es independiente de $\boldsymbol{\nu}$. De forma equivalente, se tiene la expresión

$$\boldsymbol{\theta} = \mathbf{a} + \boldsymbol{\omega}, \quad \boldsymbol{\omega} \sim \mathbf{N}[\mathbf{0}, \mathbf{R}],$$

donde $\boldsymbol{\omega}$ es independiente de $\boldsymbol{\nu}$.

Notemos que $(\boldsymbol{\nu}^\top, \boldsymbol{\omega}^\top)^\top$ tiene una distribución normal multivariada. Como $\boldsymbol{\theta} = \mathbf{a} + \boldsymbol{\omega}$ y $\mathbf{Y} = \mathbf{F}^\top \boldsymbol{\theta} + \boldsymbol{\nu} = \mathbf{F}^\top \mathbf{a} + \mathbf{F}^\top \boldsymbol{\omega} + \boldsymbol{\nu}$, entonces $(\mathbf{Y}^\top, \boldsymbol{\theta}^\top)^\top$ es una transformación lineal de $(\boldsymbol{\nu}^\top, \boldsymbol{\omega}^\top)^\top$. Por lo tanto, $(\mathbf{Y}^\top, \boldsymbol{\theta}^\top)^\top$ también tiene una distribución normal multivariada.

Entonces,

- $E[\boldsymbol{\theta}] = \mathbf{a}$;
- $V[\boldsymbol{\theta}] = \mathbf{R}$;
- $E[\mathbf{Y}] = E[\mathbf{F}^\top \boldsymbol{\theta} + \boldsymbol{\nu}] = \mathbf{F}^\top E[\boldsymbol{\theta}] + E[\boldsymbol{\nu}] = \mathbf{F}^\top \mathbf{a}$;
- $V[\mathbf{Y}] = V[\mathbf{F}^\top \boldsymbol{\theta} + \boldsymbol{\nu}] = \mathbf{F}^\top V[\boldsymbol{\theta}] \mathbf{F} + V[\boldsymbol{\nu}] = \mathbf{F}^\top \mathbf{R} \mathbf{F} + \mathbf{V}$;
- $C[\mathbf{Y}, \boldsymbol{\theta}] = C[\mathbf{F}^\top \boldsymbol{\theta} + \boldsymbol{\nu}, \boldsymbol{\theta}] = \mathbf{F}^\top C[\boldsymbol{\theta}, \boldsymbol{\theta}] + C[\boldsymbol{\nu}, \boldsymbol{\theta}] = \mathbf{F}^\top \mathbf{R}$.
- $C[\boldsymbol{\theta}, \mathbf{Y}] = C[\mathbf{Y}, \boldsymbol{\theta}]^\top = \mathbf{R}^\top \mathbf{F} = \mathbf{R} \mathbf{F}$, debido a que \mathbf{R} es una matriz de covarianza y por lo tanto, es simétrica.

Por lo tanto,

$$\begin{pmatrix} \mathbf{Y} \\ \boldsymbol{\theta} \end{pmatrix} \sim \mathbf{N} \left[\begin{pmatrix} \mathbf{F}^\top \mathbf{a} \\ \mathbf{a} \end{pmatrix}, \begin{pmatrix} \mathbf{F}^\top \mathbf{R} \mathbf{F} + \mathbf{V} & \mathbf{F}^\top \mathbf{R} \\ \mathbf{R} \mathbf{F} & \mathbf{R} \end{pmatrix} \right].$$

Utilizando la distribución condicional presentada en A.1, se obtiene la siguiente distribución condicional:

$$(\boldsymbol{\theta} \mid \mathbf{Y}) \sim \mathbf{N}[\mathbf{m}, \mathbf{C}],$$

donde

$$\mathbf{m} = \mathbf{a} + \mathbf{R} \mathbf{F} [\mathbf{F}^\top \mathbf{R} \mathbf{F} + \mathbf{V}]^{-1} [\mathbf{Y} - \mathbf{F}^\top \mathbf{a}]$$

y

$$\mathbf{C} = \mathbf{R} - \mathbf{R} \mathbf{F} [\mathbf{F}^\top \mathbf{R} \mathbf{F} + \mathbf{V}]^{-1} \mathbf{F}^\top \mathbf{R}$$

Si se define $\mathbf{e} = \mathbf{Y} - \mathbf{F}^\top \mathbf{a}$, $\mathbf{Q} = \mathbf{F}^\top \mathbf{R} \mathbf{F} + \mathbf{V}$ y $\mathbf{A} = \mathbf{R} \mathbf{Q} \mathbf{R}^\top$, entonces las ecuaciones son

$$\mathbf{m} = \mathbf{a} + \mathbf{A} \mathbf{e} \quad \text{y} \quad \mathbf{C} = \mathbf{R} - \mathbf{A} \mathbf{Q} \mathbf{A}^\top.$$

A.3. Distribución normal-gamma

Una variable aleatoria $\phi > 0$ tiene una distribución gamma con parámetros $n > 0$ y $d > 0$, denotada como $\phi \sim \mathbf{G}[n, d]$, si

$$f(\phi) = d^n \Gamma(n)^{-1} \phi^{n-1} \exp(-\phi d) \propto \phi^{n-1} \exp(-\phi d)$$

Con $E[\phi] = n/d$ y $V[\phi] = n/d^2$.

Un caso de interés es cuando $\phi \sim \mathbf{G}[n/2, d/2]$ si n es un entero positivo. Con esta distribución se tiene que $d\phi \sim \chi_n^2$, una ji cuadrada con n grados de libertad.

Se tiene $\phi \sim \mathbf{G}[n/2, d/2]$ para cualquier $n > 0$ y $d > 0$, y una variable aleatoria X que condicionada en ϕ tiene una distribución normal ($X | \phi$) $\sim \mathbf{N}[m, C\phi^{-1}]$ para algunas m y C . Es importante notar que X es una variable univariada. Así, se obtiene la distribución conjunta, conocida como normal-gamma:

Como $f(X | \phi) = \frac{f(X, \phi)}{f(\phi)}$, entonces $f(X, \phi) = f(X | \phi)f(\phi)$:

$$\begin{aligned} f(X, \phi) &= \left(\frac{\phi}{2\pi C} \right)^{1/2} \exp \left[-\frac{\phi(X - m)^2}{2C} \right] \times \frac{d^{n/2}}{2^{n/2} \Gamma(\frac{n}{2})} \phi^{\frac{n}{2}-1} \exp \left[-\frac{\phi d}{2} \right] \\ &\propto \phi^{(\frac{n+1}{2})-1} \exp \left[-\frac{\phi}{2} \left(\frac{(X - m)^2}{C} + d \right) \right]. \end{aligned}$$

Entonces,

$$f(\phi | X) = \frac{f(X, \phi)}{f(X)} \propto f(X, \phi) \propto \phi^{(\frac{n+1}{2})-1} \exp \left[-\frac{\phi}{2} \left(\frac{(X - m)^2}{C} + d \right) \right]$$

Por lo tanto,

$$(\phi \mid X) \sim \mathbf{G} \left[\frac{n^*}{2}, \frac{d^*}{2} \right],$$

con $n^* = n + 1$, $d^* = \frac{(X-m)^2}{C} + d$.

Por último, tras simplificaciones algebraicas y manipular la constante de proporcionalidad, se obtiene que

$$f(X) = \frac{f(X, \phi)}{f(\phi \mid X)} \propto \left[1 + \frac{1}{n} \frac{(X - m)^2}{R} \right]^{-(n+1)/2},$$

con $R = C(d/n)$. Esto es proporcional a la densidad de una distribución t de student generalizada con n grados de libertad, parámetro locacional m y parámetro de escala R . Entonces,

$$X \sim \mathbf{T}_n[m, R].$$

Ahora hay que considerar el caso multivariado. Si $\phi \sim \mathbf{G}[n/2, d/2]$ y \mathbf{X} es un vector $p \times 1$ con distribución normal dada ϕ , $\mathbf{X} \sim \mathbf{N}[\mathbf{m}, \mathbf{C}\phi^{-1}]$. \mathbf{m} es un vector $p \times 1$ conocido, mientras que \mathbf{C} es una matriz simétrica positiva definida $p \times p$ conocida. Entonces, cada elemento de $V[\mathbf{X}]$ esta multiplicado por ϕ . Los resultados son parecidos a los del caso univariado.

$$(\phi \mid \mathbf{X}) \sim \mathbf{G}[n^*/2, d^*/2],$$

con $n^* = n + p$ y $d^* = d + (\mathbf{X} - \mathbf{m})^\top \mathbf{C}^{-1}(\mathbf{X} - \mathbf{m})/2$.

Para la distribución de \mathbf{X} , se tiene

$$f(\mathbf{X}) \propto \left[1 + \frac{1}{n} (\mathbf{X} - \mathbf{m})^\top \mathbf{R}^{-1}(\mathbf{X} - \mathbf{m}) \right]^{-(n+p)/2},$$

con $\mathbf{R} = \mathbf{C}(d/n)$. Entonces, \mathbf{X} tiene una distribución marginal t multivariada con n grados de libertad, parámetro locacional \mathbf{m} y parámetro de escala \mathbf{R}

$$\mathbf{X} \sim \mathbf{T}_n[\mathbf{m}, \mathbf{R}].$$

Si X_i es el i -ésimo elemento de \mathbf{X} , m_i su media y C_{ii} el valor correspondiente de \mathbf{C} , entonces

$$X_i \sim \mathbf{T}_n[m_i, R_{ii}].$$

com $R_{ii} = C_{ii}(d/n)$

Bibliografía

- [1] *Historia de la moneda y del billete en México*. Banco de México. Disponible en: <https://www.banxico.org.mx/billetes-y-monedas/d/%7B1EEDFA6C-8EDB-B7AD-11B2-528C7B69CC76%7D.pdf>.
- [2] *Resultados de estudios cuantitativos y cualitativos sobre efectivo 2020*. Informe técnico, Dirección General de Emisión. Banco de México, 2020. Disponible en: <https://www.banxico.org.mx/billetes-y-monedas/estudios-e-indicadores/%7B397EE391-FF6B-A3F8-D668-6656C830E700%7D.pdf>.
- [3] *Encuesta Nacional de Inclusión Financiera 2021*. Informe técnico, Comisión Nacional Bancaria y de Valores, 2022. Disponible en: https://www.cnbv.gob.mx/Inclusi%C3%B3n/Anexos%20Inclusin%20Financiera/Reporte_Resultados_ENIF_2021.pdf.
- [4] B. Fisher, P. Köhler, F. Seitz: *The demand for Euro area currencies: past, present and future*. Informe técnico, European Central Bank, 2004. Disponible en: <https://www.oth-aw.de/files/oth-aw/Professoren/Seitz/ecbwp330.pdf>.
- [5] *Informe Trimestral Abril-Junio 2021*. Informe técnico, Banco de México, 2021. Disponible en: <https://www.banxico.org.mx/publications-and-press/quarterly-reports/%7B3E868135-A8FB-4DFE-56C7-46960617B22D%7D.pdf>.
- [6] *Agregados monetarios (metodología 2018) - (CF807)*. Sistema de Información Económica. Banco de México. Disponible en: <https://www.banxico.org.mx/SieInternet/>

consultarDirectorioInternetAction.do?sector=3&accion=consultarCuadro&idCuadro=CF807&locale=es.

- [7] Sriram, Subramanian: *Survey of Literature on Demand for Money: Theoretical and Empirical Work with Special Reference to Error-Correction Models*. IMF Working Paper No. 99/64, 1999. Disponible en: <https://ssrn.com/abstract=880591>.
- [8] Ibarra, Raul: *The Demand for Money in Mexico*. American Journal of Economics, 4(2A):73–80, 2014. Disponible en: <http://article.sapub.org/10.5923.s.economics.201401.06.html>.
- [9] Noriega-Muro, Antonio E.; Ramos-Francia, Manuel; Rodríguez-Pérez, Cid Alonso: *Estimaciones de la demanda de dinero en México y de su estabilidad 1986-2010, así como algunos ejemplos de sus usos*. Banco de México, Working Papers, No. 2015-13, 2015. Disponible en: <https://www.banxico.org.mx/publicaciones-y-prensa/documentos-de-investigacion-del-banco-de-mexico/%7B3AF49299-A804-3FBB-B812-8F55DD2C85E6%7D.pdf>.
- [10] M. West, J. Harrison: *Bayesian Forecasting and Dynamic Models*. Springer, Nueva York, segunda edición, 1997, ISBN 0-387-94725-6.