

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



Modelos lineales dinámicos con  
aplicaciones al pronóstico de la demanda  
de efectivo 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

2024

“Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada **“Modelos lineales dinámicos con aplicaciones al pronóstico de la demanda de efectivo en México”**, 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 tal divulgación una contraprestación.”

**Juan Pablo López Escamilla**

---

FECHA

---

FIRMA

*DEDICATORIA*

*A mis papás y mi hermana,  
por todo el amor y apoyo  
que me han brindado a lo  
largo de mi vida.*



# Agradecimientos

A mis papás, todo esto es posible gracias a ustedes, en todos mis logros y sueños ustedes son fundamentales, soy muy afortunado de ser su hijo. A mi hermana, por su desinteresado apoyo y por hacer divertidos todos los días. A Cau, Lola y Luca por su nobleza y compañía.

A mis abuelos, por su amor. A mi Mina, por ser uno de los pilares de mi vida. A mis tíos y primos, por su cariño. A mis tías Nina y Sofía, porque siempre están cerca.

A Tonantzin, por su sincera amistad, siempre estar presente y saber escuchar.

A Cañón, Chanona, Madrid y Pablo, por llevar casi 15 años en mi vida y compartir conmigo una fracción de todas sus semanas. Brauni.

A Ismael, Nicole y Santiago, por su esencial ayuda durante mis estudios y por seguir conviviendo después de nuestra graduación.

A los maestros que han aportado en mi formación. Es imposible mencionarlos a todos en este espacio.

A Jorge de la Vega, por asesorarme en la desafiante tarea de escribir una tesis y por sus extraordinarias clases de estadística. A Nicolás Gómez Castillo, Juan José Fernández Durán y Enrique Cuervo Guzmán, mis sinodales, por el tiempo que dedicaron en revisar mi tesis y sus valiosos comentarios.



# Índice general

<b>1. Introducción</b>	<b>1</b>
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 . . . . .	6
1.4. Objetivos de la tesis . . . . .	8
<b>2. Marco teórico</b>	<b>11</b>
2.1. Notación . . . . .	11
2.2. Modelo lineal dinámico . . . . .	12
2.3. Modelo lineal dinámico univariado . . . . .	14
2.4. Actualización de ecuaciones . . . . .	14
2.5. Distribuciones de pronóstico . . . . .	21
2.6. Varianzas desconocidas . . . . .	26
2.7. Análisis retrospectivo . . . . .	35
2.8. Intervenciones . . . . .	39
2.9. Modelo de espacio de estados . . . . .	48
<b>3. Pronóstico de la demanda agregada de efectivo</b>	<b>51</b>
3.1. Modelo lineal dinámico . . . . .	51
3.1.1. Datos . . . . .	53
3.1.2. Estimación de valores desconocidos . . . . .	54
3.1.3. Actualización . . . . .	57
3.1.4. Pronósticos . . . . .	59
3.1.5. Distribuciones filtradas . . . . .	61
3.2. Modelo lineal dinámico intervenido . . . . .	63
3.2.1. Intervención . . . . .	63

3.2.2.	Pronósticos con intervención . . . . .	66
3.2.3.	Distribuciones filtradas con intervención . . . . .	68
3.3.	VECM . . . . .	68
3.4.	Comparación . . . . .	73
3.4.1.	MSE . . . . .	74
3.4.2.	MAE . . . . .	74
3.4.3.	MAPE . . . . .	75
3.4.4.	U de Theil . . . . .	76
3.4.5.	Intervalos de pronóstico . . . . .	77
3.4.6.	Selección de modelo . . . . .	79
3.5.	Pronósticos 2023 y 2024 . . . . .	80
<b>4.</b>	<b>Conclusión</b>	<b>83</b>
4.1.	Resumen de los resultados . . . . .	83
4.2.	Limitaciones y próximos pasos . . . . .	84
<b>A.</b>	<b>Resultados de probabilidad y estadística</b>	<b>87</b>
A.1.	Distribuciones marginales y condicionales de la normal mul- tvariada . . . . .	87
A.2.	Distribución normal condicional y regresión lineal . . . . .	88
A.3.	Distribución normal-gamma . . . . .	90
<b>B.</b>	<b>VECM</b>	<b>93</b>
B.1.	Vector autorregresivo . . . . .	93
B.2.	Vector error correction model . . . . .	94
<b>C.</b>	<b>Gráficas de los modelos</b>	<b>97</b>
C.1.	Gráficas Modelo lineal dinámico . . . . .	97
C.2.	Gráficas Modelo lineal dinámico intervenido . . . . .	99
C.3.	Gráficas VECM . . . . .	100



# 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 con la mayor precisión posible 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 circulación. 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 tiempo,

debido a la importancia del Imperio Español en el contexto internacional, 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 fiduciaria<sup>1</sup> 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

---

<sup>1</sup>Una 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 para realizar transacciones.

república, la acuñación de monedas con el peso como unidad continuó. Durante el Porfiriato, gracias a un sistema bancario más avanzado y funcional, 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 privados 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 se generalizó de nuevo la aceptación de los billetes [1].

Las monedas han sido utilizadas en territorio mexicano por casi 500 años, mientras que los billetes, a pesar de ser rechazados 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 financiero<sup>2</sup> [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]. Otro problema que es necesario abordar es que las distintas denominaciones de billetes y monedas deben distribuirse en cantidades adecuadas para que sea posible realizar transacciones. Por ejemplo, la falta de billetes de \$500 o de monedas de \$10 pueden dificultar ciertas transacciones. Por esta razón, es necesario pronosticar las demandas de cada denominación.

El uso de billetes y monedas se amplificó durante la pandemia de

---

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

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. El inicio de la pandemia, en marzo de 2020, está señalado con una línea roja vertical. 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 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.

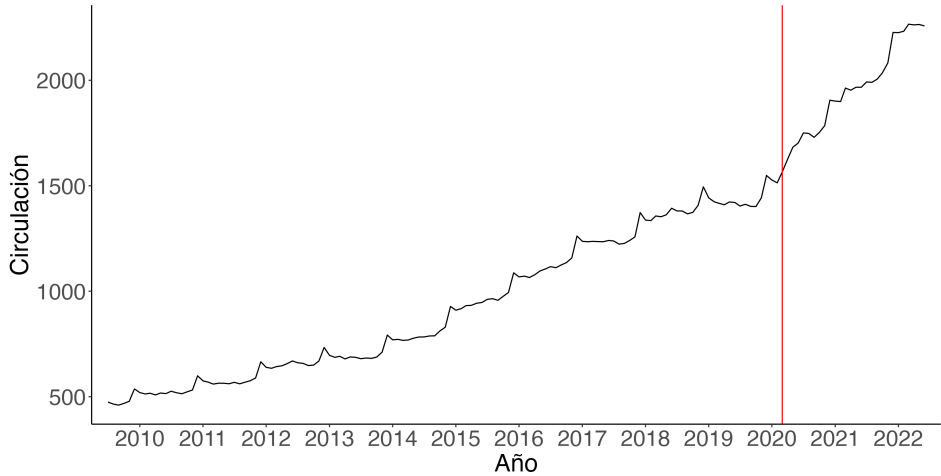


Figura 1.1: Billetes y monedas en poder del público (miles de millones de pesos).

Fuente: Banco de México [6]

### 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.

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 definición de dinero es amplia, no se refiere únicamente al efectivo. Dinero es todo bien aceptado por las personas para comprar bienes y servicios, y para pagar deudas. Un ejemplo de dinero, además del efectivo, son las tarjetas de débito y las transferencias electrónicas [7]. Por esta razón, cuando se trata de modelar la demanda de dinero, 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 <sup>3</sup>. 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 el 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

---

<sup>3</sup>Los 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]

por otros activos es limitada y es posible que no existan datos de las tasas de interés. La gran mayoría de los estudios estiman que la relación entre demanda de efectivo e inflación es negativa porque las personas prefieren adquirir activos reales para cubrirse de la inflación en lugar de efectivo. Sin embargo, otros estudios encuentran relaciones positivas entre dinero e inflación porque al subir los precios las personas necesitan más dinero [8, p. 10]. 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 utilizados anteriormente, los cuales se enfocaban en la teoría económica sin examinar los datos a profundidad. [9]

En la actualidad los ECM conservan su importancia en el análisis de la demanda de dinero. Ibarra [10] empleó un ECM para estimar una función de demanda de dinero en México para el periodo 1980-2005. Ese 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* [11] realizaron un estudio parecido para el periodo 2006-2010.

## 1.4. Objetivos de la tesis

La principal meta de este trabajo es introducir los modelos lineales dinámicos y mostrar su utilidad en la realización de pronósticos en contextos complejos. Se ejemplifica la aplicación de los modelos lineales dinámicos en el pronóstico de la demanda de efectivo en México. Para determinar sus



ventajas se compara su desempeño contra un modelo que es frecuentemente aplicado en el estudio de la demanda de dinero: el *vector error correction model* (VECM). La variable de interés es el agregado M0 nominal, 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.

En el capítulo 2 se presenta la notación y teoría de los modelos lineales dinámicos. En específico, se describe el proceso para la actualización de los modelos al recibir nueva información, el cálculo de pronósticos, la realización inferencias sobre periodos pasados, y se introduce una forma de intervención para incorporar información externa al modelo que permita mejorar su desempeño.

En el capítulo 3 se muestra la aplicación de los modelos lineales dinámicos al pronóstico de la demanda de efectivo en México, la precisión de los pronósticos, y se comparan tres modelos: un modelo lineal dinámico; un modelo similar, pero con una intervención al inicio de la pandemia; y un VECM. Finalmente, se escoge un modelo ganador y se pronostica la demanda de efectivo trimestral de 2023 y 2024 con este modelo.

Por último, en el capítulo 4 se resumen brevemente los resultados y se discuten las limitaciones de esta tesis y próximos pasos.



# 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 [12]. Por lo tanto, la notación empleada es la misma que la de ese 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. La serie de tiempo puede ser univariada o multivariada.

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, y por lo tanto, es dinámica, cambia a lo largo del tiempo. Con esta formulación probabilística se pueden asimilar fácilmente cambios no anticipados en las condiciones del proceso y a sus incertidumbres asociadas. Por ejemplo, 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 de espacio de estado.

**Definición 2.2.1** (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, a través de las siguientes ecuaciones:

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

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

Las secuencias de errores  $\boldsymbol{\omega}_t$  y  $\boldsymbol{\nu}_t$  son 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 o de parámetros.
- $\boldsymbol{\mu}_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t$  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  $\omega_t$  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 2.2.1 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 2.3.1** (Modelo Lineal Dinámico univariado). Para cada  $t$ , el modelo lineal dinámico univariado se define así:

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

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

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

donde  $\mathbf{m}_0$  y  $\mathbf{C}_0$  son momentos previos, las secuencias de errores  $\omega_t$  y  $\nu_t$  son independientes, y además son independientes de  $(\boldsymbol{\theta}_0 \mid D_0)$ .

## 2.4. 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 previa  $D_0$  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 2.4.1.** *Para el Modelo lineal dinámico univariado de la definición 2.3.1, 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. *Previa 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*

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

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

*Demostración.* [12, pp. 104-105]

La prueba es por inducción. En la demostración,  $\mathbf{Y}_t$  es un vector debido a que la prueba es valida 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:

$$\begin{aligned}\mathbf{Y}_1 &= \mathbf{F}_1^\top \boldsymbol{\theta}_1 + \boldsymbol{\nu}_1, & \boldsymbol{\nu}_1 &\sim \mathbf{N}[\mathbf{0}, \mathbf{V}_1]. \\ \boldsymbol{\theta}_1 &= \mathbf{G}_1 \boldsymbol{\theta}_0 + \boldsymbol{\omega}_1, & \boldsymbol{\omega}_1 &\sim \mathbf{N}[\mathbf{0}, \mathbf{W}_1].\end{aligned}$$

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

Entonces,

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

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  $\mathbf{Y}_1$ :

$$\left( \begin{array}{c} \boldsymbol{\theta}_1 \\ \mathbf{Y}_1 \end{array} \middle| D_0 \right) \sim \mathbf{N} \left[ \left( \begin{array}{c} \mathbf{a}_1 \\ \mathbf{F}_1^\top \mathbf{a}_1 \end{array} \right), \left( \begin{array}{cc} \mathbf{R}_1 & \mathbf{R}_1 \mathbf{F}_1 \\ \mathbf{F}_1^\top \mathbf{R}_1 & \mathbf{F}_1^\top \mathbf{R}_1 \mathbf{F}_1 + \mathbf{V}_1 \end{array} \right) \right].$$

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

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

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

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

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

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

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

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

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

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:

$$\begin{aligned} \mathbf{Y}_{t+1} &= \mathbf{F}_{t+1}^\top \boldsymbol{\theta}_{t+1} + \boldsymbol{\nu}_{t+1}, & \boldsymbol{\nu}_{t+1} &\sim \mathbf{N}[\mathbf{0}, \mathbf{V}_{t+1}]. \\ \boldsymbol{\theta}_{t+1} &= \mathbf{G}_{t+1} \boldsymbol{\theta}_t + \boldsymbol{\omega}_{t+1}, & \boldsymbol{\omega}_{t+1} &\sim \mathbf{N}[\mathbf{0}, \mathbf{W}_{t+1}]. \end{aligned}$$

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  $\mathbf{Y}_{t+1}$ :

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

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  $\mathbf{C}_{t+1} = \mathbf{R}_{t+1} - \mathbf{A}_{t+1} \mathbf{Q}_{t+1} \mathbf{A}_{t+1}^\top$ .

□

Este mecanismo de actualización de los parámetros cuando se obtienen nuevas observaciones se conoce como Filtro de Kalman en el contexto de los modelos de espacio de estado, introducidos por primera vez por Rudolf E. Kalman en 1960. El desarrollo de estos modelos se originó en ingeniería, pero con el paso de las décadas comenzaron a ser estudiados y aplicados en los campos de estadística y economía [13, pp. 51-52].

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 previa. Así es como la distribución posterior se ajusta con base en la  $Y_t$  observada.

**Ejemplo 2.4.1.** En este ejemplo se utilizará la base de datos **freeny** disponible en el paquete **datasets** de R para ilustrar el Teorema 2.4.1. En este conjunto de datos se tiene una variable  $y$  que representa la serie de tiempo de las ganancias trimestrales desde 1962 T2 hasta 1971 T4. 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

$$Y_t = \text{Ganancias en } t, \quad \boldsymbol{\theta}_t = \begin{pmatrix} \text{Intercepto en } t \\ \text{Ingreso en } t \\ \text{Precio en } t \end{pmatrix}.$$

En este contexto ficticio, estamos ubicados en el cuarto trimestre de 1966. La serie empieza en 1962 T2, entonces 1966 T4 equivale a  $t = 19$ . Las ganancias en 1966 T4 ya fueron observadas, mientras que de 1967 T1 en adelante no se han observado, por lo que solo se tiene la información  $D_{19}$ , que incluye las observaciones de 1966 T4 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:

$$(\boldsymbol{\theta}_{19} \mid D_{19}) \sim \mathbf{N} \left[ \begin{pmatrix} 1.5 \\ 1.8 \\ -0.7 \end{pmatrix}, \begin{pmatrix} 0.00002 & 0.00001 & -0.00002 \\ 0.00001 & 0.00003 & -0.00001 \\ -0.00002 & -0.00001 & 0.00002 \end{pmatrix} \right].$$

Ahora se debe obtener la distribución previa 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. Se obtiene la siguiente distribución:

$$(\boldsymbol{\theta}_{20} \mid D_{19}) \sim \mathbf{N} \left[ \begin{pmatrix} 1.501 \\ 1.8 \\ -0.7 \end{pmatrix}, \begin{pmatrix} 0.00003 & 0.00001 & -0.00002 \\ 0.00001 & 0.00004 & -0.00002 \\ -0.00002 & -0.00002 & 0.00007 \end{pmatrix} \right].$$

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

$$(Y_{20} \mid D_{19}) \sim \mathbf{N}[9.254, 0.001821].$$

Con esta distribución, es posible calcular un intervalo de pronóstico del 95 % para  $Y_{20}$ :  $[9.17, 9.338]$ .

Supongamos que, unos meses después, se observan las ganancias en 1967 T1 (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  $t = 20$ . La distribución posterior de los parámetros en  $t = 20$  es

$$(\boldsymbol{\theta}_{20} \mid D_{20}) \sim \mathbf{N} \left[ \begin{pmatrix} 1.5015 \\ 1.8053 \\ -0.6943 \end{pmatrix}, \begin{pmatrix} 0.00003 & 0.00001 & -0.00002 \\ 0.00001 & 0.000026 & -0.000036 \\ -0.00002 & -0.000036 & 0.000053 \end{pmatrix} \right].$$

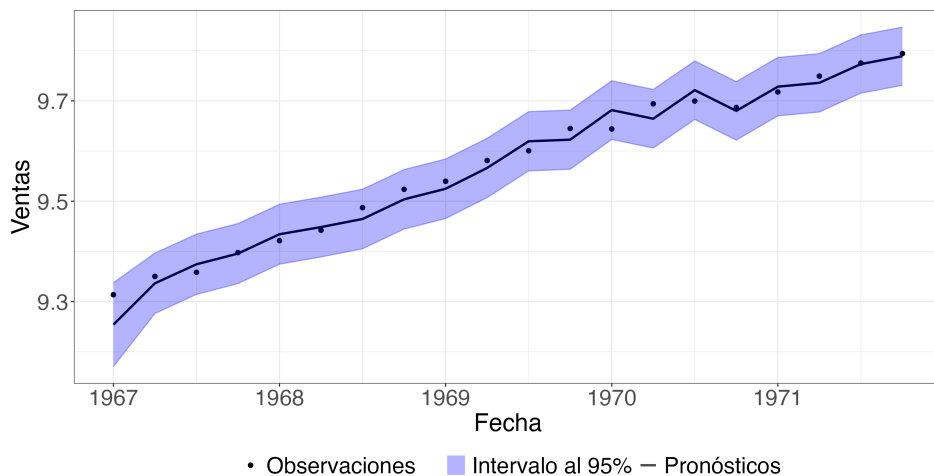


Figura 2.1: Pronósticos a un paso ( $Y_t | D_{t-1}$ ).

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

Con el mismo procedimiento se pueden obtener los pronósticos a un paso para toda  $t$  una vez que se cuente con la información  $D_{t-1}$ . En la figura 2.1 se muestran los pronósticos a un paso ( $Y_t | D_{t-1}$ ) con los datos de este ejemplo desde 1967 T1 ( $t = 20$ ) hasta 1971 T3 ( $t = 39$ ). En esta gráfica también se muestra el intervalo al 95 %. Se supone que  $\mathbf{G}_t$ ,  $\mathbf{W}_t$  y  $V_t$  son constantes a lo largo de toda la serie.

Por último, en la figura 2.2 se observa la evolución de los parámetros a lo largo del tiempo. Los pronósticos fueron bastante acertados, por lo tanto el modelo no realizó grandes cambios en el valor de los parámetros.

## 2.5. 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.

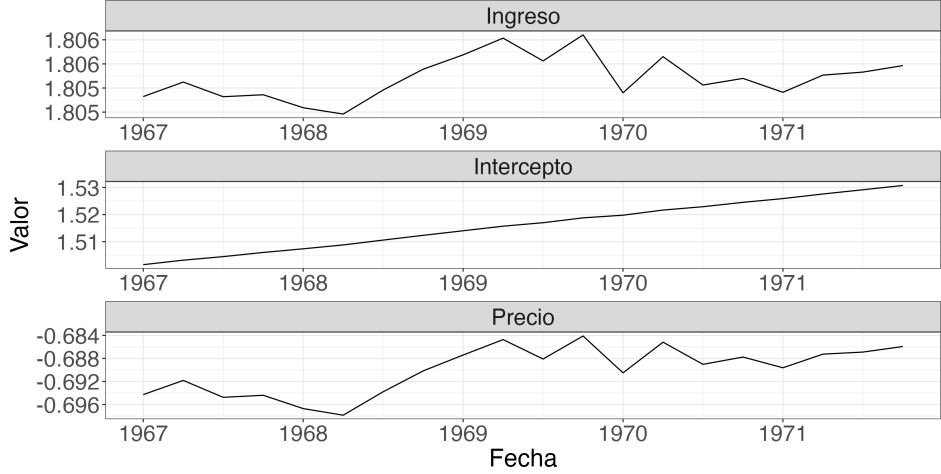


Figura 2.2: Valores de  $E[\theta_t \mid D_t] = \mathbf{m}_t$ .

**Definición 2.5.1** (Función de pronóstico). La función de pronóstico  $f_t(k)$ , en cualquier tiempo  $t$ , se define para todos los enteros  $k \geq 0$  como

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

donde  $\mu_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t$  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  $Y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t$ , el cual nunca será igual que  $\mathbf{F}_t^\top \boldsymbol{\theta}_t$  por la presencia del error  $\nu_t$  con distribución normal.

La definición de la función de pronóstico esta formulada en términos de  $\mu_{t+k}$  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 2.5.2.** *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.* [12, 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  $k$  veces hacia atrás, se obtiene

$$\boldsymbol{\theta}_{t+k} = \mathbf{H}_{t+k}(k) \boldsymbol{\theta}_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r) \boldsymbol{\omega}_{t+r}.$$

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

$$E[\boldsymbol{\theta}_{t+k} \mid D_t] = E \left[ \mathbf{H}_{t+k}(k) \boldsymbol{\theta}_t + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r) \boldsymbol{\omega}_{t+r} \mid D_t \right]$$

$$\begin{aligned}
&= \mathbf{H}_{t+k}(k)E[\boldsymbol{\theta}_t \mid D_t] + \sum_{r=1}^k \mathbf{H}_{t+k}(k-r) \cancel{E[\boldsymbol{\omega}_{t+r} \mid D_t]} \xrightarrow{0} \\
&= \mathbf{H}_{t+k}(k)\mathbf{m}_t \quad (\text{por Teorema 2.4.1}).
\end{aligned}$$

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

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

Entonces,

$$\mathbf{R}_t(k) = V[\boldsymbol{\theta}_{t+k} \mid D_t] = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k}.$$

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 2.4.1.

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,

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

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

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

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

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

□

**Ejemplo 2.5.1.** Se parte del final del ejemplo pasado. Estamos ubicados en 1967 T1, 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 el ejemplo pasado. Después se debe definir la distribución de  $(\boldsymbol{\theta}_{21} \mid D_{20})$  y posteriormente se calcula la distribución de  $(Y_{21} \mid D_{20})$ . La

distribución es  $(Y_{21} \mid D_{20}) \sim \mathcal{N}[9.336, 0.0009445]$ . Ahora, con los resultados de  $Y_{21}$ , se obtiene la distribución de pronóstico de  $Y_{22}$ , y así sucesivamente durante  $k$  periodos futuros.

En la figura 2.3 se muestran los pronósticos  $(Y_{20+k} \mid D_{20})$  para  $k = 1, \dots, 19$ . Es evidente que los intervalos se vuelven más anchos entre más lejanos son los pronósticos. Este comportamiento es normal debido a que hay más incertidumbre en los periodos futuros más distantes.

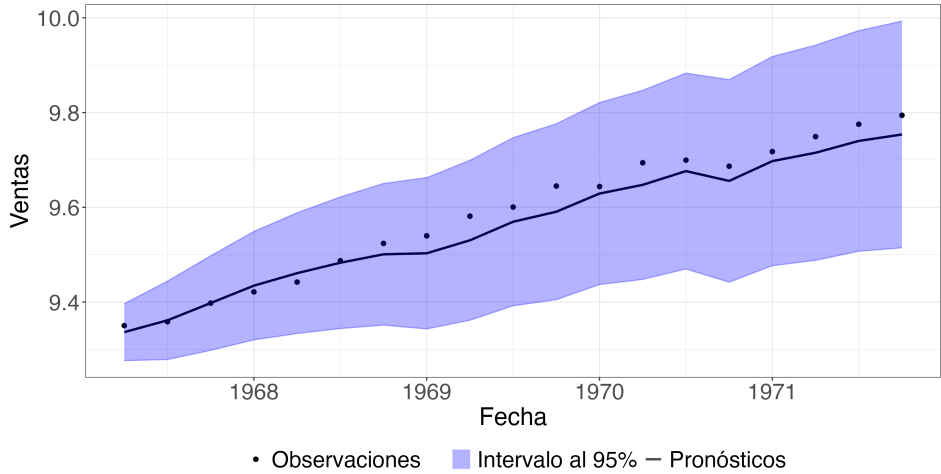


Figura 2.3: Pronósticos a  $k$  pasos  $(Y_{20+k} \mid D_{20})$ .

## 2.6. 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. 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 2.6.1.** Para cada  $t$ , si  $\phi = V^{-1}$ , el modelo lineal dinámico se define como

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

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

$$\begin{aligned} \text{Información inicial:} \quad & (\boldsymbol{\theta}_0 \mid D_0, \phi) \sim \mathbf{N}[\mathbf{m}_0, V \mathbf{C}_0^*], \\ & (\phi \mid D_0) \sim \mathbf{G} \left[ \frac{n_0}{2}, \frac{n_0 S_0}{2} \right]. \end{aligned}$$

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 2.3.1, 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 2.6.2.** *Se tienen las siguientes distribuciones para el modelo lineal dinámico de la definición 2.6.1:*

*A. Condicional en  $V$ :*

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

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  $\phi = V^{-1}$ :

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

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

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

C. No condicionado en  $V$ :

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

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

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

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

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

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

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

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t, \quad \mathbf{C}_t = \frac{S_t}{S_{t-1}} (\mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t^\top Q_t).$$

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

*Demostración.* [12, pp. 110-111]

El punto A. del teorema es el mismo caso del teorema 2.4.1. 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. Primero se debe demostrar la validez para  $t = 1$ . Se omite esta parte.

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

El punto A. es equivalente al teorema 2.4.1.

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,

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

Por el teorema de Bayes, la distribución posterior de  $\phi$  es

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

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

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

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

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

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

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  $\phi$  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  $V$ .

La ecuación de actualización de  $S_t$  tiene implicaciones bastante lógicas e interesantes de analizar. Se puede ver que la magnitud del ajuste depende de  $\frac{e_t^2}{Q_t} - 1$ . Entonces, si el error es grande en comparación con la varianza de  $Y_t$ , el valor de  $S_t$  será considerablemente mayor. Otro punto importante es que la magnitud del ajuste es inversamente proporcional a los grados de libertad. Entonces, entre más información por observaciones pasadas se tenga, la magnitud del ajuste será menor.

**Ejemplo 2.6.1.** 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 2.4.1.

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

$$(\boldsymbol{\theta}_{19} \mid D_{19}) \sim \mathbf{T}_{19.5} \left[ \begin{pmatrix} 1.5 \\ 1.8 \\ -0.7 \end{pmatrix}, \begin{pmatrix} 0.00002 & 0.00001 & -0.00002 \\ 0.00001 & 0.00003 & -0.00001 \\ -0.00002 & -0.00001 & 0.00002 \end{pmatrix} \right].$$

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 previa de los parámetros en  $t = 20$ :

$$(\boldsymbol{\theta}_{20} \mid D_{19}) \sim \mathbf{T}_{19.5} \left[ \begin{pmatrix} 1.501 \\ 1.8 \\ -0.7 \end{pmatrix}, \begin{pmatrix} 0.00003 & 0.00001 & -0.00002 \\ 0.00001 & 0.00004 & -0.00002 \\ -0.00002 & -0.00002 & 0.00007 \end{pmatrix} \right],$$

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  $V = 0.00005$ , aquí supongamos que  $S_{19}$  fue obtenida anteriormente y su valor es  $S_{19} = 0.00005$  con el fin de ilustrar lo parecidos que son ambos casos. Entonces,

$$(Y_{20} \mid D_{19}) \sim \mathbf{T}_{19.5}[9.254, 0.001821].$$

Ahora es posible calcular el intervalo de pronóstico del 95 % para  $Y_{20}$ : [9.165, 9.343].

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  $S_{20}$ . Entonces, la distribución posterior de los parámetros en  $t = 20$  es



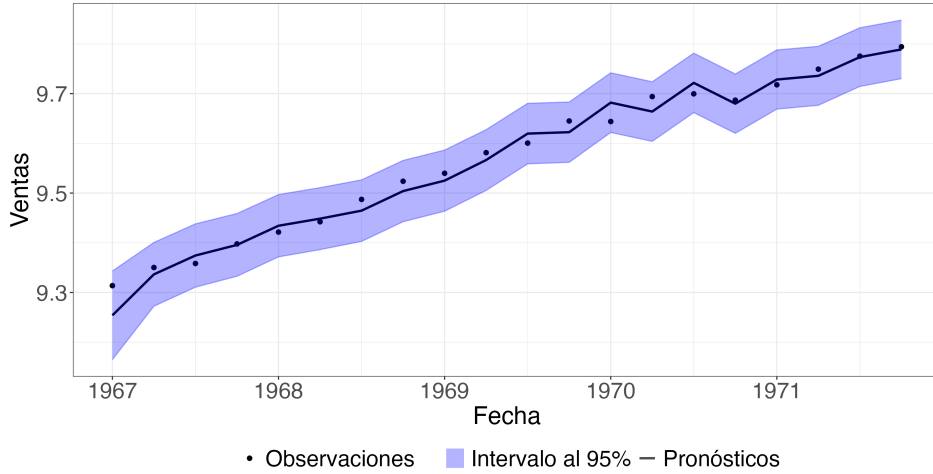


Figura 2.4: Pronósticos a un pasos ( $Y_t | D_{t-1}$ ) con  $V$  desconocida.

$$(\theta_{20} | D_{20}) \sim \mathbf{T}_{20.5} \left[ \begin{pmatrix} 1.5015 \\ 1.8053 \\ -0.6943 \end{pmatrix}, \begin{pmatrix} 0.000031 & 0.000001 & -0.000021 \\ 0.000001 & 0.000027 & -0.000037 \\ -0.000021 & -0.000037 & 0.000056 \end{pmatrix} \right].$$

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  $V$ .

En la figura 2.4 se muestran los pronósticos a un paso de todos los periodos. Esta gráfica es extremadamente parecida a la del caso con  $V$  conocida, es difícil apreciar visualmente diferencias entre ambas. Por último, en la figura 2.5 se muestra la evolución de la estimación de  $V$ , a lo largo de la serie. La estimación disminuye en casi toda la serie debido a que los errores de pronóstico son pequeños en comparación con la variabilidad de la distribución de  $Y_t$  (las observaciones reales ni siquiera se acercan a los límites del intervalo en la figura 2.4).

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

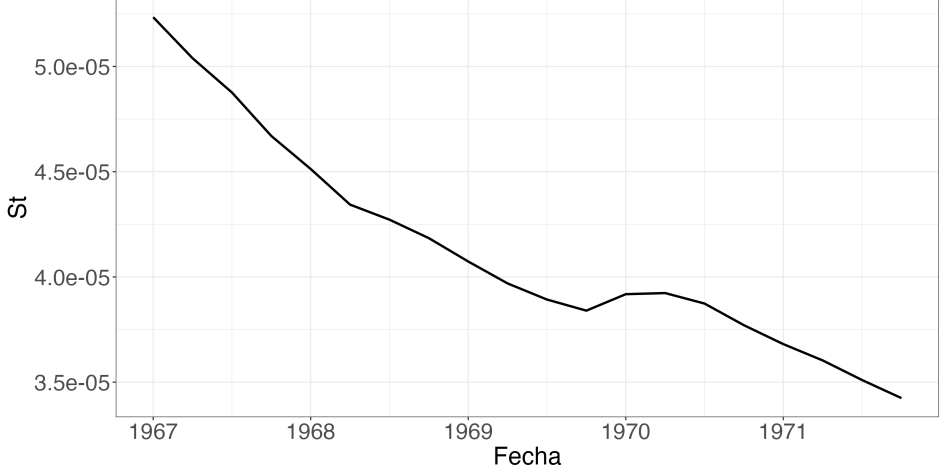


Figura 2.5: Actualización de  $S_t$ .

$V$  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  $V_{t+k}$  cuando se conoce  $D_t$ . 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.

**Corolario 2.6.1.** *Con el modelo de la definición 2.6.1 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:

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

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

## 2.7. 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 para las distribuciones de pronóstico.

**Teorema 2.7.1.** *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.* [12, pp. 114-115]

La demostración es por inducción. Se omite la demostración de la base de inducción.

**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,

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

Entonces, para obtener  $f(\boldsymbol{\theta}_{t-k-1} \mid D_t)$  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 2.4.1 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:

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

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

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

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

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 2.7.1.** *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  $\mu_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t$  a través de las distribuciones filtradas  $(\mu_{t-k} \mid D_t)$  se le conoce como suavizamiento.

**Corolario 2.7.2.** *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 2.7.1.** Supongamos que pasaron los años, nos encontramos en 1971 T3 y ya se conoce el valor de  $Y_{39}$ . Recordemos que  $\mathbf{a}_{39}(0) = \mathbf{m}_{39}$  y  $\mathbf{R}_{39}(0) = \mathbf{C}_{39}$ , entonces para obtener la distribución filtrada  $(\boldsymbol{\theta}_{38} \mid D_{39})$  se necesitan los valores que se obtuvieron de  $\mathbf{m}_{39}$ ,  $\mathbf{m}_{38}$ ,  $\mathbf{C}_{39}$ ,  $\mathbf{C}_{38}$ ,  $\mathbf{a}_{39}$  y  $\mathbf{R}_{39}$ . Entonces se llega a

$$(\boldsymbol{\theta}_{38} \mid D_{39}) \sim \mathbf{T}_{39.5} \left[ \begin{pmatrix} 1.5292 \\ 1.8059 \\ -0.6869 \end{pmatrix}, \begin{pmatrix} 0.000166 & 0.000003 & -0.000043 \\ -0.000003 & 0.000155 & -0.000224 \\ -0.000043 & -0.000224 & 0.000336 \end{pmatrix} \right].$$

Ahora es posible obtener fácilmente la distribución suavizada para la respuesta media, solo es necesario utilizar el valor de las variables explicativas en  $t = 38$ :  $F_{38} = [1, 6.19377, 4.27839]^\top$ . La distribución suavizada de la respuesta media es

$$(\mu_{38} \mid D_{39}) \sim \mathbf{T}_{39.5}[9.775, 0.000031].$$

De manera iterativa se pueden calcular la distribución suavizada de la respuesta media para los periodos anteriores. Los resultados se muestran en la figura 2.6.

## 2.8. 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

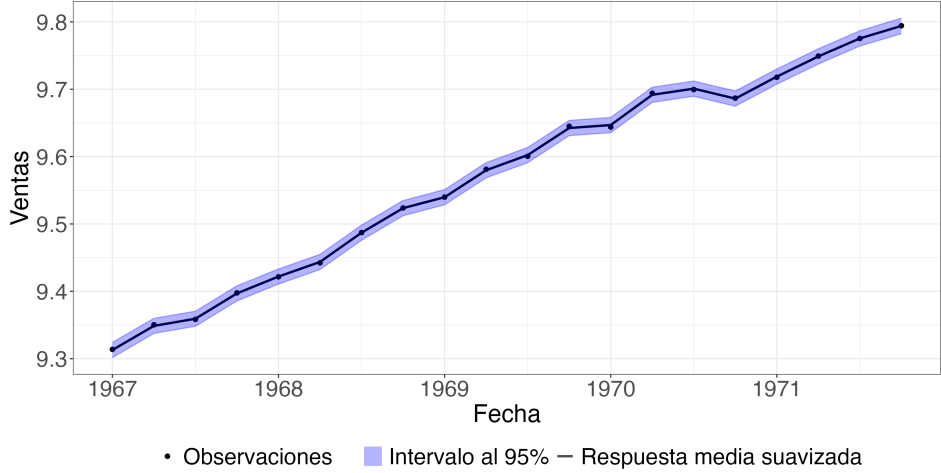


Figura 2.6: Respuesta media suavizada ( $\mu_{39-k} \mid D_{39}$ ).

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 2.3.1 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 2.4.1 se obtuvo que la distribución previa de los parámetros en  $t$  es

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

La intervención consiste en cambiar los momentos de la distribución previa de  $\boldsymbol{\theta}_t$  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 previa de los parámetros post-intervención es

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

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  $\boldsymbol{\theta}_t$  y  $\boldsymbol{\theta}_{t-1}$  condicionadas en  $D_{t-1}$  y en  $I_t$ , pero las modificaciones arbitrarias en los momentos de  $\boldsymbol{\theta}_t$  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 2.8.1.** *Sea  $\mathbf{K}_t$  una matriz cuadrada, triangular inferior, y  $\mathbf{h}_t$  cualquier vector. Se define*

$$\boldsymbol{\theta}_t^* = \mathbf{K}_t \boldsymbol{\theta}_t + \mathbf{h}_t,$$

*donde  $E[\boldsymbol{\theta}_t] = \mathbf{a}_t$  y  $V[\boldsymbol{\theta}_t] = \mathbf{R}_t$ . Entonces  $\boldsymbol{\theta}_t^*$  tiene momentos  $\mathbf{a}_t^*$  y  $\mathbf{R}_t^*$  si  $\mathbf{K}_t$  y  $\mathbf{h}_t$  se escogen de la siguiente manera:*

$$\mathbf{K}_t = \mathbf{U}_t \mathbf{Z}_t^{-1},$$

$$\mathbf{h}_t = \mathbf{a}_t^* - \mathbf{K}_t \mathbf{a}_t,$$

donde  $\mathbf{U}_t$  y  $\mathbf{Z}_t$  son las factorizaciones de Cholesky de  $\mathbf{R}_t^*$  y  $\mathbf{R}_t$ , entonces  $\mathbf{R}_t^* = \mathbf{U}_t \mathbf{U}_t^\top$  y  $\mathbf{R}_t = \mathbf{Z}_t \mathbf{Z}_t^\top$ . Las matrices  $\mathbf{U}_t$  y  $\mathbf{Z}_t$  son triangular inferior, únicas e invertibles.

*Demostración.* [12, p. 379]

$\mathbf{R}_t$  y  $\mathbf{R}_t^*$  son matrices de covarianza, entonces son simétricas y positivas definidas. Por lo tanto, su descomposición de Cholesky es  $\mathbf{R}_t^* = \mathbf{U}_t \mathbf{U}_t^\top$  y  $\mathbf{R}_t = \mathbf{Z}_t \mathbf{Z}_t^\top$ , donde  $\mathbf{U}_t$  y  $\mathbf{Z}_t$  son triangular inferior, únicas e invertibles.

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

$$E[\boldsymbol{\theta}_t^*] = \mathbf{a}_t^* = \mathbf{K}_t E[\boldsymbol{\theta}_t] + \mathbf{h}_t = \mathbf{K}_t \mathbf{a}_t + \mathbf{h}_t.$$

Entonces,

$$\mathbf{h}_t = \mathbf{a}_t^* - \mathbf{K}_t \mathbf{a}_t$$

para cualquier  $\mathbf{K}_t$  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,  $\mathbf{K}_t = \mathbf{U}_t \mathbf{Z}_t^{-1}$ .  $\square$

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 2.8.1.** *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 2.8.1. Entonces, la distribución previa 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 2.8.1 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  $\boldsymbol{\omega}_t^*$  es

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

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

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

Por independencia de  $\boldsymbol{\theta}_{t-1}$  y  $\boldsymbol{\omega}_t^*$ ,

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

Por lo tanto,

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

□

**Ejemplo 2.8.1.** 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 2.8.1, se realizará una intervención a lo realizado en el ejemplo 2.6.1 en  $t = 30$ . Supongamos que ya se observó  $Y_{29}$ , por lo que ya se cuenta con  $D_{29}$ . La distribución de  $\theta_{30}$  es

$$(\theta_{30} \mid D_{29}) \sim \mathbf{T}_{29.5} \left[ \begin{pmatrix} 1.5172 \\ 1.8065 \\ -0.6847 \end{pmatrix}, \begin{pmatrix} 0.000113 & 0.000006 & -0.000032 \\ 0.000006 & 0.000106 & -0.000145 \\ -0.000032 & -0.000145 & 0.000246 \end{pmatrix} \right].$$

Ahora, supongamos durante el tercer trimestre de 1969 ( $t = 30$ ) 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 1.8065 a 1.9. 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 0.000106 a 0.0002. Es decir,

$$\mathbf{a}_{30} = \begin{pmatrix} 1.5172 \\ 1.8065 \\ -0.6847 \end{pmatrix}, \quad \mathbf{R}_{30} = \begin{pmatrix} 0.000113 & 0.000006 & -0.000032 \\ 0.000006 & 0.000106 & -0.000145 \\ -0.000032 & -0.000145 & 0.000246 \end{pmatrix}.$$

$$\mathbf{a}_{30}^* = \begin{pmatrix} 1.5172 \\ 1.9 \\ -0.6847 \end{pmatrix}, \quad \mathbf{R}_{30}^* = \begin{pmatrix} 0.000113 & 0.000006 & -0.000032 \\ 0.000006 & 0.0002 & -0.000145 \\ -0.000032 & -0.000145 & 0.000246 \end{pmatrix}.$$

Por lo tanto,

$$(\theta_{30} \mid D_{29}, I_{30}) \sim \mathbf{T}_{29.5} \left[ \begin{pmatrix} 1.5172 \\ 1.9 \\ -0.6847 \end{pmatrix}, \begin{pmatrix} 0.000113 & 0.000006 & -0.000032 \\ 0.000006 & 0.0002 & -0.000145 \\ -0.000032 & -0.000145 & 0.000246 \end{pmatrix} \right].$$

Con esta intervención, se calculará la distribución del pronóstico de  $Y_{30}$  con el procedimiento usual, solo que ahora se utilizan  $\mathbf{a}_{30}^*$  y  $\mathbf{R}_{30}^*$  en lugar de

$\alpha_{30}$  y  $R_{30}$ . Entonces, la distribución del pronóstico de  $Y_{30}$  es

$$(Y_{30} \mid D_{29}, I_{30}) \sim \mathbf{T}_{29.5}[10.19, 0.004455].$$

Con esta distribución se calcula el siguiente intervalo de pronóstico del 95 % para  $Y_{30}$ :  $[10.06, 10.33]$ .

Aquí se aprecia el resultado de la intervención. Sin intervención, en el ejemplo 2.6.1 el intervalo para  $Y_{30}$  hubiera sido  $[9.559, 9.68]$ . Con la intervención, los valores del intervalo son 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  $Y_{30}$ . El valor observado fue de 10.0556, ligeramente debajo del intervalo. Entonces, ahora que se cuenta con la información  $D_{30}$ , hay que actualizar la distribución de  $\theta_{30}$ .

$$(\theta_{30} \mid D_{30}) \sim \mathbf{T}_{30.5} \left[ \begin{pmatrix} 1.5169 \\ 1.8813 \\ -0.6897 \end{pmatrix}, \begin{pmatrix} 0.000125 & 0.000005 & -0.000036 \\ 0.000005 & 0.000132 & -0.000184 \\ -0.000036 & -0.000184 & 0.000267 \end{pmatrix} \right].$$

Como  $Y_{30}$  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 de 1.9 a 1.88.

En la figura 2.7 se muestran los pronósticos a un paso ( $Y_t \mid D_{t-1}$ ) con la intervención. Es evidente que la intervención permite al modelo capturar correctamente el brinco que la serie dio en 1969 T3. Por otro lado, la figura 2.8 muestra los pronósticos a un paso del modelo sin intervenirlo. Este modelo no tiene forma de anticipar el gran brinco de las ventas, por lo que el error fue muy grande en ese periodo, aunque la naturaleza dinámica del modelo le permite ajustarse correctamente para el siguiente periodo.

Hasta ahora no se han utilizado los resultados del lema 2.8.1 y del teorema 2.8.1 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 como se muestra a continuación.

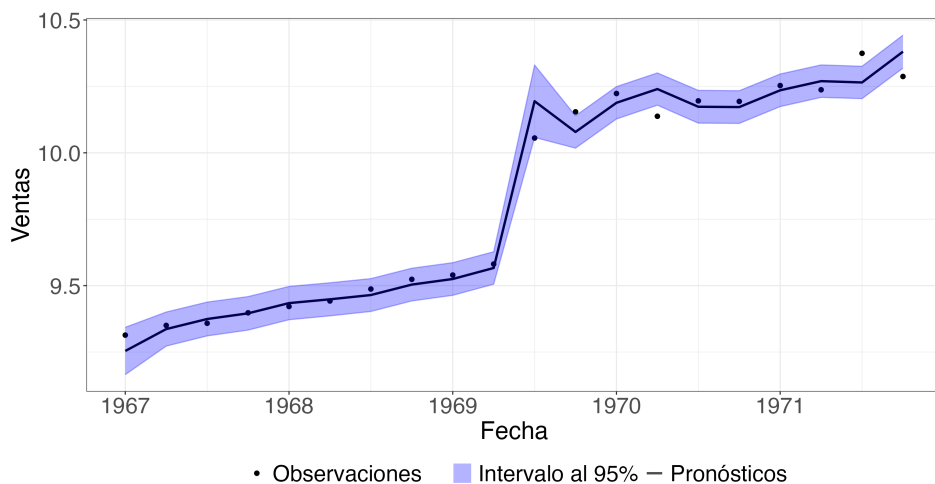


Figura 2.7: Pronósticos a un paso ( $Y_t | D_{t-1}$ ) con intervención.

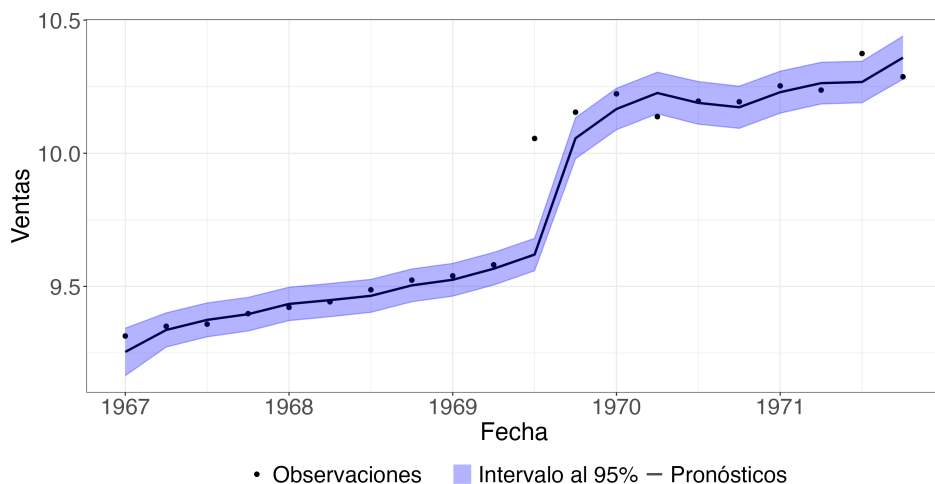


Figura 2.8: Pronósticos a un paso ( $Y_t | D_{t-1}$ ) sin intervención.

Supongamos que pasaron los meses y ya se conoce el valor de  $Y_{39}$ . El objetivo es obtener la distribución filtrada ( $\theta_{29} | D_{39}$ ) y la respuesta media suavizada ( $\mu_{29} | D_{39}$ ) considerando la intervención que se realizó en  $t = 30$ .

Primero hay que calcular las descomposiciones de Cholesky de  $\mathbf{R}_{30}$  y  $\mathbf{R}_{30}^*$ , con esto es posible calcular  $\mathbf{K}_{30}$ . Posteriormente, se calcula  $\mathbf{G}_{30}^*$  y se utiliza en lugar de  $\mathbf{G}_{30}$ . Con esto, se obtiene la distribución

$$(\boldsymbol{\theta}_{29} \mid D_{39}) \sim \mathbf{T}_{39.5} \left[ \begin{pmatrix} 1.5160 \\ 1.7960 \\ -0.6706 \end{pmatrix}, \begin{pmatrix} 0.000584 & 0.000042 & -0.000192 \\ 0.000042 & 0.000363 & -0.000523 \\ -0.000192 & -0.000523 & 0.000788 \end{pmatrix} \right].$$

Después se obtiene la siguiente distribución suavizada para la respuesta media  $(\mu_{29} \mid D_{39})$ :

$$(\mu_{29} \mid D_{39}) \sim \mathbf{T}_{39.5}[9.578, 0.0001].$$

En la figura 2.9 se muestran las respuestas medias suavizadas. El brusco incremento en las ventas se capturó correctamente.

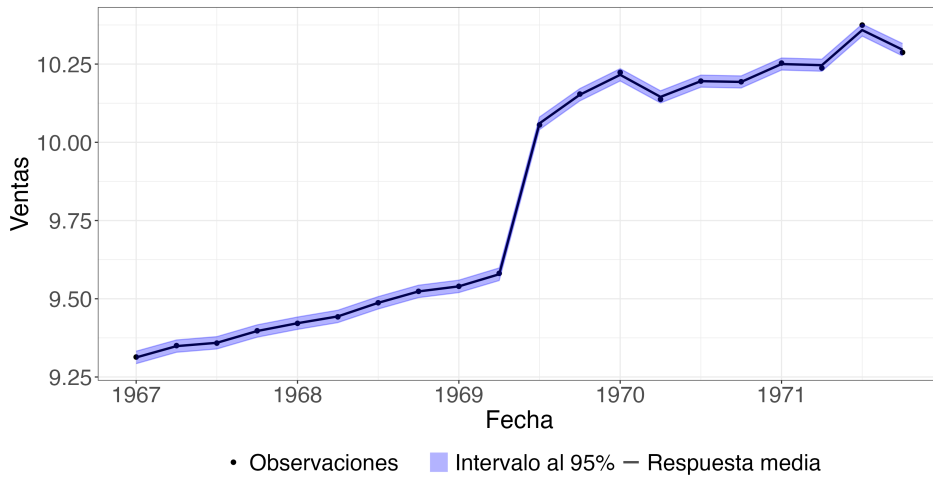


Figura 2.9: Respuesta media suavizada  $(\mu_{39-k} \mid D_{39})$  con intervención.

En realidad los procedimientos para obtener las distribuciones filtradas cuando se interviene el modelo y cuando no hay intervención son muy parecidos. La única diferencia es que cuando hay una intervención es necesario calcular una  $\mathbf{G}_t^*$  para que la ecuación de evolución coincida con la intervención. Se utiliza  $\mathbf{G}_t^*$  en lugar de  $\mathbf{G}_t$  y se procede con el método usual.

## 2.9. Modelo de espacio de estados

El modelo lineal dinámico es un modelo bayesiano, todos los elementos del modelo de los que se tenga incertidumbre se miden mediante probabilidades. Estas probabilidades son subjetivas, representan los conocimientos y creencias del modelador. El modelo lineal dinámico tiene su equivalente frecuentista conocido como modelo de espacio de estados. Este modelo se puede definir de la misma manera que el modelo lineal dinámico en 2.3.1. Para las actualizaciones del modelo de espacio de estados se utiliza el filtro de Kalman. Igual que en el teorema 2.4.1, se inicia con que

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

Una vez que se conoce el valor de  $Y_t$ , se definen  $e_t = Y_t - \mathbf{F}_t^\top \mathbf{a}_t$ ,  $Q_t = \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t + V_t$ , y  $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t Q_t^{-1}$ . Entonces, se obtienen  $\mathbf{a}_{t+1}$  y  $\mathbf{R}_{t+1}$  de la siguiente manera:

$$\mathbf{a}_{t+1} = \mathbf{G}_{t+1}[\mathbf{a}_t + \mathbf{A}_t e_t], \quad \mathbf{R}_{t+1} = \mathbf{G}_{t+1}[\mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{A}_t^\top] \mathbf{G}_{t+1}^\top + \mathbf{W}_{t+1}.$$

Esto es equivalente a los resultados presentados en el teorema 2.4.1, donde se obtuvo que  $\mathbf{a}_{t+1} = \mathbf{G}_{t+1} \mathbf{m}_t$ ,  $\mathbf{R}_{t+1} = \mathbf{G}_{t+1} \mathbf{C}_t \mathbf{G}_{t+1}^\top \mathbf{W}_{t+1}$ . En el modelo de espacio de estados no se hace la separación entre las distribuciones previa y posterior para actualizar el vector de estados, en este caso son parte del mismo paso. Las actualizaciones cuando se conocen todos los componentes del modelo son exactamente iguales. Sin embargo, el enfoque cambia cuando hay elementos desconocidos. Por ejemplo, cuando la varianza de observación es desconocida, en el modelo lineal dinámico se define una distribución previa  $(\phi \mid D_0)$ , con  $\phi = V^{-1}$ , que se actualiza cada vez que se tiene más información, como se muestra en 2.6.2. Sin embargo, en el modelo de espacio de estados los elementos desconocidos se estiman por máxima verosimilitud, posteriormente las cantidades desconocidas son sustituidas por sus estimaciones.

En los pronósticos los resultados son iguales, pero se tienen diferencias conceptuales importantes. El pronóstico a un paso en el modelo lineal dinámico es una variable aleatoria  $(Y_t \mid D_{t-1}) \sim \mathbf{N}[f_t, Q_t]$ . Por otro lado, en el modelo de espacio de estados el pronóstico no es una variable



aleatoria, sino que es el valor  $\hat{Y}_t$  que minimiza el error cuadrático medio  $E[(\hat{Y}_t - Y_t)(\hat{Y}_t - Y_t)^\top \mid D_{t-1}]$ . Como  $Y_t$  es una variable aleatoria con media  $f_t$ , el valor que minimiza el error cuadrático medio es  $\hat{Y}_t = f_t$ , por lo tanto, este es el pronóstico. Entonces, el error cuadrático medio mínimo es  $E[(\hat{Y}_t - Y_t)(\hat{Y}_t - Y_t)^\top] = V[Y_t] = Q_t$ . Por lo tanto, aunque los resultados son idénticos, en el modelo lineal dinámico se habla de distribuciones de pronóstico, mientras que en el modelo de espacio de estados el pronóstico es un valor [13].



## Capítulo 3

# Pronóstico de la demanda agregada de efectivo

En esta sección se aplica la teoría desarrollada anteriormente sobre los modelos lineales dinámicos con el objetivo de pronosticar la demanda agregada de efectivo nominal (M0) en México. Se presentan los resultados de dos modelos lineales dinámicos: uno no intervenido y otro intervenido al inicio de la pandemia. Además, se comparan estos dos modelos con el *vector error correction model* (VECM), el cual es comúnmente utilizado en este contexto. Para más información sobre los VECM, se presenta teoría en el apéndice B.1 y B.2. El código utilizado para ajustar los modelos, obtener sus pronósticos y medir sus desempeños se puede encontrar aquí:

<https://github.com/JP-LopezEsc/Tesis>

### 3.1. Modelo lineal dinámico

En esta sección se ajusta un modelo lineal dinámico sin ninguna intervención. Por facilidad, a partir de ahora se referirá al modelo lineal dinámico como DLM por sus siglas en inglés, *dynamic linear model*. Como se mencionó en la sección 1.3, para modelar la demanda de efectivo se suele usar una variable de escala, como el PIB, y una variable de costo de oportunidad, como las tasas de interés o la inflación. Lo más común es modelar los agregados monetarios reales como función del PIB real y una

tasa de interés, entre otras variables. En el DLM obtenido en este trabajo se modeló el agregado M0 nominal con el PIB real, la inflación (INPC) y factores estacionales trimestrales debido a que la demanda de efectivo incrementa en el último trimestre de todos los años. No se empleó la tasa de interés porque al intentar agregar la tasa de interés interbancaria (TIIE), empeoraba de gran manera el desempeño del modelo. Aunque esta función de demanda de dinero no es frecuente, existen precedentes de modelos parecidos. Hay casos de países, como Nueva Zelanda y Reino Unido, que modelaron agregados monetarios nominales con el PIB real como variable de escala. También hay ejemplos de países que no usaron tasa de interés en sus modelos como variable de oportunidad, pero sí usaron la inflación, entre ellos Argentina y Bolivia [14].

Se considera el modelo presentado en la sección 2.6, donde la varianza de observación es una constante desconocida. El problema es que se asume que los valores de  $\mathbf{G}_t$  y  $\mathbf{W}_t$  son conocidos, pero en realidad no es así. Los valores de  $\mathbf{G}_t$ , y  $\mathbf{W}_t$  se estimarán por máxima verosimilitud utilizando el paquete MARSS de R. Para facilitar la estimación, se asume que  $\mathbf{G}_t$  y  $\mathbf{W}_t$  son constantes, entonces se debe estimar  $\mathbf{G}$  y  $\mathbf{W}$ . Se cuenta con datos trimestrales del agregado M0 desde 2001 T1 hasta 2022 T4. Se utilizan los datos de 2001 T1 a 2011 T4 para estimar estos valores. Además, se utilizará la distribución posterior de los parámetros de 2011 T4 como la información inicial  $(\boldsymbol{\theta}_0 | D_0) \sim \mathbf{T}_{n_0}[\mathbf{m}_0, \mathbf{C}_0]$  para los periodos de 2012 T1 en adelante. Es decir,  $\mathbf{m}_{2011T4} = \mathbf{m}_0$  y  $\mathbf{C}_{2011T4} = \mathbf{C}_0$ . Entonces  $t = 0$  es 2011 T4,  $t = 1$  es 2012 T1, ...,  $t = 44$  es 2022 T4. Debido a que en este modelo se asume que  $V$  es desconocida, se necesita una estimación inicial de  $V$ ,  $S_0$ , que se actualizará en todos los periodos a partir de 2012 T1. La  $S_0$  obtenida es la estimación de  $V$  calculada con los datos de 2001 T1 a 2011 T4. Solo se usan datos de 2001 T1 a 2011 T4 para estimar todos los valores desconocidos porque las medidas de desempeño de los pronósticos se calcularán empleando únicamente los resultados del DLM de 2012 T1 a 2022 T4. Es importante determinar si el DLM funciona bien para pronosticar valores que no fueron empleados en la estimación de los valores desconocidos. En la figura 3.1 se muestran en rojo los datos empleados en la estimación de los valores desconocidos, y en azul los datos utilizados en el DLM para medir el desempeño de los pronósticos. Es evidente que el periodo de estimación y

el del DLM son muy diferentes. Por esto es importante medir el desempeño de los pronósticos en periodos que no fueron empleados en la estimación.

El paquete **MARSS** asume que la varianza de observación es conocida, incluso si en realidad es desconocida y se tuvo que estimar. Además, no se pueden hacer intervenciones al modelo. Para enfrentar esas limitaciones, se programó una función capaz de aplicar el modelo del teorema 2.6.2 y las intervenciones mostradas en la sección 2.8.

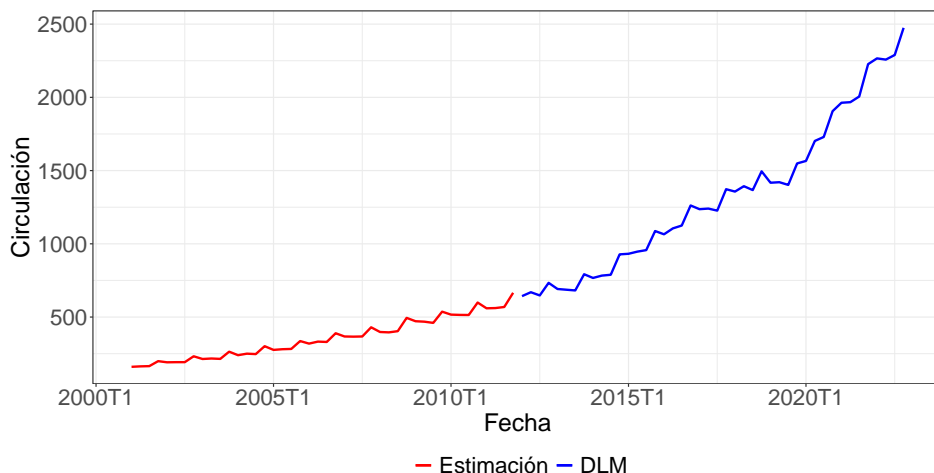


Figura 3.1: Datos empleados para estimación y DLM.

### 3.1.1. Datos

El objetivo es modelar la demanda de efectivo a través del PIB real y la inflación. Para la demanda de efectivo se utilizó el agregado monetario M0 nominal calculado por el Banco de México. El modelo es trimestral, entonces se toma el agregado M0 del último mes de cada trimestre. Los datos se pueden encontrar en <https://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?sector=3&accion=consultarCuadro&idCuadro=CF807&locale=es>, dentro del agregado M1 se tomó la serie Billetes y monedas en poder del público. Para simplificar los resultados, la circulación de efectivo se muestra en miles de millones de pesos.

El PIB es calculado por el INEGI. Se empleó el PIB trimestral a valores constantes, disponibles en <https://www.inegi.org.mx/temas/pib/#Tabulados>.

La inflación trimestral se calculó utilizando datos de la inflación mensual del INEGI, basada en el Índice Nacional de Precios al Consumidor (INPC). Estos datos se encuentran en <https://www.inegi.org.mx/app/indicadores/?tm=0&t=10000500#D10000500#D628201#D628201>, dentro de Indicadores económicos de coyuntura, en la sección de Índices de precios.

En el modelo se utilizó el logaritmo natural de la demanda de efectivo y del PIB, mientras que la inflación se dejó en la escala original. Entonces,

$$Y_t = \ln \text{Efectivo}_t, \quad \mathbf{F}_t^T = [1, \ln \text{PIB}_t, \text{Inflación}_t, 1, 0, 0, 0].$$

En  $\mathbf{F}_t^T$  el primer 1 es para el intercepto. Los últimos cuatro valores, 1, 0, 0, 0, son para la estacionalidad trimestral, siempre en ese orden, la posición del 1 no cambia según el trimestre del año. Posteriormente se explicará el porqué.

### 3.1.2. Estimación de valores desconocidos

La estimación de  $\mathbf{G}$  y  $\mathbf{W}$  se hace por máxima verosimilitud empleando el paquete de MARSS de R. Documentación sobre este paquete se puede encontrar aquí [15]. La notación es diferente, pero es útil para realizar la estimación. Un ejemplo muy claro en donde se emplea MARSS para estimar un DLM se explica en [16, pp. 255-277], además se contrasta la notación de MARSS con la misma notación empleada en esta tesis. Es importante recalcar que estas estimaciones se hacen empleando únicamente datos de 2001 T1 a 2011 T4.

Las matrices  $\mathbf{G}$  y  $\mathbf{W}$  son de dimensión  $7 \times 7$ , debido a que se cuenta con un intercepto, el PIB, la inflación y cuatro trimestres como variables explicativas. Para no estimar 49 entradas en cada matriz, la mayoría de las entradas serán 0 o 1, así solo se tendrán que estimar pocos valores. En el

caso de  $\mathbf{G}$  se tiene la siguiente matriz:

$$\mathbf{G} = \begin{bmatrix} G_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & G_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & G_{3,3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

De esta forma, como  $\mathbf{F}_t^\top = [1, \ln \text{PIB}_t, \text{Inflación}_t, 1, 0, 0, 0]$ ,  $G_{1,1}$  modela la dinámica del intercepto,  $G_{2,2}$  la del parámetro del PIB y  $G_{3,3}$  el parámetro de la inflación. Los 1 a partir de la cuarta fila se definieron de esa forma para que los factores estacionales se reacomoden. Por ejemplo, si el periodo  $r$  se encuentra en el primer trimestre, el intercepto es  $\alpha_r$ , el parámetro del PIB es  $\beta_r$ , el parámetro de la inflación es  $\gamma_r$  y el parámetro del  $i$ -ésimo trimestre del año es  $Ti$ , entonces se tiene

$$\boldsymbol{\theta}_{r+1} = \mathbf{G}\boldsymbol{\theta}_r + \boldsymbol{\omega}_{r+1} = \mathbf{G} \begin{bmatrix} \alpha_r \\ \beta_r \\ \gamma_r \\ T1_r \\ T2_r \\ T3_r \\ T4_r \end{bmatrix} + \boldsymbol{\omega}_{r+1} = \begin{bmatrix} G_{1,1}\alpha_r \\ G_{2,2}\beta_r \\ G_{3,3}\gamma_r \\ T2_r \\ T3_r \\ T4_r \\ T1_r \end{bmatrix} + \boldsymbol{\omega}_{r+1} = \begin{bmatrix} \alpha_{r+1} \\ \beta_{r+1} \\ \gamma_{r+1} \\ T2_{r+1} \\ T3_{r+1} \\ T4_{r+1} \\ T1_{r+1} \end{bmatrix}.$$

Aquí se nota claramente que  $G_{1,1}$ ,  $G_{2,2}$  y  $G_{3,3}$  sirven para modelar la evolución del intercepto, del parámetro del PIB y del parámetro de la inflación respectivamente. Como en el periodo  $r$  se encuentra en el primer trimestre del año, entonces T1 es el primer factor trimestral en el vector de estados  $\boldsymbol{\theta}_r$ . El periodo  $r + 1$  está en el segundo trimestre del año, la  $\mathbf{G}$  está definida para que los factores trimestrales se reacomoden para que ahora el primero sea T2. De esta forma el trimestre en turno siempre es el primer factor trimestral en el vector de estados. Por esta razón,  $\mathbf{F}_t^\top = [1, \ln \text{PIB}_t, \text{Inflación}_t, 1, 0, 0, 0]$  para toda  $t$ . Los últimos cuatro valores, que están asociados a los factores trimestrales, siempre aparecen en el orden 1,0,0,0, ya que el primer factor estacional es el del trimestre en

turno. Entonces,

$$\begin{aligned}\ln \text{Efectivo}_r &= \mathbf{F}_r^\top \boldsymbol{\theta}_r + \nu_r = \alpha_r + \beta_r \ln \text{PIB}_r + \gamma_r \text{Inflación}_r + T1_r + \nu_r, \\ \ln \text{Efectivo}_{r+1} &= \alpha_{r+1} + \beta_{r+1} \ln \text{PIB}_{r+1} + \gamma_{r+1} \text{Inflación}_{r+1} + T2_{r+1} + \nu_{r+1}.\end{aligned}$$

Con esas  $\mathbf{G}$  y  $\mathbf{F}_t$  se puede modelar la estacionalidad. Solo es cuestión de estimar  $G_{1,1}$ ,  $G_{2,2}$  y  $G_{3,3}$  utilizando MARSS. Por otro lado para  $\mathbf{W}$  se define la matriz

$$\mathbf{W} = \begin{bmatrix} W_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & W_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & W_{3,3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & W_{4,4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Es importante recordar que  $\mathbf{W}$  es la matriz de covarianza del vector aleatorio de errores  $\boldsymbol{\omega}_t \sim \mathbf{N}[\mathbf{0}, \mathbf{W}_t]$ . Entonces para cualquier  $t$  se tendrá que  $\boldsymbol{\omega}_t = [\omega_1, \omega_2, \omega_3, \omega_4, 0, 0, 0]^\top$ . Así solo se añade ruido a los coeficientes del intercepto, PIB, inflación y trimestre en turno, pero no a los trimestres no observados. Cuando se actualizan los parámetros al tener nuevas observaciones, todos los parámetros cambian, incluso los de los trimestres que no están en turno. Sin embargo, al calcular las distribuciones filtradas de los parámetros pasados, solo cambia el parámetro del trimestre en turno de cada periodo, los demás trimestres permanecen iguales. Entonces, solo es necesario estimar  $W_{1,1}$ ,  $W_{2,2}$ ,  $W_{3,3}$  y  $W_{4,4}$ .

Para iniciar el modelo cuando se asume que  $V$  constante y desconocida se necesita una  $S_0$ . Para escogerla, se estimó  $V$  con MARSS y esa fue la  $S_0$  utilizada. También se necesita una distribución inicial. Para obtenerla, se ajustó el DLM de 2001 T1 a 2011 T4 utilizando las estimaciones de  $\mathbf{G}$ ,  $\mathbf{W}$  y  $V$  mencionadas anteriormente. La distribución posterior en 2011 T4 será la distribución inicial para el modelo de 2012 T1 a 2022 T4. Además, entre 2001 T1 y 2011 T4 hay 44 observaciones, por lo tanto, la distribución inicial tendrá 44 grados de libertad. Entonces,  $(\boldsymbol{\theta}_0 | D_0) = (\boldsymbol{\theta}_{2011T4} | D_{2011T4}) \sim \mathbf{T}_{44}[\mathbf{m}_{2011T4}, \mathbf{C}_{2011T4}]$ . Con esto ya se puede iniciar el DLM.



### 3.1.3. Actualización

Una vez estimados los valores desconocidos empleando los datos de 2001 T1 a 2011 T4 y definida la distribución inicial, se aplica el modelo de la sección 2.6 con los datos de 2012 T1 a 2022 T4. La evolución de los parámetros se muestra en la figura 3.2.

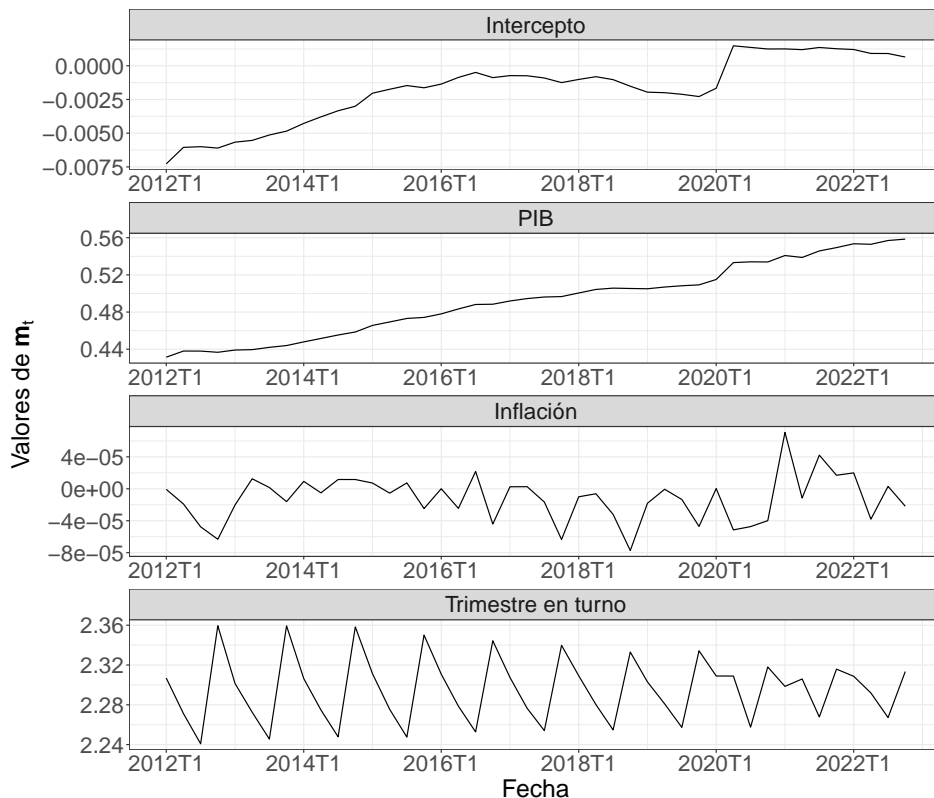


Figura 3.2: Parámetro locacional de  $(\theta_t | D_t)$ ,  $m_t$ .

El intercepto se acerca a 0 a lo largo del tiempo, tuvo un incremento fuerte en 2020 T2 para corregir el aumento inesperado en la demanda de efectivo provocado por la pandemia. El parámetro del PIB ha aumentado en todo el periodo y es positivo, como indica la literatura. En 2020 T2 el parámetro del PIB tuvo un incremento importante para ajustar el

incremento en la demanda de efectivo, pero esto no es deseable debido a que se asocia el aumento del efectivo exclusivamente a la pandemia, por lo que el parámetro del PIB no debería subir tanto. La inflación toma valores muy pequeños y ha fluctuado alrededor del 0. Como se mencionó en 1.3, se espera que la relación entre demanda de efectivo e inflación sea negativa, pero en ocasiones puede ser positiva. Aunque en este caso sus valores son extremadamente pequeños y no parece ser importante en el modelo, normalmente se utiliza una variable de costo de oportunidad, por eso se decidió dejarla. Además, en caso de que la inflación tome relevancia en este problema, puede ser útil mantenerla para intervenir el modelo. En cuanto a la estacionalidad, el modelo detectó que la demanda de efectivo se incrementa en el cuarto trimestre del año. Sin embargo, el incremento en el cuarto trimestre se redujo a lo largo del tiempo. En la figura 3.1 se puede observar que la estacionalidad en los últimos años está distorsionada en comparación con años anteriores, antes se tenía un pico en los cuartos trimestres del año y posteriormente disminuía. En los últimos años el mayor incremento se da en el cuarto trimestre, pero después ya no disminuye como antes, y esto se ve reflejado en la evolución del parámetro del trimestre en turno.

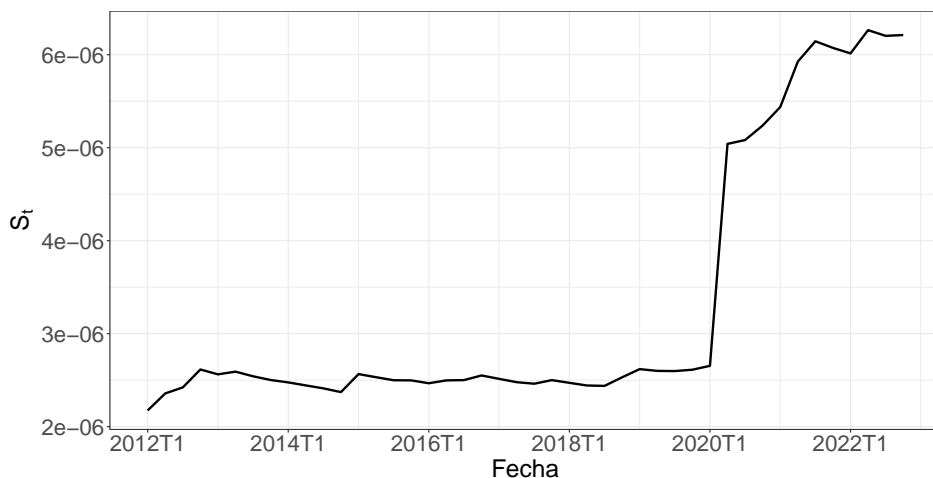


Figura 3.3: Evolución de  $S_t$ .

Otro punto interesante es analizar el cambio de la estimación de la varianza de observación, es decir,  $S_t$ . En la figura 3.3 se puede ver que  $S_t$  no tenía cambios bruscos hasta antes de la pandemia. Sin embargo, al inicio de la pandemia  $S_t$  se duplicó en un periodo, debido a que el modelo no fue capaz de pronosticar adecuadamente el incremento en la circulación de efectivo provocado por la incertidumbre asociada a la pandemia. Como la varianza que se está estimando teóricamente es constante, la estimación no debería cambiar tanto. Esto refuerza la necesidad de realizar intervenciones en el modelo para incluir situaciones externas.

### 3.1.4. Pronósticos

La realización de los pronósticos presenta una dificultad. Dada la información  $D_t$ , el modelo supone que  $\mathbf{F}_{t+k}$  es conocida. Sin embargo, en la práctica no se tiene conocimiento del PIB e inflación en el futuro. Una posibilidad para medir el desempeño de pronósticos pasados es emplear los valores históricos que actualmente ya se conocen como  $\mathbf{F}_{t+k}$ . Por ejemplo, si es de interés obtener pronósticos a 2 pasos dada la información hasta 2016 T1, podrían usarse los valores del PIB y de la inflación de 2016 T2 y 2016 T3. El problema es que esto haría injusta la comparación de modelos debido a que el VECM es un modelo multivariado que hace sus propios pronósticos de todas las variables. Además, usar los valores históricos puede reflejar un poder predictivo que el modelo en realidad no tiene, pues no se están condicionando los pronósticos hasta  $D_t$ , sino que se está utilizando información futura. Para enfrentar este problema, se emplearon los resultados de las encuestas sobre las expectativas de los especialistas en economía del sector privado recopiladas por Banxico. En estas encuestas se tiene información de pronósticos que instituciones privadas realizan sobre el PIB y la inflación en todos los meses. Entonces, volviendo al ejemplo de calcular pronósticos a 2 pasos dada la información de 2016 T1, se emplea la media de las expectativas que se tenían en 2016 T1 sobre la inflación y el PIB en 2016 T2 y 2016 T3. Así, los pronósticos del modelo realmente están condicionados únicamente hasta la información presente  $D_t$  en cada periodo. Los pronósticos externos son una forma de información disponibles en todos los periodos, por lo tanto es válido incorporarlos al modelo en  $\mathbf{F}_{t+k}$ .

En general, se usó la media de las expectativas del PIB y de la inflación en cada periodo  $t$  como  $\mathbf{F}_{t+k}$ . En las encuestas se reportan las expectativas de hasta 8 periodos futuros. En caso de que se reporten menos periodos futuros, se utilizó la última expectativa de crecimiento del PIB y de la inflación reportadas para completar los pronósticos. Para actualizar el DLM se utilizaron los valores reales del PIB y de la inflación una vez que se conocen. Lo de sus expectativas únicamente es para poder realizar los pronósticos. Las expectativas se pueden consultar en <https://www.banxico.org.mx/publicaciones-y-prensa/encuestas-sobre-las-expectativas-de-los-especialis/encuestas-expectativas-del-se.html>.

Se obtuvieron las distribuciones de pronóstico  $(Y_{t+k} | D_t)$  con  $k = 1, \dots, 8$ , para las observaciones a partir de 2012 T1. Los pronósticos puntuales son los parámetros locacionales de las distribuciones  $(Y_{t+k} | D_t)$ ,  $f_t(k)$ . Es importante recordar que  $Y_t = \ln \text{Efectivo}_t$ , y lo que se desea es calcular los pronósticos de  $\text{Efectivo}_{t+k}$ .

Sea  $Z_t = \exp(Y_t) = \text{Efectivo}_t$ , entonces, si se tienen  $b_{t+k}$ ,  $c_{t+k}$  tales que  $F_{Y_{t+k}}(b_{t+k}) = 0.025$  y  $F_{Y_{t+k}}(c_{t+k}) = 0.975$ , como la función exponencial es monótona creciente,

$$\begin{aligned} 0.95 &= P(b_{t+k} \leq Y_{t+k} \leq c_{t+k}) = P(Y_{t+k} \leq c_{t+k}) - P(Y_{t+k} \leq b_{t+k}) \\ &= P(Z_{t+k} \leq \exp(c_{t+k})) - P(Z_{t+k} \leq \exp(b_{t+k})) \\ &= P(\exp(b_{t+k}) \leq Z_{t+k} \leq \exp(c_{t+k})). \end{aligned}$$

Por lo tanto,  $[\exp(b_{t+k}), \exp(c_{t+k})]$  es un intervalo del 95 % para el pronóstico de  $\text{Efectivo}_{t+k}$ .

En cuanto al pronóstico puntual  $f_t(k)$ , como es igual al parámetro locacional de una  $t$  de Student, entonces  $f_t(k)$  también es la mediana,  $F_{Y_{t+k}}(f_t(k)) = 0.5$ . Entonces,

$$0.5 = P(Y_{t+k} \leq f_t(k)) = P(Z_{t+k} \leq \exp(f_t(k)))$$

Por lo tanto, el pronóstico de  $\text{Efectivo}_{t+k}$  es equivalente a la mediana de la distribución de  $Z_{t+k}$ , que es igual a  $\exp(f_t(k))$ .



Figura 3.4: Pronósticos a un paso,  $(Z_{t+1} | D_t)$ .

En la figura 3.4 se muestran los pronósticos de  $(Z_{t+1} | D_t)$  y sus intervalos al 95 %. En general los pronósticos a un paso son acertados, con los valores reales dentro de los intervalos de pronóstico. El error más grande ocurrió en 2020 T2 debido a que el modelo no pudo anticipar el aumento en la demanda de dinero provocado por la pandemia. En el apéndice C.1 se muestran gráficas para los pronósticos desde  $k = 2$  hasta  $k = 8$  pasos. Además, en la sección 3.4 se reportan las medidas de desempeño de los pronósticos del DLM, y se comparan contra el DLM intervenido y el VECM.

### 3.1.5. Distribuciones filtradas

En la sección 2.7 se mostró que es posible obtener distribuciones filtradas para los parámetros,  $(\theta_{t-k} | D_t)$ . En este modelo la última observación es 2022 T4, que es equivalente a  $t = 44$ . Entonces las distribuciones filtradas

están condicionadas en  $D_{44}$ . En la figura 3.5 se pueden ver los parámetros locacionales de las distribuciones filtradas de los parámetros.

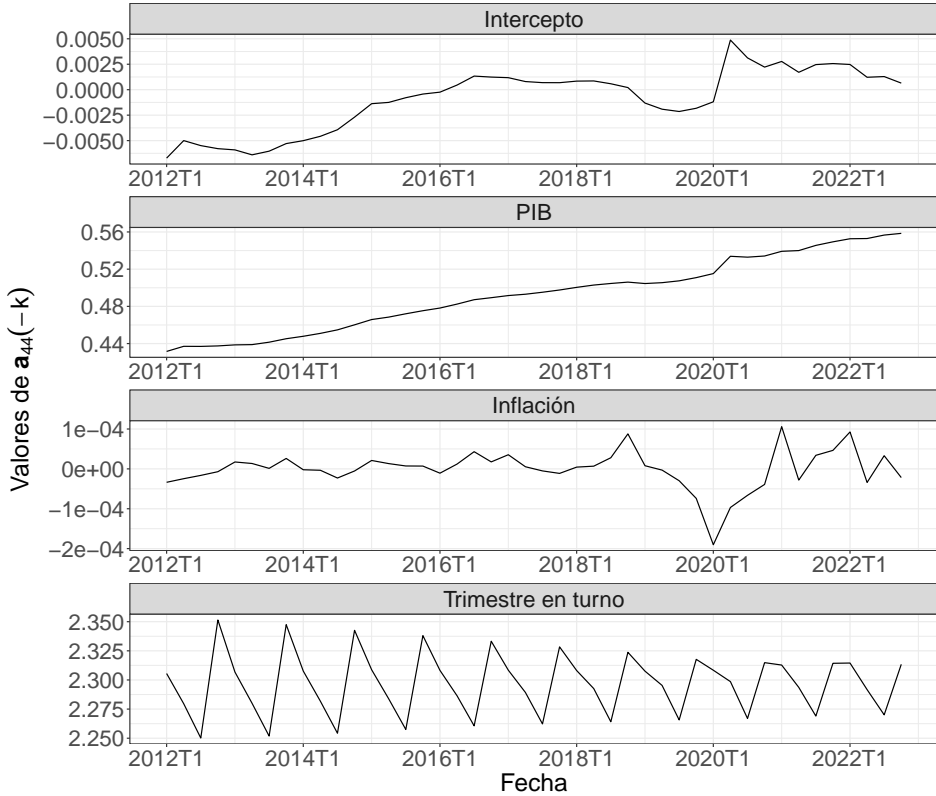


Figura 3.5: Parámetro locacional de  $(\theta_{44-k} | D_{44})$ ,  $a_{44}(-k)$ .

Las gráficas 3.5 y 3.2 son bastante parecidas. Una de las principales diferencias es que en 3.5 el incremento en el intercepto en 2020 T2 es mayor. Considerando información hasta 2020 T2 el intercepto en ese periodo era 0.0015, pero con información hasta 2022 T4 el intercepto en 2020 T2 es 0.0049. Ahora que se conoce el efecto que toda la pandemia tuvo en la circulación de efectivo, el modelo triplicó el valor del intercepto en 2020 T2 en las distribuciones filtradas. Esto indica que el modelo considera que

el intercepto en 2020 T2 no fue lo suficientemente grande para modelar correctamente el aumento en la demanda de efectivo en los periodos posteriores. Otra diferencia es que en los últimos años en 3.2 el cuarto trimestre era el de mayor circulación, pero en 3.5 el primer y el cuarto trimestre están en el mismo nivel.

## 3.2. Modelo lineal dinámico intervenido

Este modelo es muy parecido al presentado anteriormente. Las estimaciones de  $\mathbf{W}$ ,  $\mathbf{G}$ ,  $S_0$  y  $(\boldsymbol{\theta}_0 \mid D_0)$  son exactamente iguales. El modelo es el mismo hasta 2020 T2, que es cuando se realiza la intervención. A partir de ahí, los modelos son diferentes, pero antes son idénticos.

### 3.2.1. Intervención

Anteriormente se mostró que el DLM tuvo un error grande en 2020 T2, debido a que el modelo no tuvo forma de pronosticar el gran aumento en la demanda de efectivo provocado por la pandemia. La pandemia es información externa al modelo, y la única forma de realizar un pronóstico acertado es incluir esta información externa. En 2020 T1 ya se sabía que la pandemia estaba generando una gran incertidumbre, incrementando la demanda de efectivo. En el informe del primer trimestre de 2020 del Banco de México se menciona que durante marzo 2020 hubo una mayor demanda de efectivo en los hogares por precaución ante la pandemia [17, p. 47]. El periodo 2020 T2 corresponde a  $t = 34$ , entonces se intervendrá  $(\boldsymbol{\theta}_{34} \mid D_{33}) \sim \mathbf{T}_{77}[\mathbf{a}_{34}, \mathbf{R}_{34}]$  como se muestra en la sección 2.8. En el DLM se tiene

$$\mathbf{a}_{34} = \begin{bmatrix} -0.0015 \\ 0.5182 \\ 7.8\text{e}-08 \\ 2.2844 \\ 2.2593 \\ 2.3307 \\ 2.3090 \end{bmatrix}, \quad \mathbf{R}_{34} = \begin{bmatrix} 7.58\text{e}-05 & \dots \\ \vdots & \ddots \end{bmatrix}.$$

Es necesario proponer una distribución  $(\boldsymbol{\theta}_{34} \mid D_{33}, I_{34}) \sim \mathbf{T}_{77}[\mathbf{a}_{34}^*, \mathbf{R}_{34}^*]$

que represente las creencias que se tienen en 2020 T1 sobre el incremento en la demanda de efectivo en 2020 T2 debido a la pandemia. Una forma sencilla de sugerir una intervención es comparar la diferencia entre 2018 T4 y 2019 T1 contra la diferencia entre 2019 T4 y 2020 T1 para tener una idea de qué tan anormal es 2020 T1. El logaritmo de la circulación de efectivo en 2018 T4 es 0.0535 unidades mayor que en 2019 T1. Por otro lado, 2019 T4 es 0.0109 unidades menor que 2020 T1. Esto quiere decir que el primer trimestre de 2020 está 0.0644 unidades por encima de lo esperado según lo observado en 2019. Tiene sentido suponer que en 2020 T2 pasaría algo parecido. Entonces al intercepto de  $\mathbf{a}_{34}$  se le sumará 0.0644. Obviamente hay mucha incertidumbre por la pandemia y no se sabe bien qué tan grande será el incremento. Por esta razón, la varianza asociada al intercepto se multiplicará por 10. El resto de la distribución se dejará igual. Entonces para  $(\boldsymbol{\theta}_{34} \mid D_{33}, I_{34}) \sim \mathbf{T}_{77}[\mathbf{a}_{34}^*, \mathbf{R}_{34}^*]$  se tiene que

$$\mathbf{a}_{34}^* = \begin{bmatrix} 0.0629 \\ 0.5182 \\ 7.8\text{e}-08 \\ 2.2844 \\ 2.2593 \\ 2.3307 \\ 2.3090 \end{bmatrix}, \quad \mathbf{R}_{34}^* = \begin{bmatrix} 7.58\text{e}-04 & \dots \\ \vdots & \ddots \end{bmatrix}.$$

Con esto la distribución del intercepto en  $t = 34$  es  $\mathbf{T}_{77}[0.0629, 7.58\text{e}-04]$ , mientras que sin intervención la distribución era  $\mathbf{T}_{77}[-0.0015, 7.58\text{e}-05]$ . Si se calculan intervalos al 95 % para el intercepto, el intervalo sin intervención es  $[-0.0189, 0.0158]$ . Por último el intervalo para el intercepto intervenido es  $[0.0080, 0.1177]$ . En el intervalo se ve reflejado el aumento en la incertidumbre que se desea modelar, pues es considerablemente más ancho que el no intervenido.

En la figura 3.6 se ve claramente el efecto de la intervención en el intercepto en  $\mathbf{m}_t$ . El valor del intercepto en  $\mathbf{a}_{34}^*$  es 0.0629, pero una vez que se observa  $Y_{34}$  y se actualizan los parámetros, el valor del intercepto sube a 0.1307 en  $\mathbf{m}_{34}$ . Es decir, incluso con la intervención el incremento en la demanda de efectivo en 2020 T2 fue mayor a lo que se esperaba, por eso el intercepto se incrementó aún más. Contrario a lo sucedido en el modelo no



intervenido, no se aprecia ningún cambio fuerte en el parámetro del PIB en 2020 T2, que es lo esperado debido a que no se conoce ningún cambio en la relación entre PIB y efectivo generado por la pandemia.

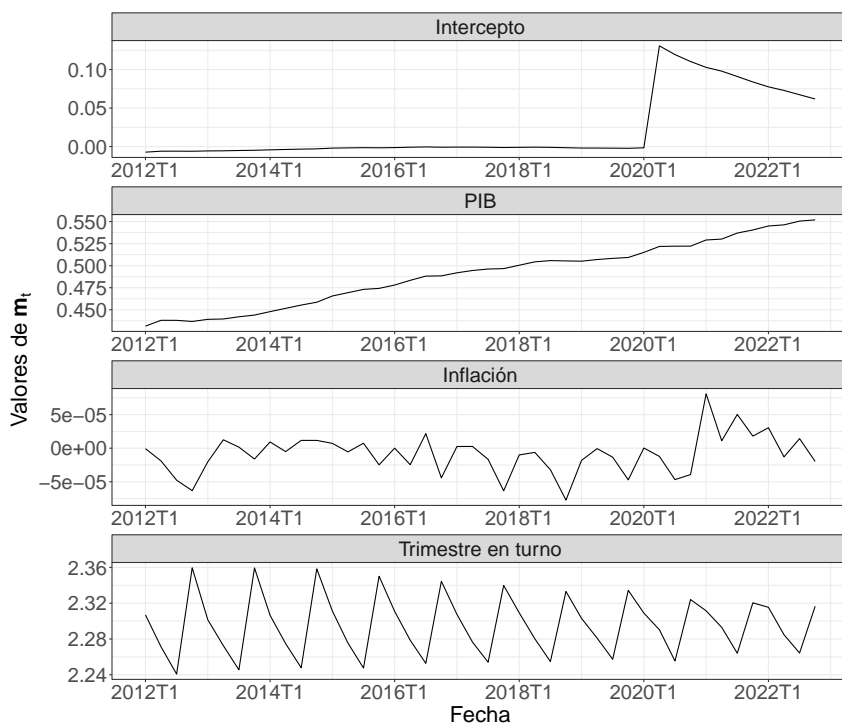


Figura 3.6: Parámetro locacional de  $(\theta_t | D_t)$ ,  $m_t$ , con intervención.

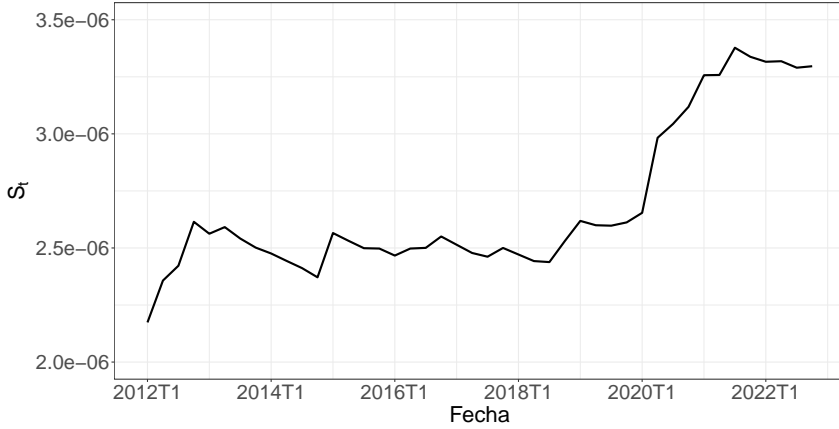


Figura 3.7: Evolución de  $S_t$  con intervención.

La evolución de la estimación de la varianza,  $S_t$ , se muestra en la figura 3.7. El incremento de  $S_t$  en 2020 T2 fue grande debido a que el DLM intervenido no captura por completo el incremento en la demanda de dinero, pero es considerablemente menor que la estimación de la varianza del modelo no intervenido, como se muestra en 3.3. Sin intervención, la estimación de la varianza es casi el doble.

### 3.2.2. Pronósticos con intervención

Para los pronósticos se procedió de la misma manera que en el DLM, la única diferencia es la intervención en 2020 T2. Solo es necesario considerar que se tuvo conocimiento del efecto de la pandemia en la demanda de efectivo hasta 2020 T1. Por ejemplo, si dada la información en 2019 T4 ( $t = 32$ ) se desea hacer un pronóstico de 2 pasos para 2020 T2 ( $t = 34$ ), no se puede intervenir porque todavía no se sabía esto. Es decir, la distribución de pronóstico ( $Y_{34} \mid D_{32}$ ) no considera la intervención. Pero si en 2020 T1 ( $t = 33$ ) se quiere calcular el pronóstico a 2 pasos para 2020 T3 ( $t = 35$ ), se puede considerar la intervención para obtener el pronóstico a un paso ( $Y_{34} \mid D_{33}, I_{34}$ ) para posteriormente calcular la distribución del pronóstico a 2 pasos ( $Y_{35} \mid D_{33}, I_{34}$ ). Aunque la intervención solo se realiza en el pronóstico de  $Y_{34}$ , también afecta el resultado de todos los pronósticos

a  $k$  pasos dado  $D_{33}$ .

En la figura 3.8 se muestran los pronósticos a un paso en la escala original. Es evidente que gracias a la intervención se tiene un pronóstico mucho más acertado, ahora el valor real de la circulación de efectivo está dentro de los intervalos del pronóstico. Sin intervención, el pronóstico puntual a un paso es 1530 y los intervalos de pronóstico al 95 % son  $[1468, 1594]$ . Por otro lado, el pronóstico puntual a un paso del modelo intervenido es 1632, con intervalo  $[1527, 1744]$ . El valor real fue 1702, está dentro del intervalo del modelo intervenido. Otro punto importante es que la longitud del intervalo del modelo intervenido es 217, mientras que el intervalo del modelo no intervenido es 126. Este era otro objetivo de la intervención, modelar el incremento en la incertidumbre en 2020 T2.

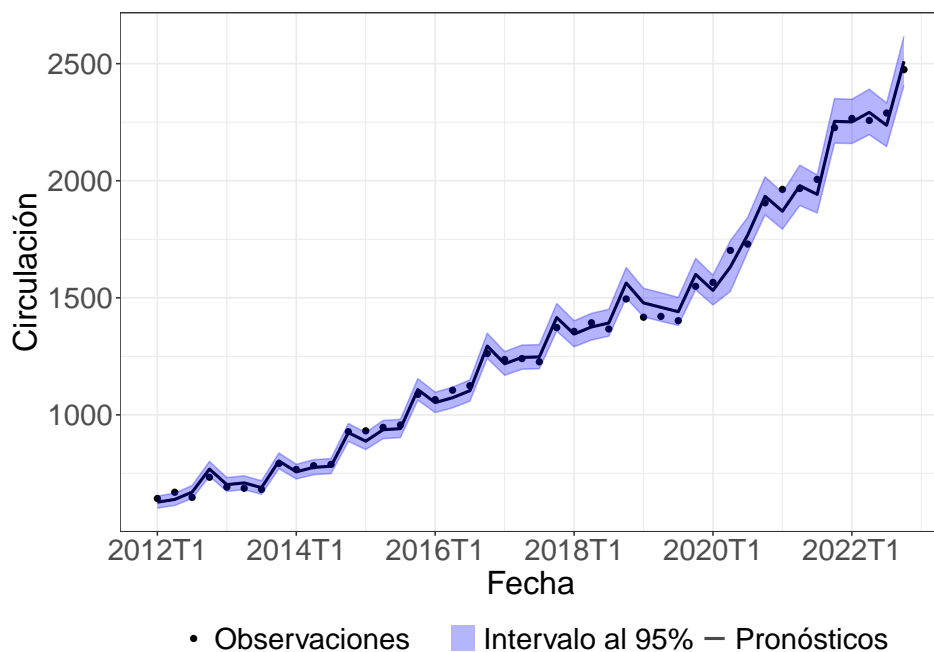


Figura 3.8: Pronósticos a un paso  $(Z_{t+1} | D_t)$ , con intervención.

En la sección 3.4 se muestran las medidas de desempeño de los pronósticos del DLM intervenido y se comparan contra el DLM y el VECM. En el apéndice C.2 se muestran gráficas de los pronósticos desde 2 pasos hasta 8 pasos.

### 3.2.3. Distribuciones filtradas con intervención

El objetivo de las intervenciones no es únicamente obtener mejores pronósticos. También son correctivas, cuando el desempeño de los pronósticos es inadecuado se utiliza información externa retroactivamente para ajustar el modelo. Hasta ahora se mostró la intervención en 2020 T2 para ejemplificar la importancia de las intervenciones en los pronósticos, pero en realidad se conoce el valor real del efectivo en 2020 T2, entonces no tiene sentido pronosticar este valor. Sin embargo, esta intervención es necesaria para corregir el modelo en observaciones pasadas, y así funcione mejor en el presente. Ya se mencionó anteriormente el efecto positivo que tuvo la intervención en  $S_t$ , ahora la varianza del modelo es menor.

Las distribuciones filtradas de los parámetros con la intervención se muestran en la figura 3.9. El intercepto en  $\mathbf{m}_{34}$  tiene un valor de 0.1307, mientras que el de  $\mathbf{a}_{44}(-10)$  es 0.1398. Es decir, ya que se conocen todos los datos no hubo un cambio grande en el intercepto de 2020 T2, solo es ligeramente más alto. Es necesario recordar que en el modelo no intervenido el intercepto de  $\mathbf{a}_{44}(-10)$  triplica el valor del intercepto en  $\mathbf{m}_{34}$ . Esto significa que si no se interviene el modelo se tendrá un intercepto más bajo de lo que se necesita no solo en 2020 T2, sino durante toda la pandemia. Además, como ya se mencionó anteriormente, el modelo no intervenido trató de ajustar el crecimiento del efectivo incrementando el parámetro del PIB, lo que no es correcto. Esto es una muestra de cómo realizar una intervención en el pasado es útil para corregir el modelo.

## 3.3. VECM

Los VECM son ampliamente utilizados para modelar la demanda de dinero. Para ajustar un VECM apropiado se realizó un procedimiento parecido al reportado por el Banco Central Europeo en [4, pp. 29-37]. Primero

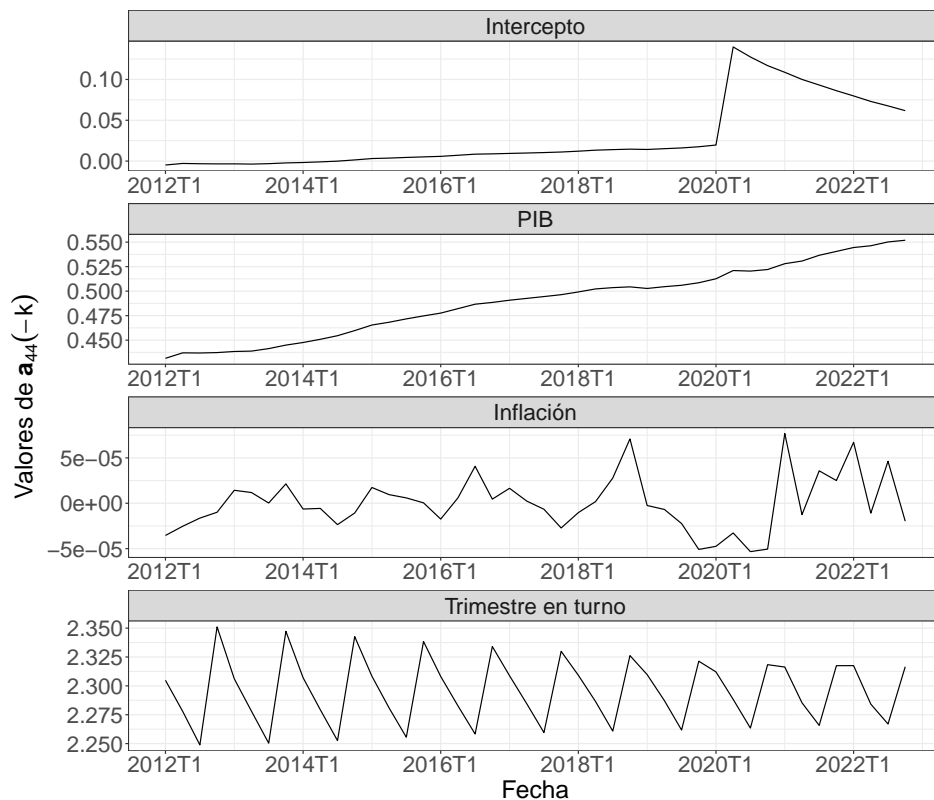


Figura 3.9: Parámetro locacional de  $(\theta_{44-k} | D_{44})$ ,  $a_{44}(-k)$ , con intervención.

se probará el orden de integración de las variables, después se escogerá el número de rezagos  $p$  y el número de relaciones de cointegración  $r$ . Se utilizaron los datos de 2001 T1 a 2011 T4 para escoger  $p$  y  $r$ . En el periodo 2012 T1 a 2022 T4, en cada trimestre se estimó un VECM con los datos disponibles en cada fecha, pero el número de rezagos y de relaciones de cointegración siempre se deja igual.

Las variables involucradas en el VECM tienen que ser integradas de orden 1. Entonces, el primer paso es verificar que las variables incluidas en

el VECM sean  $I(1)$ . Para probar el orden de integración, se aplicó la prueba aumentada Dickey-Fuller. En el cuadro 3.1 se muestran los resultados.

Variable	Rezagos	Estadístico	p-value
$\ln$ Efectivo	4	-1.589	0.737
$\Delta \ln$ Efectivo	4	-3.509	0.053
$\ln$ PIB	2	-2.695	0.299
$\Delta \ln$ PIB	2	-6.057	$\leq 0.01$
Inflación	4	-2.182	0.502
$\Delta$ Inflación	4	-3.605	0.044
$\ln$ THIE	2	-1.953	0.59
$\Delta \ln$ THIE	2	-2.233	0.484

Cuadro 3.1: Prueba aumentada Dickey-Fuller.

En el caso del logaritmo del efectivo, logaritmo del PIB y de la inflación se acepta la hipótesis nula de que hay raíces unitarias en las series en niveles. Sin embargo, al calcular sus diferencias, se rechaza la hipótesis nula con un nivel de significancia de 10 % (el p-value de la diferencia de la inflación es ligeramente mayor a 0.05). Por lo tanto, las diferencias son estacionarias, es decir, el logaritmo del efectivo, el logaritmo del PIB y la inflación son  $I(1)$ . El logaritmo de la THIE no es  $I(0)$  ni  $I(1)$ , por lo que no se empleará. En el DLM tampoco se utilizó debido a que impactaba negativamente el desempeño del modelo.

El siguiente paso es determinar el número de rezagos en el VECM. Para esto, se tomará el número de rezagos que minimiza los criterios de Aikaike, Hannan-Quinn, Schwarz y *final prediction error* de un VAR. En este caso, con 2 rezagos se minimiza el Aikaike, Hannan-Quinn y *final prediction error*, mientras que con un rezago se minimiza Schwarz. Por lo tanto, se selecciona el VAR con 2 rezagos. Debido a que el interés es obtener el modelo en forma de VECM se debe considerar que el equivalente a un VAR de 2 rezagos es un VECM de un rezago. Por lo tanto, el VECM seleccionado tendrá un rezago.

Después se debe escoger el número de relaciones de cointegración del

VECM. Para esto, se aplica el procedimiento de Johansen con eigenvalores. No se incluyeron intercepto ni tendencia en las relaciones de cointegración debido a que el modelo fue más acertado sin estos. En el cuadro 3.2 se reportan los resultados. El estadístico de prueba cuando la hipótesis nula es  $r = 0$  supera el valor crítico, por lo tanto se rechaza. Cuando  $r \leq 1$  es la hipótesis nula, el estadístico obtenido no supera el valor crítico, por lo tanto se acepta. Como se rechazó que  $r = 0$ , entonces se concluye que se tiene  $r = 1$ .

H0	Estadístico	Valor crítico 95 %
$r \leq 2$	5.64	8.18
$r \leq 1$	7.63	14.90
$r = 0$	40.82	21.07

Cuadro 3.2: Procedimiento de Johansen.

En el DLM fue importante considerar la estacionalidad, por lo que en el VECM también se agregarán los factores estacionales. Todos los VECM estimados tienen  $p = 1$ ,  $r = 1$ . Entonces los VECM estimados en todos los periodos tienen la forma

$$\Delta \mathbf{Y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}^T \mathbf{Y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{Y}_{t-1} + \mathbf{C} \mathbf{d}_t + \mathbf{u}_t,$$

con

$$\mathbf{Y}_t = \begin{bmatrix} \ln \text{Efectivo}_t \\ \ln \text{PIB}_t \\ \text{Inflación}_t \end{bmatrix}, \quad \mathbf{d}_t = \begin{bmatrix} 1 \\ Q1_t \\ Q2_t \\ Q3_t \end{bmatrix}.$$

Los factores estacionales y el intercepto se encuentran en  $\mathbf{d}_t$ , mientras que  $\boldsymbol{\beta}$  es el vector de cointegración, que se puede interpretar como el equilibrio a largo plazo,  $\beta_1 \ln \text{Efectivo}_t + \beta_2 \ln \text{PIB}_t + \beta_3 \text{Inflación}_t$ .

Una ventaja del VECM es que es un modelo multivariado que realiza sus propios pronósticos de todas las variables, mientras que el DLM empleado utiliza pronósticos externos de las variables explicativas. Esto no es un problema en este caso debido a que el PIB y la inflación son extremadamente importantes y es fácil obtener pronósticos de estas variables

realizados por instituciones fiables. En cambio, si se desean incluir variables cuyos pronósticos no suelen ser publicados por agentes externos sería necesario realizar pronósticos propios para incluirlos en el DLM, lo que implicaría un esfuerzo adicional. Por otro lado, los VECM tienen la limitación de que las variables empleadas deben ser  $I(1)$  y deben estar cointegradas para que el equilibrio de largo plazo se cumpla; esto no es necesario en los DLM debido a que son dinámicos y pueden modelar cambios abruptos en las variables, en los DLM se pueden incorporar variables no estacionarias sin diferenciar. Esto es una ventaja debido a que diferenciar las variables puede distorsionar las relaciones entre ellas y dificultar la interpretación del modelo [12, pp. 299-300].

Para medir el desempeño de los pronósticos del DLM se actualiza el modelo y se hacen los pronósticos con toda la información disponible en cada periodo. Para hacer justa la comparación, el VECM se reestimó en todos los periodos con los datos disponibles en esos periodos. Siempre se toma un rezago y una relación de integración, pero los valores  $\alpha$ ,  $\beta$ ,  $\Gamma_1$  y  $C$  se reestimaron en todos los periodos con máxima verosimilitud. En el DLM solo se estiman los valores desconocidos con máxima verosimilitud una vez, después gracias a su naturaleza dinámica solo es necesario actualizar los parámetros cuando hay nuevas observaciones. Por lo tanto, la actualización del DLM es mucho más sencilla debido a que no es necesario reestimar los parámetros en cada periodo, solo se actualizan. En la figura 3.10 se muestran los pronósticos a un paso del VECM siguiendo este procedimiento.

Es evidente que se tuvo un error grande en el pronóstico de 2020 T2, al igual que en el DLM. Visualmente es imposible determinar qué modelo tiene mayor poder predictivo. Para averiguarlo, en la siguiente sección se reportan y se comparan las medidas de precisión de los pronósticos de los tres modelos empleados.





Figura 3.10: Pronósticos a un paso VECM.

### 3.4. Comparación

En esta sección se compara el poder predictivo del DLM, DLM intervenido y VECM. Se consideran cuatro medidas de precisión de pronósticos: el error cuadrático medio (MSE por sus siglas en inglés), error absoluto medio (MAE), error de porcentaje medio absoluto (MAPE) y U de Theil. Se puede encontrar teoría sobre estas medidas en [18, pp. 42-50]. Se presentan los resultados de las medidas de precisión para los pronósticos desde un paso hasta ocho pasos, esto con el objetivo de comparar los pronósticos de los modelos en el corto y mediano plazo. Es importante mencionar que la escala en la que se modeló la demanda de dinero M0 nominal fue en miles de millones de pesos, y en esta misma escala se calcularon las medidas de precisión.

### 3.4.1. MSE

El pronóstico de  $Y_{t+k}$  dada la información  $D_t$  es  $f_t(k)$ . Si el error de pronóstico es  $e_t = Y_{t+k} - f_t(k)$  y se tienen pronósticos de  $n$  periodos futuros, entonces

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2.$$

Al elevar al cuadrado los errores más grandes adquieren más relevancia y tienen un efecto mayor en la métrica. En el cuadro 3.3 se reporta el MSE de los tres modelos, de uno a ocho pasos.

$k$	DLM	DLM int.	VECM
1	2,172.68	1,240.21	4,222.47
2	3,288.09	2,746.81	5,006.28
3	4,689.36	4,219.50	6,978.05
4	5,837.19	5,416.24	6,915.17
5	9,317.32	8,654.27	12,846.78
6	10,322.76	10,272.56	14,445.70
7	11,983.37	11,740.79	16,955.17
8	13,450.90	13,057.38	16,947.03

Cuadro 3.3: Comparación MSE de pronósticos a  $k$  pasos.

Es evidente que el peor modelo para todos los  $k$  pasos es el VECM. Por otro lado, el DLM intervenido es superior al DLM, sobre todo cuando  $k = 1$ . Esto tiene sentido porque la intervención se realizó a un paso, pero también ayuda a mejorar la precisión de pronósticos a más de un paso, aunque en menor medida.

### 3.4.2. MAE

Esta medida tiene la ventaja de estar en la misma escala que los datos, por lo que es fácil de interpretar. A diferencia del MSE, el MAE le otorga el mismo peso a todos los errores. Se promedia el valor absoluto de los errores,

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t|.$$

En el cuadro 3.4 se muestra el MAE de los tres modelos. El VECM tiene menor precisión en todos los horizontes de pronósticos que los DLM. En los pronósticos a un paso, el DLM intervenido tiene un error medio 5 mil millones de pesos menor que el DLM no intervenido. Conforme se incrementan los  $k$  pasos, esta diferencia se vuelve más pequeña. En  $k = 8$ , el DLM tuvo una precisión ligeramente mayor a la del modelo intervenido. Como se mencionó en los resultados del MSE, tiene sentido que la diferencia en la precisión entre el DLM y el DLM intervenido sea menor conforme aumenta  $k$  debido a que la intervención fue a un paso.

$k$	DLM	DLM int.	VECM
1	34.28	29.02	46.62
2	41.94	39.68	50.99
3	52.90	49.34	61.50
4	57.65	59.30	65.66
5	75.44	73.29	87.24
6	77.29	79.14	95.23
7	85.67	84.70	103.56
8	94.87	95.64	105.08

Cuadro 3.4: Comparación MAE de pronósticos a  $k$  pasos.

### 3.4.3. MAPE

Un inconveniente del MSE y el MAE es que sus magnitudes dependen de la escala de la serie que se desea pronosticar, pero es importante conocer qué tan grandes son los errores en comparación con los valores de la serie. Para esto, se utiliza el MAPE, que se calcula así:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \left( \frac{Y_{t+k} - f_t(k)}{Y_{t+k}} \right) \times 100 \right|.$$

Entonces, el MAPE indica el error relativo porcentual promedio. En el cuadro 3.5 se presenta el MAPE de los modelos. Para todos los horizontes

de pronóstico presentados el VECM tiene los peores desempeños. En el caso de los DLM, en todos los horizontes de tiempo reportados el error relativo es menor al 7 % de la demanda de efectivo. El MAPE del DLM intervenido es mejor cuando  $k$  es pequeña, pero cuando el horizonte de pronóstico es de un año o más, casi no hay diferencia entre ambos modelos. De hecho, en  $k = 4, 6, 8$  el modelo no intervenido tuvo un desempeño ligeramente mejor.

$k$	DLM	DLM int.	VECM
1	2.51 %	2.23 %	3.23 %
2	3.21 %	3.10 %	3.74 %
3	3.89 %	3.73 %	4.41 %
4	4.35 %	4.43 %	4.86 %
5	5.41 %	5.31 %	6.14 %
6	5.65 %	5.72 %	6.66 %
7	6.12 %	6.08 %	7.13 %
8	6.62 %	6.66 %	7.23 %

Cuadro 3.5: Comparación MAPE de pronósticos a  $k$  pasos.

#### 3.4.4. U de Theil

Las tres medidas anteriores son útiles para comparar los modelos, pero no indican si de verdad emplear estos modelos mejora la precisión de los pronósticos en comparación con algún otro método más sencillo. Para esto, la U de Theil compara el pronóstico de un modelo formal contra los de un modelo *naive* o ingenuo, que toma la última observación conocida como pronóstico. Esta medida calcula el cuadrado de los errores, por lo que al igual que el MSE le da mayor peso a errores grandes. En realidad la U de Theil está definida para pronósticos a un paso, se modificó para poder comparar los modelos con el modelo *naive* para pronósticos a más de un paso, pero es importante mencionar que su definición solo es para un paso. La U de Theil modificada para  $n$  pronósticos a  $k$  pasos se calculó como se muestra a continuación:

$$U = \sqrt{\frac{\sum_{t=1}^n \left( \frac{f_t(k) - Y_{t+k}}{Y_t} \right)^2}{\sum_{t=1}^n \left( \frac{Y_{t+k} - Y_t}{Y_t} \right)^2}}.$$

El numerador y el denominador son parecidos al MAPE del modelo formal y del modelo *naive* respectivamente. Entonces,  $U = 1$  significa que el modelo formal y el *naive* tienen la misma precisión de pronósticos. Si  $U < 1$  entonces el método formal es mejor que el *naive*, y entre más se acerque  $U$  a 0 mejor. Por último, si  $U > 1$  no tiene sentido usar el modelo formal por ser peor que el *naive*.

La  $U$  de Theil de los modelos se muestra en el cuadro 3.6. Todos los modelos empleados son mejores que un modelo *naive*. Obviamente, conforme aumentan los pasos futuros mejores son los modelos en comparación con el *naive*. El DLM intervenido es mejor que el resto en todos los pasos (con la excepción del empate en  $k = 7$  con el DLM).

$k$	DLM	DLM int.	VECM
1	0.48	0.40	0.63
2	0.48	0.46	0.57
3	0.43	0.41	0.49
4	0.41	0.40	0.45
5	0.39	0.38	0.44
6	0.36	0.36	0.41
7	0.33	0.33	0.38
8	0.32	0.31	0.35

Cuadro 3.6:  $U$  de Theil modificada de pronósticos a  $k$  pasos.

### 3.4.5. Intervalos de pronóstico

Es imposible que un pronóstico puntual sea exactamente igual al valor de interés una vez observado. Por esto, es necesario obtener pronósticos en

forma de intervalos, con base en cierta probabilidad. En esta tesis se calcularon los intervalos de pronóstico al 95 %. Para analizar qué tan correctos son los intervalos de cada modelo, se calculó el porcentaje de observaciones  $Efectivo_{t+k}$  que están dentro del intervalo de pronóstico de la variable aleatoria  $(Z_{t+k} | D_t)$  con probabilidad de 0.95. Los resultados se presentan en el cuadro 3.7.

k	DLM	DLM int.	VECM
1	81.82 %	86.36 %	70.45 %
2	83.72 %	86.05 %	69.77 %
3	80.95 %	83.33 %	66.67 %
4	80.49 %	82.93 %	70.73 %
5	82.50 %	85.00 %	65.00 %
6	79.49 %	82.05 %	61.54 %
7	81.58 %	81.58 %	68.42 %
8	83.78 %	86.49 %	64.86 %

Cuadro 3.7: % Observaciones dentro de intervalos de pronóstico al 95 %.

Antes que nada, al ser intervalos al 95 %, se espera que en todos los modelos el porcentaje de observaciones dentro de los intervalos sea cercana al 95 %. Claramente, el VECM está muy lejos de esto. Los DLM tienen porcentajes más altos, aunque por debajo del 95 %. Esto indica que los modelos están subestimando la varianza en sus pronósticos. Sin embargo, es importante recordar que de  $t = 1$  a  $t = 33$  el DLM y el DLM intervenido son exactamente iguales, las diferencias se encuentran de  $t = 34$  a  $t = 44$ . A pesar de que la intervención afecta únicamente los últimos 11 periodos, se puede notar su efecto en los intervalos de pronóstico, pues el DLM intervenido tiene porcentajes más altos que el DLM en el cuadro 3.7.

Se realizaron 11 pronósticos a un paso después de la intervención, pero conforme aumentan los pasos son menos los pronósticos disponibles para medir su precisión. Por ejemplo, a 8 pasos solo se puede medir la precisión de 4 pronósticos después de la intervención. Aunque se tienen muy pocos pronósticos después de la intervención para determinar si los intervalos funcionan bien, es bueno saber que desde el inicio de la pandemia únicamente

en 2021 T1 cuando  $k = 1$  se tuvo un pronóstico fuera del intervalo, en todos los demás pronósticos con  $k = 1, \dots, 8$  los valores observados están dentro de los intervalos de pronóstico. En total se tienen 60 pronósticos después de la intervención (11 con  $k = 1$ , 10 con  $k = 2$ , ..., 4 con  $k = 8$ ), y solo uno falló. Por lo tanto, la intervención ayudó bastante a obtener mejores pronósticos por intervalos. En la figura 3.11 se muestra en qué casos los intervalos del DLM y DLM intervenido durante la pandemia contienen al valor real y en qué casos no. Es interesante ver que en el DLM, con la excepción de  $k = 7$ , siempre fallaron los intervalos de  $(Z_{33+k} | D_{33})$ , ya que  $t = 33$  es 2020 T1 y no se incluyó información sobre la pandemia. Posteriormente los intervalos mejoran debido a que el modelo fue capaz de corregir su error una vez que se observa el efecto de la pandemia. El DLM intervenido no tiene este problema, incluir la información externa sobre la pandemia en 2020 T1 para el pronóstico de 2020 T2 ayudó a que los intervalos fueran correctos desde el inicio.

### 3.4.6. Selección de modelo

Es claro que los DLM superaron la precisión de los pronósticos del VECM en todas las medidas consideradas. Entre los DLM, el intervenido es mejor sobre todos en el corto plazo, aunque conforme se incrementa el horizonte de pronóstico ambos modelos parecen tener desempeños parecidos. Debido a que en la intervención realizada también se incrementó la varianza para reflejar la incertidumbre creada por la pandemia, los intervalos de pronóstico del modelo intervenido funcionaron mejor durante la pandemia. Como se mencionó anteriormente, las intervenciones tienen como objetivo mejorar pronósticos y corregir el modelo cuando tiene problemas de ajuste en el pasado. Es claro que la intervención realizada cumple con ambas. Además de mejorar los pronósticos puntuales y los intervalos, en las secciones pasadas se mostró que con la intervención ya no se tiene el enorme incremento en  $S_t$  en 2020 T2 y la evolución del parámetro del PIB al inicio de la pandemia tiene más sentido. Por estas razones, se concluye que el mejor modelo es el DLM intervenido.

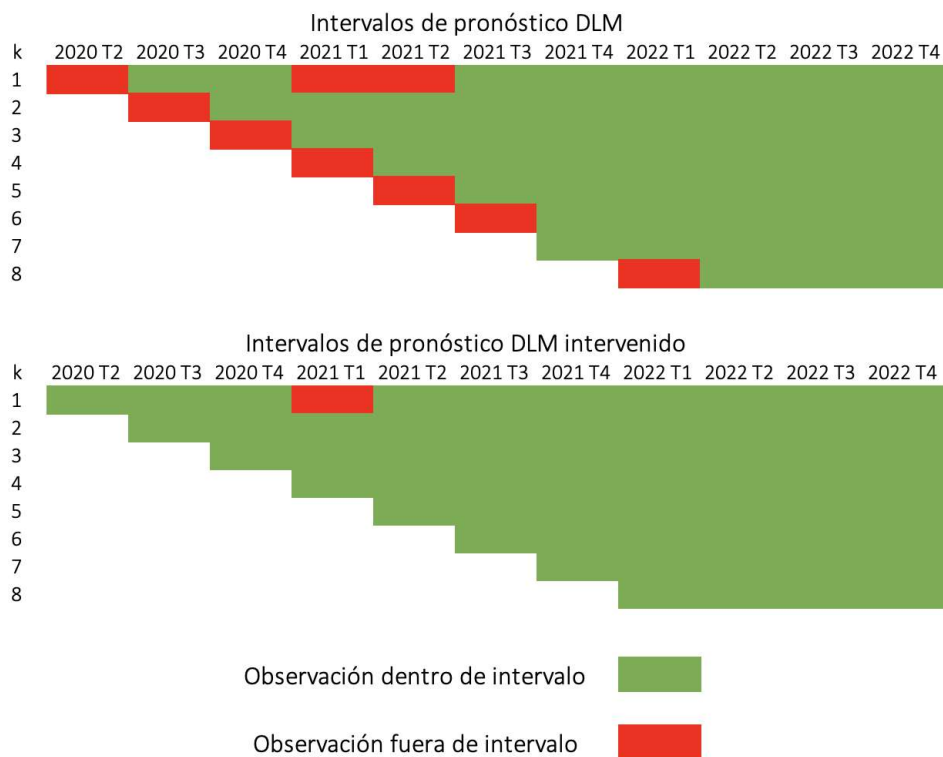


Figura 3.11: Cobertura de intervalos de pronóstico a  $k$  pasos durante la pandemia.

### 3.5. Pronósticos 2023 y 2024

Ahora que se definió el modelo ganador, es de interés realizar verdaderos pronósticos para la demanda de dinero futura con el DLM intervenido dada la información disponible en 2022 T4. Los pronósticos puntuales y sus intervalos en miles de millones de pesos se muestran en el cuadro 3.8. Los mismos pronósticos se graficaron en la figura 3.12 para tener una representación visual. Se puede ver que el modelo espera que el crecimiento en la demanda de efectivo provocado por la pandemia y los grandes incrementos en los cuartos trimestres sin disminuir en los primeros trimestres continúen.



Trimestre	Pronóstico	Límite inf.	Límite sup.
2023 T1	2536.62	2400.15	2680.84
2023 T2	2560.11	2397.55	2733.71
2023 T3	2573.11	2392.80	2767.01
2023 T4	2857.08	2626.44	3107.97
2024 T1	2877.38	2625.24	3153.75
2024 T2	2919.79	2644.73	3223.46
2024 T3	2943.37	2652.84	3265.72
2024 T4	3277.84	2928.38	3669.00

Cuadro 3.8: Pronósticos 2023 y 2024 (miles de millones de pesos).

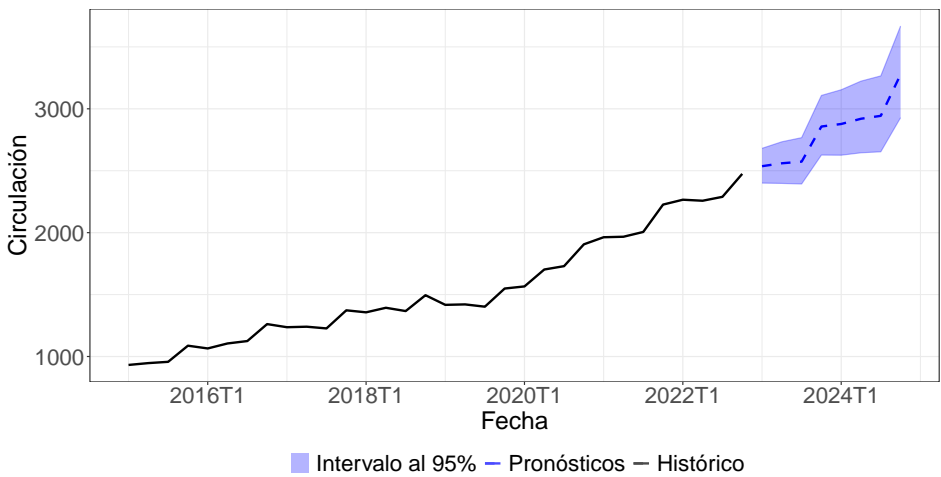


Figura 3.12: Pronósticos 2023 y 2024 (miles de millones de pesos).



## Capítulo 4

# Conclusión

### 4.1. Resumen de los resultados

El modelo lineal dinámico es una buena alternativa para la realización de pronósticos de la demanda de efectivo en el mediano plazo. Tiene características que lo hacen muy útil, como tener parámetros dinámicos, la actualización sencilla del modelo cuando se observan nuevas observaciones, y la facilidad para incluir información externa. En el DLM las intervenciones son muy flexibles y naturales, se pueden incrementar y disminuir los parámetros y sus varianzas con base en el conocimiento y creencias que se tienen. Otro aspecto positivo de las intervenciones en el DLM es que son consistentes con el resto del modelo. Así se pueden obtener distribuciones filtradas para realizar análisis retrospectivo del modelo intervenido sin ningún problema. En 2020 T2 se realizó una intervención con el conocimiento que se tenía en 2020 T1 sobre el impacto de la pandemia en la demanda de efectivo. Se incrementó el intercepto y su varianza para ajustar el incremento en la demanda de efectivo y en la incertidumbre. Gracias a esta intervención el modelo mejoró. Los pronósticos puntuales y los intervalos del modelo intervenido fueron más precisos. Además, las distribuciones filtradas de los parámetros y la evolución de la estimación de la varianza de observación muestran que la intervención ayudó al modelo a ser coherente con lo que se esperaba. Por último, se realizaron pronósticos para 2023 y 2024 considerando la distribución de parámetros posterior en 2022 T4 del modelo intervenido.

## 4.2. Limitaciones y próximos pasos

Una limitación del DLM empleado es que se necesitan pronósticos de las variables explicativas. Es sencillo obtener pronósticos de instituciones fiables sobre el PIB y la inflación debido a que son variables muy importantes, pero si se deseara utilizar alguna variable que no suele ser pronosticada por agentes externos sería necesario desarrollar otros modelos para pronosticar esas variables y así poder incluirlas en el DLM presentado. Además, estos pronósticos tienen una varianza que no se considera en el DLM. Sin embargo, en el caso del PIB y la inflación tiene sentido suponer que los pronósticos de instituciones especializadas son mejores que las que se podrían hacer en este trabajo, por eso se decidió tomar la media de esos pronósticos y no crear modelos propios.

La teoría económica detrás de la demanda de dinero es compleja y tiene una larga historia. El objetivo de esta tesis es mostrar el valor de los DLM en la realización de pronósticos, en este caso en la demanda de dinero. Aunque se hizo una revisión de la literatura para incluir variables que históricamente han formado parte de estos modelos, el enfoque de la tesis no es la teoría económica. El DLM propuesto tiene una función de demanda de dinero sencilla y es adecuada para ilustrar las ventajas de los DLM. Sin embargo, tiene algunos puntos criticables desde el punto de vista de la teoría económica. Un modelo de demanda de dinero por lo general modela un agregado monetario real y las tasas de interés son variables de costo de oportunidad muy importantes, pero en el DLM empleado se modeló un agregado monetario nominal y no se incluyó ninguna tasa de interés. Se mostró que los DLM pueden ser muy útiles en este contexto, pero es necesario emplear una función de demanda de dinero que sea coherente con la teoría económica y con la realidad de México aunque el modelo resultante sea más complejo.

El problema de la demanda de efectivo en México no se resuelve únicamente pronosticando la demanda agregada, en realidad es mucho más complejo. También es necesario considerar la demanda de cada denominación de billetes y monedas, la demanda en las distintas regiones del país y los billetes deteriorados que deben ser reemplazados. Para esto, un próxi-

mo paso sería aplicar un DLM multivariado para pronosticar la demanda por denominación o región. Los resultados para la actualización, pronósticos y distribuciones filtradas del modelo multivariado son muy parecidas al univariado. La principal dificultad es que si se asume que la matriz de covarianza de observación  $\mathbf{V}_t$  es desconocida, el modelo multivariado se complica. Si se desea evitar las complejidades de este caso podría usarse una estimación, pero esto implicaría subestimar la varianza del modelo aún más. Dado que el DLM univariado tuvo buenos resultados con la demanda agregada, sería razonable probar un DLM multivariado para pronosticar la demanda por denominación o región.

El DLM multivariado se definió en 2.2.1. Es muy parecido al univariado, las diferencias son que  $\mathbf{Y}_t$  es un vector  $r \times 1$ ,  $\mathbf{F}_t$  es una matriz  $n \times r$  en vez de un vector  $n \times 1$ , y  $\mathbf{V}_t$  es una matriz  $r \times r$  en lugar de un número. En  $\mathbf{F}_t$  las distintas denominaciones pueden tener diferentes valores en las variables explicativas. No tendría sentido considerar un diferente PIB o inflación en las denominaciones porque son a nivel nacional, pero si se modela la demanda a nivel región podría ser útil considerar el PIB y la inflación de cada región. Otro punto importante son las intervenciones. Se sabe que la pandemia no afectó por igual a todas las denominaciones, los billetes de \$500 aumentaron más que el resto. Sería útil que la intervención considerara que la pandemia provocó que los billetes de denominaciones altas tuvieron un incremento mayor que los de menor valor. Para modelar esto una posibilidad es tener un intercepto diferente para cada grupo de denominación (por ejemplo, un intercepto común para billetes de \$200 a \$1000, otro para billetes de \$20 a \$100 y otro para monedas). Así se podrían intervenir los interceptos con base en lo que se espera de cada grupo de denominación. Obviamente se tienen que modelar las denominaciones por separado, pero formar grupos para que compartan interceptos ayudaría a que el modelo sea menos complejo. La flexibilidad de los DLM puede ser muy útil para resolver problemas multivariados de pronóstico.



## Apéndice A

# Resultados de probabilidad y estadística

La teoría de esta sección proviene de [12, pp. 637-642].

### 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  $\omega$  es independiente de  $\nu$ .

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

Entonces,

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

Por lo tanto,

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

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

$$(\theta \mid Y) \sim N[m, C],$$

donde

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

y

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

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

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

### 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  $\phi \sim \chi_n^2$ , 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  $\phi$  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,

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

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

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

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  $\phi$ ,  $\mathbf{X} \sim \mathbf{N}[\mathbf{m}, \mathbf{C}\phi^{-1}]$ .  $\mathbf{m}$  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  $V[\mathbf{X}]$  está multiplicado por  $\phi$ . Los resultados son parecidos a los del caso univariado.

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

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

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

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



# Apéndice B

## VECM

### B.1. Vector autorregresivo

Un proceso VAR( $p$ ) tiene la siguiente forma:

$$\mathbf{Y}_t = \boldsymbol{\nu} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \dots + \mathbf{A}_p \mathbf{Y}_{t-p} + \mathbf{v}_t,$$

donde  $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{Kt})^\top$  es un vector aleatorio ( $K \times 1$ ),  $\mathbf{A}_i$  son matrices de coeficientes ( $K \times K$ ),  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_K)^\top$  es un vector de intercepto ( $K \times 1$ ). Por último,  $\mathbf{v}_t = (v_{1t}, \dots, v_{Kt})^\top$  es un vector ( $K \times 1$ ) de ruido blanco, con  $E[\mathbf{v}_t] = \mathbf{0}$ ,  $E[\mathbf{v}_t \mathbf{v}_t^\top] = \boldsymbol{\Sigma}_v$ , y  $E[\mathbf{v}_t \mathbf{v}_s^\top] = \mathbf{0}$  para  $t \neq s$ .

Se dice que un proceso VAR( $p$ ) es estable si su polinomio característico no tiene raíces dentro o sobre el círculo unitario complejo:

$$\det(\mathbf{I}_K - \mathbf{A}_1 z - \dots - \mathbf{A}_p z^p) \neq 0 \quad \text{para } |z| \leq 1.$$

Además, un proceso VAR( $p$ ) estable es estacionario. Los primeros y segundos momentos de los procesos estacionarios no varían a lo largo del tiempo. El problema es que muchas series económicas son no estacionarias. Sin embargo, un proceso VAR puede generar tendencias si su polinomio característico tiene raíces unitarias ( $z = 1$ ). Las variables generadas por estos procesos están integradas. Un proceso univariado con  $d$  raíces unitarias se dice que está integrado de orden  $d$ , denotado como  $I(d)$ . Si un proceso VAR es inestable únicamente por la existencia de raíces unitarias,

se puede volver estable diferenciando sus componentes. Sin embargo, diferenciar puede traer complicaciones. Por ejemplo, si todos los elementos de un  $\text{VAR}(p)$  son  $I(d)$ , un posible problema es que al diferenciar cada uno de los componentes se puede distorsionar la relación de las variables originales. Por esta razón, es útil encontrar otra forma de modelar estas variables.

## B.2. Vector error correction model

Cuando algunas variables son integradas, es posible que exista una combinación lineal entre ellas que es integrada de un orden menor. En ese caso, se dice que las variables están cointegradas. Un proceso  $\mathbf{Y}_t$  que es  $I(d)$  está cointegrado si existe una combinación lineal  $\beta^\top \mathbf{Y}_t$  con  $\beta \neq \mathbf{0}$  que es  $I(d-b)$ . A  $\beta$  se le conoce como vector de cointegración. Por ejemplo, si se tiene el vector cointegrado  $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{Kt})^\top$ , donde todas las series son  $I(1)$ , entonces  $\beta^\top \mathbf{Y}_t = \beta_1 Y_{1t} + \dots + \beta_K Y_{Kt}$  es estacionaria. Por esta razón, se puede interpretar  $\beta^\top \mathbf{Y}_t$  como un equilibrio a largo plazo. En cualquier periodo habrá desviaciones de este equilibrio, pero siempre cerca del equilibrio.

Si las variables individuales son  $I(1)$  o  $I(0)$ , entonces el proceso  $\text{VAR}(p)$  tiene la representación de VECM (vector error correction model),

$$\Delta \mathbf{Y}_t = \Pi \mathbf{Y}_{t-1} + \Gamma_1 \Delta \mathbf{Y}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{Y}_{t-p+1} + \mathbf{u}_t,$$

con  $\Pi = -(\mathbf{I}_K - \mathbf{A}_1 - \dots - \mathbf{A}_p)$ , y  $\Gamma_i = -(\mathbf{A}_{i+1} - \dots - \mathbf{A}_p)$  para  $i = 1, \dots, p-1$ . Si  $\Pi$  es de rango  $r$ , entonces  $\Pi = \alpha \beta^\top$ , con  $\alpha$  y  $\beta$  de dimensiones  $(K \times r)$  y rango  $r$ . Por lo tanto, el VECM también puede representarse como

$$\Delta \mathbf{Y}_t = \alpha \beta^\top \mathbf{Y}_{t-1} + \Gamma_1 \Delta \mathbf{Y}_{t-1} + \dots + \Gamma_{p-1} \Delta \mathbf{Y}_{t-p+1} + \mathbf{u}_t.$$

Si se reacomodan los términos de la siguiente manera,

$$\alpha \beta^\top \mathbf{Y}_{t-1} = \Delta \mathbf{Y}_t - \Gamma_1 \Delta \mathbf{Y}_{t-1} - \dots - \Gamma_{p-1} \Delta \mathbf{Y}_{t-p+1} - \mathbf{u}_t,$$

todos los elementos del lado derecho de la ecuación son estacionarios, pues todas las variables son  $I(1)$  o  $I(0)$ , y el ruido blanco también es estacionario. Como el lado derecho es estacionario, entonces, por la igualdad,  $\alpha \beta^\top \mathbf{Y}_{t-1}$  también debe ser estacionario, y sigue siendo estacionaria si se multiplica

por  $(\boldsymbol{\alpha}^\top \boldsymbol{\alpha})^{-1} \boldsymbol{\alpha}^\top$ . Por lo tanto,  $\boldsymbol{\beta}^\top \mathbf{Y}_{t-1}$  es estacionario y representa una relación de cointegración. Si solo se hubieran obtenido las primeras diferencias del proceso VAR( $p$ ) se eliminaría el término de cointegración, por lo que se perderían relaciones importantes. [19, pp. 237-249]





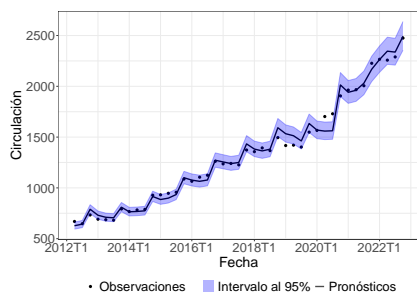
# Apéndice C

## Gráficas de los modelos

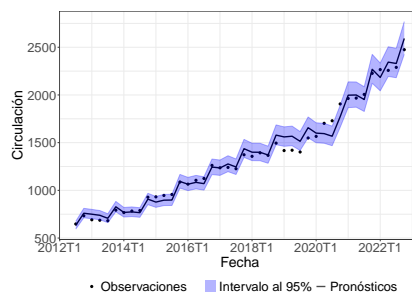
Se presentan los pronósticos y los intervalos de

$$Z_{t+k} = \exp(Y_{t+k}) = \text{Efectivo}_{t+k}.$$

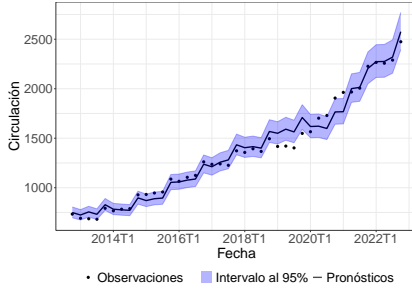
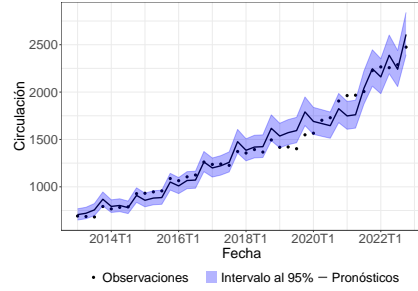
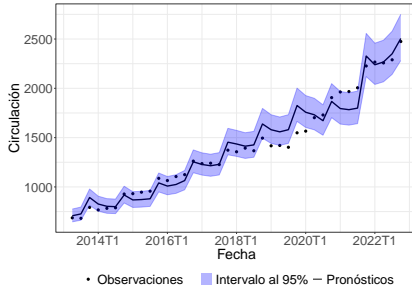
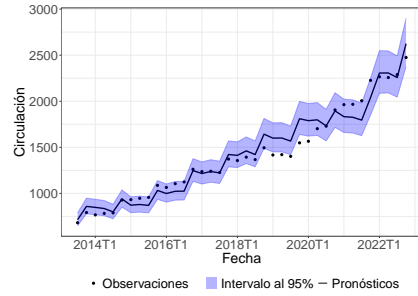
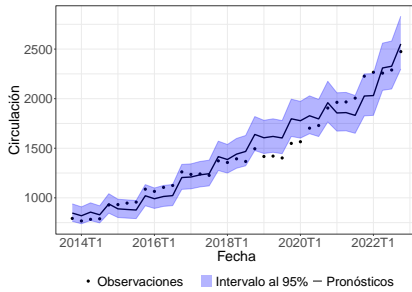
### C.1. Gráficas Modelo lineal dinámico



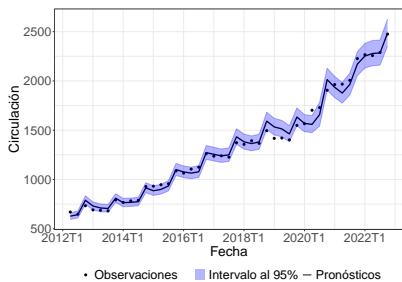
Pronósticos ( $Z_{t+2} \mid D_t$ ).



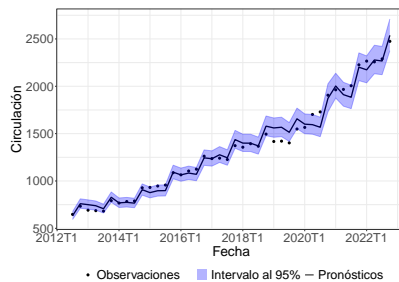
Pronósticos ( $Z_{t+3} \mid D_t$ ).

Pronósticos  $(Z_{t+4} \mid D_t)$ .Pronósticos  $(Z_{t+5} \mid D_t)$ .Pronósticos  $(Z_{t+6} \mid D_t)$ .Pronósticos  $(Z_{t+7} \mid D_t)$ .Pronósticos  $(Z_{t+8} \mid D_t)$ .

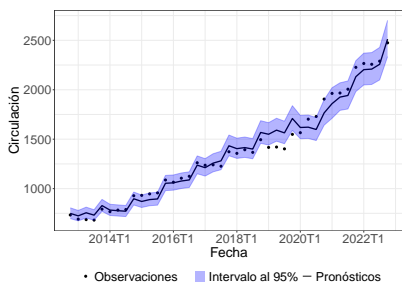
## C.2. Gráficas Modelo lineal dinámico intervenido



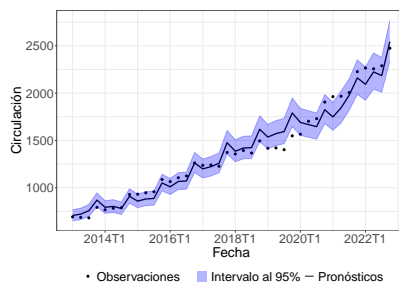
Pronósticos  $(Z_{t+2} | D_t)$ .



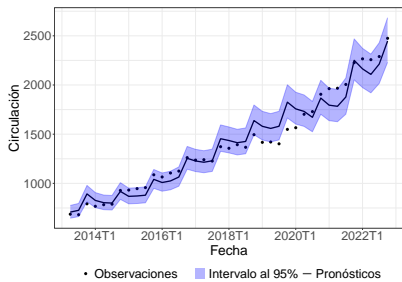
Pronósticos  $(Z_{t+3} | D_t)$ .



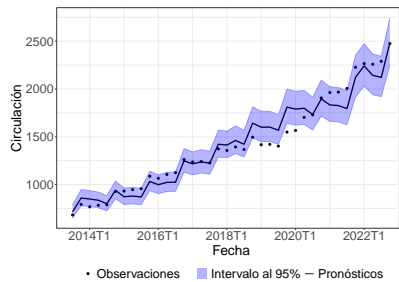
Pronósticos  $(Z_{t+4} | D_t)$ .



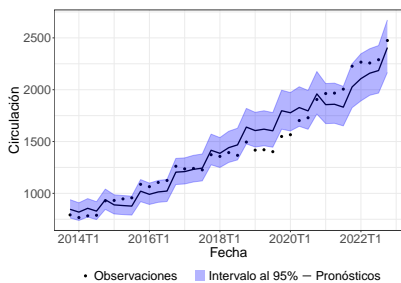
Pronósticos  $(Z_{t+5} | D_t)$ .



Pronósticos  $(Z_{t+6} | D_t)$ .



Pronósticos  $(Z_{t+7} | D_t)$ .

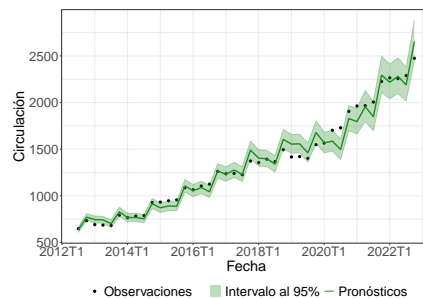


Pronósticos ( $Z_{t+8} \mid D_t$ ).

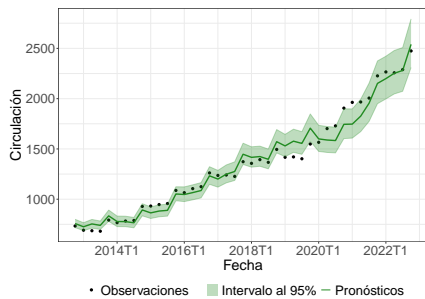
### C.3. Gráficas VECM



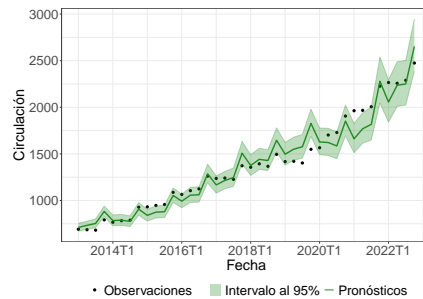
Pronósticos a 2 pasos.



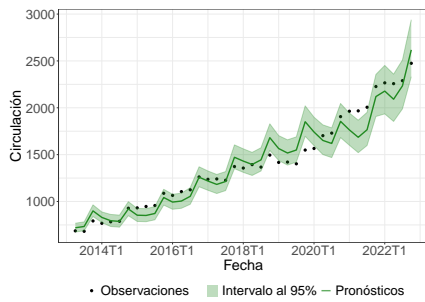
Pronósticos a 3 pasos.



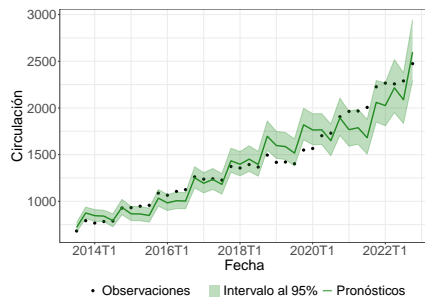
Pronósticos a 4 pasos.



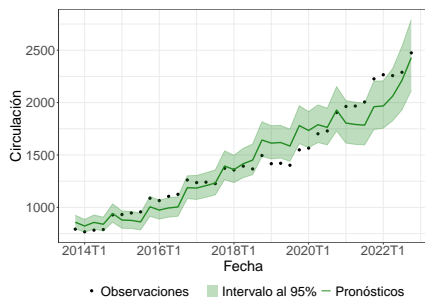
Pronósticos a 5 pasos.



Pronósticos a 6 pasos.



Pronósticos a 7 pasos.



Pronósticos a 8 pasos.



# 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](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/consultarDirectorioIn>

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

- [7] *Importancia de los billetes y monedas en la vida diaria*. Banco de México. Disponible en: <https://anterior.banxico.org.mx/divulgacion/billetes-y-monedas/importancia-billetes-monedas-.html#Queeseldinero>.
- [8] Sriram, Subramanian: *Demand for M2 in an Emerging-Market Economy: An Error-Correction Model for Malaysia*. IMF Working Paper No. 99/173, 1999. Disponible en: <https://www.imf.org/external/pubs/ft/wp/1999/wp99173.pdf>.
- [9] 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>.
- [10] 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>.
- [11] 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>.
- [12] M. West, J. Harrison: *Bayesian Forecasting and Dynamic Models*. Springer, Nueva York, segunda edición, 1997, ISBN 0-387-94725-6.
- [13] J. Durbin, S.J. Koopman: *Time Series Analysis by State Space Methods*. Oxford University Press, primera edición, 2005, ISBN 0-19-852354-8.
- [14] Sriram, Subramanian: *A Survey of Recent Empirical Money Demand Studies*. IMF Staff Papers Vol. 47, No. 3, 2001. Disponible en: <https://www.imf.org/external/pubs/ft/stp/2001/v47n3/stp200103a.pdf>.



- [//www.imf.org/external/pubs/ft/staffp/2001/01/pdf/Sriram.pdf](http://www.imf.org/external/pubs/ft/staffp/2001/01/pdf/Sriram.pdf).
- [15] Elizabeth Eli Holmes, Eric J. Ward, Mark D. Scheuerell Kellie Wills: *MARSS: Multivariate Autoregressive State-Space Modeling*, 2023. <https://CRAN.R-project.org/package=MARSS>, R package version 3.11.8.
- [16] Elizabeth Eli Holmes, Eric J. Ward, Mark D. Scheuerell: *Applied Time Series Analysis for Fisheries and Environmental Sciences*. 2021. Disponible en: [https://atsa-es.github.io/atsa-labs/Applied\\_Time\\_Series\\_Analysis.pdf](https://atsa-es.github.io/atsa-labs/Applied_Time_Series_Analysis.pdf).
- [17] *Informe Trimestral Enero-Marzo 2020*. Informe técnico, Banco de México, 2020. Disponible en: <https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7B23C2DCA8-4AD3-FBE0-B0BF-4D30C8066B84%7D.pdf>.
- [18] Spyros G. Makridakis, Steven C. Wheelwright, Rob J. Hyndman: *Forecasting: Methods and Applications*. John Wiley Sons, Inc., tercera edición, 1998, ISBN 978-0-471-53233-0.
- [19] Lütkepohl, H.: *New Introduction to Multiple Time Series Analysis*. Springer, primera edición, 2005, ISBN 3-540-40172-5.