



UNIVERSIDAD NACIONAL DE COLOMBIA

# Paneles cointegrados Bayesianos: una aplicación en *márketing* digital

*Bayesian cointegrating panels: an application in digital marketing*

**Juan David Carranza Sánchez**

Universidad Nacional de Colombia

Facultad de ciencias

Departamento de estadística

Bogotá, Colombia

2023



# Paneles cointegrados Bayesianos: una aplicación en *márketing* digital

*Bayesian cointegrating panels: an application in digital marketing*

**Juan David Carranza Sánchez**

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:

**Magíster en Ciencias - Estadística**

Director:

Ph.D. Juan Camilo Sosa Martínez

Línea de Investigación:

Estadística Bayesiana

Universidad Nacional de Colombia

Facultad de ciencias

Departamento de estadística

Bogotá, Colombia

2023



## Dedicatoria

Dedico este trabajo a mi madre, que nunca dejó de creer en mí

A mi abuela, que me aguantó las interminables noches de trabajo.

A mi padre, que me ofreció su apoyo durante esta etapa de mi vida.

A mis gatos: Aguacate, Malteada y Mango, que siempre me mantuvieron con actitud positiva.

A ustedes.



# Agradecimientos

Agradezco el apoyo de todas las personas que me apoyaron durante esta etapa de mi vida, de manera directa o indirecta, con sus consejos, tiempo, sabiduría y compañía. Gracias también a esas pequeñas cosas que me ayudaron a no rendirme y a mantener siempre un ritmo constante y disciplinado, a mostrarme que no importa cuán dura sea la situación, es posible adaptarse y continuar a pesar de las adversidades. Gracias a From Software por sus obras que me motivaron a tomar un camino fuera de mi zona de confort.





## Resumen

En el presente documento se desarrolla la metodología de paneles cointegrados Bayesianos de Koop, Leon-Gonzalez y R. Strachan (2008) en el ámbito de *marketing* digital, particularmente en la modelación de un sistema donde se modelen métricas de rendimiento de inversión como clicks e impresiones de una campaña digital dada, la inversión planeada y las sesiones obtenidas. Así, se quiere evaluar cómo reacciona el sistema ante perturbaciones en inversión por cambios en la estrategia de *flighting* o épocas y temporadas de inversión, y su repercusión en la visibilidad de las campañas dadas. De esta manera, se utiliza el modelo con un conjunto de datos reales de *marketing* con diferentes campañas de inversión sobre un mismo territorio geográfico. Por medio de la descomposición de la varianza del error de predicción, los resultados indican que los clicks e impresiones tienen un impacto significativo en la generación de sesiones. También se evalúa la capacidad de estimabilidad del algoritmo de estimación propuesto por medio de un estudio exhaustivo de simulación que considera diferentes procesos. Los resultados indican que el algoritmo goza de capacidades importantes en términos de estimabilidad y precisión.

**Palabras clave:** modelos Bayesianos, modelos multiecuacionales, series de tiempo, datos longitudinales, espacios de cointegración, *marketing* digital.

# Abstract

This paper develops the Bayesian cointegrated panel methodology of the Koop, Leon-Gonzalez y R. Strachan (2008) paper in the field of digital marketing, particularly in the modeling of a system where key ROI metrics such as clicks or impressions of a given digital campaign are taken into account. Thus, the objective is to evaluate how the system reacts to investment perturbations due to changes in the investment strategy and its impact on the visibility of the given campaigns. In this way, the model is used with a set of real marketing data with different investment campaigns over the same geographic territory. By the using of forecast error variance decomposition, the results indicate that clicks and impressions have a significant impact on session generation. The estimability of the proposed estimation algorithm is also evaluated through a comprehensive simulation study that considers different processes. The results indicate that the algorithm has significant capabilities in terms of estimability and accuracy.

**Keywords:** Bayesian models, multiequational models, time series, longitudinal data, cointegration spaces, digital marketing.

# Declaración

Me permito afirmar que he realizado la presente tesis de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en la propia tesis. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los he reconocido en el presente trabajo. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de tesis.

En este trabajo, bajo la supervisión del profesor Juan Camilo Sosa Martínez, es de mi autoría:

- La estimación de los coeficientes de un modelo panel cointegrado mediante las ecuaciones matriciales presentadas en Koop, Leon-Gonzalez y R. Strachan (2008) que describen los procesos multivariados y univariados de los parámetros, según corresponda.
- La modificación del proceso de actualización de las matrices de varianzas y covarianzas de los coeficientes del modelo para la actualización adecuada de los factores de largo plazo en Koop, Leon-Gonzalez y R. Strachan (2008).
- El desarrollo, implementación, optimización y evaluación del muestreador de Gibbs para ajustar el modelo.
- El estudio de simulación realizado en el capítulo 5, con sus respectivos anexos y análisis estadísticos.
- El análisis, anonimización y estudio de datos reales presentados en el capítulo 6, con su interpretación y recomendaciones.
- El código con el que fue posible realizar el modelo, gráficos y análisis, junto con las cadenas de los parámetros sin el uso de librerías.

Bogotá, D.C., 01.11.2023



# Lista de Figuras

3-1. Tasa de Desempleo Enero 2015 - Enero 2023. <b>Fuente:</b> DANE. . . . .	10
3-2. Tasa de Desempleo Enero 2015 - Enero 2023, diferenciada. <b>Fuente:</b> DANE. . .	12
3-3. Tasa de Desempleo vs ISE Enero 2015 - Enero 2023. <b>Fuente:</b> DANE . . . . .	14
4-1. DAG modelo, elaboración propia. . . . .	39
6-1. Full funnel, tomado de Amazon Ads. . . . .	48
6-2. Métricas en las etapas del <i>marketing funnel</i> , elaboración propia. . . . .	49
6-3. Campaña 1, variables del sistema. . . . .	52
6-4. Campaña 2, variables del sistema. . . . .	52
6-5. Campaña 3, variables del sistema. . . . .	53
6-6. Montos relativos de cada componente en el total. . . . .	53
6-7. ACF de cada serie en primeras diferencias, por campaña. . . . .	55
6-8. Estadísticos posteriores del modelo. . . . .	58
6-9. Descomposición de la varianza del error de predicción, canal 1. . . . .	60
6-10.Descomposición de la varianza del error de predicción, canal 2. . . . .	62
6-11.Descomposición de la varianza del error de predicción, canal 3. . . . .	63
6-12.Tamaños efectivos de muestra de los parámetros. . . . .	64
6-13.Errores de Monte Carlo de los parámetros, omitiendo $\nu = 3.12$ para efectos de visualización. . . . .	65
A-1. Resultados Caso 1 Simulación: Comparación posteriores con valores simula- dos. . . . .	69

---

<b>A-2.</b> Resultados Caso 2 Simulación: Comparación posteriores con valores simulados. . . . .	70
<b>A-3.</b> Resultados Caso 3 Simulación: Comparación posteriores con valores simulados. . . . .	71
<b>A-4.</b> Tamaños Efectivos de Muestra, todas las simulaciones. . . . .	72
<b>A-5.</b> Errores de Monte Carlo, todas las simulaciones. . <b>NOTA:</b> $\nu$ descartado por efectos de visualización, EMC = 123, 32, 57 respectivamente . . . . .	72

# Lista de Tablas

<b>3-1.</b> Resultados ADF Test - Serie de desempleo. Elaboración propia. . . . .	11
<b>3-2.</b> Resultados ADF Test - Serie de desempleo, diferenciada. Elaboración propia.	12
<b>3-3.</b> Resultados ADF Test - Residuales Tasa de Desempleo vs. ISE. Elaboración propia. . . . .	14
<b>6-1.</b> P-valores para prueba ADF tipo 1 con tres rezagos. . . . .	54
<b>6-2.</b> P-valores encontrados por medio de la metodología de Tsay, con sus respectivos rezagos. . . . .	54
<b>6-3.</b> Resultados de la log-verosimilitud de acuerdo con el número de relaciones de cointegración usadas. . . . .	57
<b>6-4.</b> Criterios de información para evaluar el número de rezagos adecuados para el modelo. Elaboración propia. . . . .	58
<b>6-5.</b> Diferencia entre el modelo estimado por medios frecuentistas vs. metodología Bayesiana. . . . .	65
<b>A-1.</b> Evaluación de la precisión de las simulaciones. . . . .	68

# Acrónimos

En el presente documento se usan los siguientes acrónimos, incluyendo la universalidad de algunas siglas en inglés.

ACF	Función de autocorrelación
ADF	Test de Dickey-Fuller aumentado
AIC	Criterio de información Akaike
BIC	Criterio de información Bayesiano
DAG	Grafo acíclico dirigido
DANE	Departamento administrativo nacional de estadística
DIC	Criterio de información de devianza
FEVD	Descomposición de varianza del error de predicción
GPU	Unidad de procesamiento gráfico
i.i.d.	Independientes e idénticamente distribuidas
IRF	Función de impulso respuesta
ISE	Índice de seguimiento económico
MAE	Error medio absoluto
MCMC	Cadenas de Markov de Monte Carlo
MMM	Modelo de mix de medios
PIB	Producto interno bruto
PPP	P-valores predictivos posteriores
RAM	Memoria de acceso aleatorio
RMSE	Error cuadrático medio
TD	Tasa de desempleo
VAR	Vector auto regresivo
VEC	Vector de corrección de errores
WAIC	Criterio de información de Watanabe-Akaike



# Notación

En el presente documento, se usa la siguiente notación

$x_t, x(t)$	Serie temporal
$T$	Intervalo de observación
$(\Omega, \mathcal{F}, P)$	Espacio de probabilidad
$\gamma_x(r, s), \text{Cov}(X_r, X_s)$	Covarianza entre $X$ medida en el tiempo $r$ y en el tiempo $s$
$E(x)$	Valor esperado de $x$
$\mathbb{Z}$	Conjunto de números enteros
$\mathbf{X}, \mathbf{X}^\top$	Matriz $\mathbf{X}$ , matriz transpuesta $\mathbf{X}^\top$
$\Delta y_{i,t}$	Serie temporal diferenciada observada en el individuo $i = 1, \dots, N$ en el momento $t = 1, \dots, T$
$x_t \sim \text{CI}(d, b)$	Serie $x_t$ cointegrada de orden $d$ con $b$ relaciones de cointegración
$x_t \sim \text{I}(d - b)$	Serie $x_t$ integrada de orden $d - b$
$r$	Relaciones de cointegración
$N$	Número de individuos en el panel
$n$	Número de variables en el sistema
$\hat{y}_{k,t}$	Serie de tiempo multivariada, observada en el momento $t = 1, \dots, T$ de la colección $k = 1, \dots, K$
$Y_{i,t}$	Datos longitudinales o panel observados en el individuo $i = 1, \dots, N$ en el momento $t = 1, \dots, T$
$\epsilon_{i,t}$	Error en el individuo $i = 1, \dots, N$ observado en el momento $t = 1, \dots, T$
$\text{N}(0, \sigma^2)$	Distribución normal univariada con media 0 y varianza $\sigma^2$
$\text{N}_{(NT) \times 1}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$	Distribución normal multivariada de tamaño $(NT) \times 1$ con vector de medias $\mathbf{X}\boldsymbol{\beta}$ y matriz de covarianzas $\sigma^2\mathbf{V}$
$\boldsymbol{\theta}$	Conjunto de parámetros

$p(\mathbf{y} \mid \boldsymbol{\theta})$	Distribución muestral de los datos $\mathbf{y}$ dado un conjunto de parámetros $\boldsymbol{\theta}$ , identificada como la verosimilitud
$p(\boldsymbol{\theta})$	Distribución previa de $\boldsymbol{\theta}$
$\int_{\Theta}$	Integral sobre el espacio de parámetros $\Theta$
$B$	Número de simulaciones
$\theta_{-j}^{t-1}$	Todos los componentes de $\boldsymbol{\theta}$ menos $\theta_j$ en el punto de simulación $t - 1$
$\tilde{y}, \hat{y}$	Datos no observados de $y$ , datos simulados $\hat{y}$ a partir del vector de simulaciones de $\boldsymbol{\theta}$
$T(\mathbf{y})$	Estadísticos de prueba
$\mathbf{y}^{rep}$	Datos hipotéticos que pudieron ser observados bajo el estado actual de conocimiento
$p_B$	Valor-p Bayesiano
$\mathfrak{p}_i$	Espacio vectorial generado por $\boldsymbol{\beta}$
$\boldsymbol{\Sigma}$	Matriz de varianzas y covarianzas
$\mathbf{I}_T$	Matriz identidad de tamaño $T$
$(\boldsymbol{\Sigma} \otimes \mathbf{I}_T)$	Producto Kronecker entre $\boldsymbol{\Sigma}$ e $\mathbf{I}_T$
$\text{tr}(\boldsymbol{\Sigma})$	Traza de la matriz $\boldsymbol{\Sigma}$
$\propto$	Operador de proporcionalidad
$\mathbf{H}_{\perp}$	Complemento ortogonal de la matriz $\mathbf{H}$
$\text{IW}(\boldsymbol{\Sigma}, T)$	Distribución Wishart inversa con matriz de escala $\boldsymbol{\Sigma}$ y $T$ grados de libertad
$G(a, b)$	Distribución Gamma con parámetros $a$ y $b$
$U(a, b)$	Distribución Uniforme con parámetros $a$ y $b$
$\text{BN}(a, b)$	Distribución Binomial Negativa con parámetros $a$ y $b$

# 1. Introducción

En la actualidad, se ha evidenciado un incremento significativo en el interés por los datos tomados a través del tiempo, ya que permiten modelar fenómenos que son indexados por un componente temporal y tienen ciertas particularidades que obligan al investigador interesado en ciertos fenómenos a tener en cuenta al momento de establecer y rechazar hipótesis.

En particular, en el *marketing* existe el interés en entender la relación a través del tiempo de un conjunto de variables que pueden guardar un cierto equilibrio, cómo pueden ser las ventas de un *retailer*<sup>1</sup> en relación con la distribución o promociones de un producto, e incluso, la inversión en medios que tiene dicho producto. En *marketing*, la metodología usual es entender cómo la inversión en un medio o canal particular afecta un indicador clave de negocio en el corto plazo como el *branding*<sup>2</sup>, el alcance, la visibilidad o directamente las conversiones o ventas, pero se obvia, o no se relaciona directamente el efecto de equilibrio o de largo plazo. En algunas ocasiones, es posible relacionar varios productos hermanos que tienen una estrecha relación por su similaridad a través de diferentes mercados, o mercados que se comportan de manera similar con algunos productos dada su naturaleza complementaria. Los análisis panel permiten incorporar efectos homogéneos o heterogéneos con la bondad de compartir información entre el conjunto de individuos preestablecido. En *marketing*, esta información es clave para gestionar y optimizar las campañas de medios asociadas, así como identificar audiencias y patrones de consumo inherentes a mercados o a sus geografías.

Este trabajo se centra en aplicar la metodología de modelos panel cointegrados, motivado en la búsqueda del establecimiento de relaciones de cointegración y relaciones de corto plazo en un sistema de múltiples ecuaciones con múltiples individuos, presentada en detalle en el

---

<sup>1</sup>Comerciantes que adquieren gran cantidad de productos o grandes superficies

<sup>2</sup>La imagen de marca que tienen los consumidores de un cierto bien o servicio

trabajo de Koop, Leon-Gonzalez y R. Strachan (2008), donde se explora una metodología que permite utilizar la inferencia Bayesiana en el contexto de series de tiempo multivariadas y cointegradas. La ventaja de estimar el modelo por medio de una metodología Bayesiana es la posibilidad de contar con la distribución empírica de los parámetros, que permite aplicar funciones entre los mismos que no podrían darse en una metodología frecuentista, así como tener una distribución de los estadísticos del modelo general con base en las distribuciones de los parámetros.

Esta metodología tiene la característica de permitir diferentes relaciones de cointegración a lo largo del panel, añadiendo así de manera directa la posibilidad de diferenciar los equilibrios entre los individuos del panel. En la sección 2 el lector encontrará todo lo relacionado a los trabajos principales que han usado metodologías similares, no sólo en *marketing*, sino también, en el ámbito de la economía, en la sección 3 encontrará todo el marco teórico al que estará sujeto este trabajo, en la sección 4 se encuentra toda la ficha metodológica y desarrollo necesarios para llevar a cabo el análisis propuesto, en la sección 5 se plantea un ejercicio de simulación con el fin de evaluar la precisión del modelo para dar paso a la sección 6 donde se hace un breve análisis descriptivo de los datos y la aplicación del modelo junto con los entregables finales, y finalmente en la sección 7 encontrará la discusión final y los próximos pasos a seguir.

## 2. Estado del arte

Dada la naturaleza de la macroeconomía, resulta de especial interés el estudio de relaciones entre diferentes variables a lo largo del tiempo; no sólo en el corto plazo, sino también en cómo el largo plazo y el equilibrio de un sistema dado moldean la realidad inmediata. Así mismo, resulta adecuado revisar si hay otros individuos que presenten la misma dinámica o al menos una dinámica similar.

Tal es el caso de los países subsaharianos, donde Chigozie Nkalu (2020) conecta el consumo de energía, el desarrollo financiero medido por medio del porcentaje del PIB en deuda del sector privado en tenencia de los bancos privados, producto interno bruto o PIB per cápita, urbanización aproximada por densidad poblacional y crecimiento poblacional. A través de una metodología basada en paneles cointegrados calculados por medio de la metodología planteada por Pedroni (1997), y entre sus hallazgos principales, se encuentra una relación de equilibrio significativa entre las variables de interés, junto con un rezago de dos períodos en el corto plazo de ajuste hacia el equilibrio general del sistema. Con estos resultados, es posible hacer una recomendación de política pública en pro del desarrollo financiero con los efectos que ello pueda tener sobre el consumo energético a largo plazo del país, con tal de mantener cierta prudencialidad en las finanzas internas del panel de países, conformado por 48 países.

Otro caso en el cual el consumo de energía es relacionado nuevamente con producción, con un enfoque más microeconómico es donde Kumar y Madheswaran (2010) plantean evaluar su relación en el mercado de cemento con un panel conformado por las 18 principales empresas de cemento de la India. En este trabajo, el eje central se basa en el uso de las relaciones de corto plazo entre las variables ya mencionadas junto con el nivel de material disponible,

el número de trabajadores y la inversión en capital bruto realizada período a período para probar la hipótesis de causalidad en el sentido de Granger. Los resultados indican que tanto el nivel de material disponible en un período dado como la inversión en capital causan en el sentido de Granger al nivel de producción y consumo de energía.

El consumo de energía es de gran interés para los hacedores de política pública, debido a que es un gran predictor de las relaciones macroeconómicas que pueden ser determinantes a la hora de medir el crecimiento económico en un país, tal cómo lo evidencian Mahadevan y Asafu-Adjaye (2007) en un panel conformado por países, no sólo desarrollados, sino en vías de desarrollo de todas las regiones del mundo. Al realizar la distinción entre países desarrollados y en vías de desarrollo, es posible ver que en términos de causalidad, los segundos explican mejor su consumo de energía con base en su tasa de crecimiento del PIB, principalmente en países que son exportadores netos de energía, mientras que los primeros tienen mayor evidencia de que la relación es inversa, mostrando que en nivel general, los países desarrollados tienden a consumir más energía para mayor producción mientras que los países en vías de desarrollo tienden a presentar el comportamiento opuesto.

Un documento más reciente y con un enfoque más occidental escrito por Máté, Somosi y Farkas (2022) donde se usa un panel de países europeos conformados por los seis miembros fundadores y algunos países de Europa Oriental, presenta la relación entre producto interno bruto o PIB per cápita y expectativa de vida con respecto al gasto gubernamental en salud pública. Se encuentra que el número de períodos esperado para la convergencia al equilibrio de largo plazo es de cuatro rezagos totales. El contenido valioso de este trabajo, es incluir las respuestas acumuladas de cambios en el gasto gubernamental con respecto al cambio en el PIB per cápita, mostrando un incremento significativo del nivel de ingreso por cada punto porcentual aumentado en la inversión en salud pública.

En un estudio con mayor relación al *marketing*, Slotegraaf y Pauwels (2008) plantean una serie de preguntas y cuestiones relativas al impacto que tiene el *brand equity*, que es consiguado en Keller (1993) conocido cómo el valor agregado que tiene una marca de un producto frente al mismo producto sin una marca bien definida, compuesto por el *awareness* que evalúa la capacidad de los clientes de reconocer la marca; la *imagen de marca* que evalúa la

percepción de los consumidores de una marca particular y factores como la *consideración* a la hora de adquirir un producto similar entre dos marcas diferentes; y las innovaciones sobre las promociones como campañas de reducción de precios conocidas como campañas de *trade*, donde se ofrecen mayor cantidad de productos a cambio del mismo precio y otro tipo de diferentes combinaciones de beneficios para los usuarios.

Los autores plantean que cuando una marca tiene un *brand equity* alto y sólido, las elasticidades relacionadas con los esfuerzos promocionales sean superiores respecto a una marca con menor *brand equity*. Se desarrolla un enfoque de modelos cointegrados con el fin de responder preguntas básicas cómo la afectación a corto y largo plazo de las elasticidades cuando se tiene un cierto nivel de *brand equity*, los impulsos dinámicos de ventas frente a esfuerzos de *marketing* y el efecto inmediato de una promoción en las ventas. Debido a que se tiene un panel de datos de 100 marcas diferentes en productos como bebidas, comida y aseo personal, se aplica una técnica de modelos VAR y VEC<sup>1</sup> dependiendo de la categoría trabajada con el fin de explorar los coeficientes y trabajar con las funciones de impulso respuesta resultantes para evaluar los impactos a corto y largo plazo promedio en las elasticidades de los productos comparando sus *brand equity*. Ya que se opta por dividir por productos, en comidas y bebidas el *brand equity* tiene un efecto a largo plazo significativo y positivo de .27 puntos de elasticidad, al igual que los productos de aseo personal con un efecto de .84 puntos. Los demás productos no tienen un efecto notorio a largo plazo, sin embargo, en el corto plazo sí bien en el siguiente período a la oferta aumentaba significativamente la elasticidad, en el período posterior la elasticidad disminuye o cae a valores negativos de la misma magnitud convergiendo al valor inicial.

Un caso de estudio hecho a marcas conocidas, es el realizado por Srinivasan y Bass (2000), ya que utilizan la marca de aderezos de sobremesa Heinz en diferentes categorías, así como otros productos de consumo alimenticio y de aseo en el hogar. En este enfoque se encargan de modelar el comportamiento de la participación de mercado, principalmente debido a que según estudios como el realizado por Dekimpe y Hanssens (1995), donde se hace una revisión a varios artículos centrados en el análisis de *marketing*, desde modelos autoregresivos hasta

---

<sup>1</sup>Acrónimo de *Vector Auto-Regressive* y *Vector Error Correction*, respectivamente

MMM,<sup>2</sup> con el fin de encontrar patrones relevantes para futuros análisis. Se encuentra que a pesar de que las ventas netas pueden evolucionar a través del tiempo, la participación de ventas de la categoría en mercados de bienes de consumo tiende a ser relativamente estacionaria. El documento se encarga de mostrar que si las ventas de la categoría están cointegradas con las ventas de la marca en cuestión, la participación de mercado se mantendrá estacionaria y establecer estrategias de respuesta entre demanda primaria (ventas de la categoría), demanda selectiva (ventas de la marca, en este caso Heinz) y demanda relativa (participación en el mercado) será prioritario ya que determinará el crecimiento a corto y largo plazo de las ventas. Para los últimos dos objetivos, se hace uso de estrategias de cointegración en diferentes períodos de tiempo y se promedian los resultados obtenidos para construir funciones de impulso-respuesta para la evaluación no sólo de choques a corto plazo sino para el estudio de las relaciones a largo plazo entre los diferentes tipos de demanda.

Los resultados indican que para los mercados de salsa de tomate, mantequilla de maní y azúcar se mantienen relaciones de equilibrio de largo plazo entre las ventas totales de la categoría y las ventas de Heinz. Este comportamiento también se nota en escenarios de subcategorías más pequeñas en constante evolución cómo es el caso de detergentes secos que impulsan las ventas de la categoría, sin embargo, dado que en este estudio no se tienen en cuenta factores relevantes para las pautas publicitarias, no se encuentra que haya una relación significativa entre los esfuerzos promocionales y las ventas a nivel de demanda selectiva, contrastando directamente con los hallazgos de Slotegraaf y Pauwels (2008).

Desde un punto de vista menos práctico y más teórico, una aproximación primaria a los modelos cointegrados estimados por medio de metodologías Bayesianas, es dada por el estudio de Kleibergen y Paap (2002), donde es posible notar que se tiene ya identificada una dificultad presentada en este documento; el muestreo de la matriz de multiplicadores de largo plazo de manera directa. En la aproximación abordada en este documento, se tiene que la matriz de multiplicadores también es descompuesta tal y como se abordará en la sección metodológica del presente documento, pero con una salvedad, ya que se utilizan dos matrices ortonormales y una matriz diagonal que contiene los valores singulares no negativos de la

---

<sup>2</sup>Sigla para *Marketing Mix Modelling*, metodología dónde se incluyen variables descriptivas de las ventas como distribución y precio, y variables de inversión en medios.



matriz de multiplicadores. Para el muestreo de las distribuciones posteriores, las distribuciones previas usadas son comunes en gran parte de la literatura: para la matriz de varianzas, se usa una distribución Wishart inversa, mientras que para los parámetros del modelo se utiliza una distribución normal multivariada. Esta aproximación, a pesar de ofrecer ciertas ventajas como la relativa facilidad a la hora de la computación de la distribución posterior, tiene complicaciones a la hora de enfrentarse a una metodología panel, debido a que no tiene en cuenta los posibles cambios entre los diferentes individuos para el muestreo conjunto de los parámetros; sin embargo, es un punto de partida valioso en caso de modelos para un solo individuo.

Para el caso univariado, los autores originales que escribieron el documento en el que se basa el presente proyecto Koop, Leon-Gonzalez y R. Strachan (2011), hacen uso del mismo proceso de descomposición de la matriz de multiplicadores de largo plazo usando matrices semi-ortogonales, que se relacionan con las matrices originales, diferenciándose del artículo de Kleibergen y Paap (2002) ya que estas matrices pueden ser relacionadas con las originales por medio de una matriz intermediaria que tiene ciertas características como que es definida positiva y además puede interactuar con los coeficientes de corto plazo. Ya que para este artículo se plantea la aplicación de la metodología para estudiar el efecto Fisher en la economía británica, que relaciona el ingreso, la tasa de desempleo y la inflación; y su utilidad en la capacidad de permitirle a los parámetros cierto movimiento a través del tiempo, permitiendo tener cierta flexibilidad en la parametrización del espacio de cointegración definido por las posibles combinaciones lineales entre las variables dependientes del sistema, de manera que dicho espacio puede caracterizarse por medio de una distribución de probabilidad que tiene la facultad de tener su punto óptimo en el momento anterior del tiempo, comportándose como una cadena de Markov donde el proceso descrito “no tiene memoria”, es decir, el valor actual sólo depende del valor previo. Esta técnica se aplica en el caso panel haciendo unos pequeños cambios, ya que las distribuciones previas siguen la misma forma funcional que en el caso presentado y el proceso de muestreo posterior tiene el paso añadido de diferenciar por individuos las matrices de multiplicadores de largo plazo, así como los conjuntos de parámetros determinísticos como tendencia o ciclo y parámetros

de corto plazo.

## 3. Marco teórico

A continuación se hará mención de cada uno de los temas relevantes para el presente proyecto, de manera que permita dar suficiente contexto sobre el trabajo que se llevará a cabo en la sección de metodología, donde se desarrollará un modelo de tipo panel cointegrado, estimado por medio de metodologías Bayesianas.

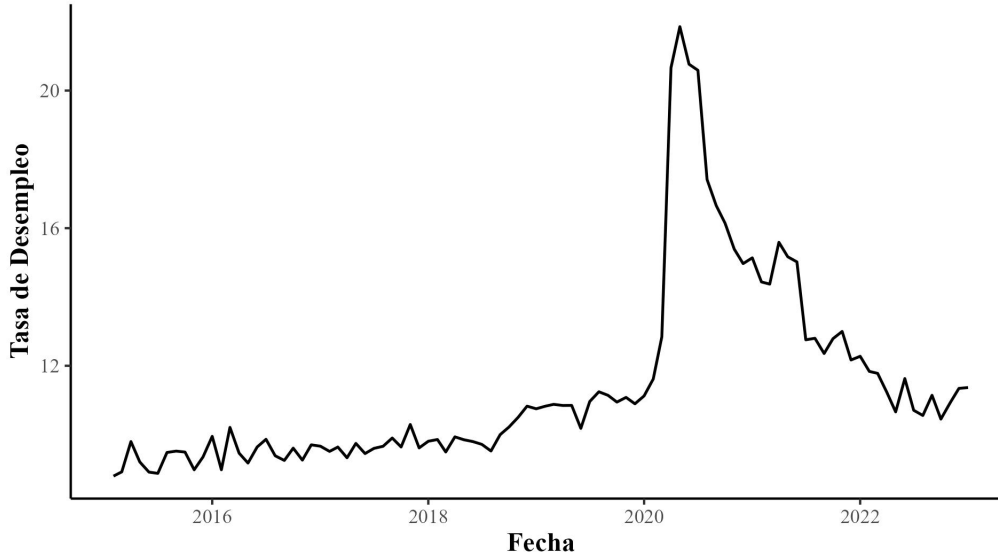
### 3.1. Serie temporal

Tal como se define en Brockwell y Davis (2006), una serie de tiempo es un conjunto de observaciones denotadas por medio de  $x_t$  tal que cada una de ellas es obtenida en un tiempo específico  $t$ . Éstas pueden ser de tipo discreto, donde se tienen  $T_0$  observaciones o de tipo continuo, donde las observaciones son obtenidas continuamente en un intervalo  $T_0 = [0, 1]$ . Usualmente puede usarse la notación  $x_t$  o  $x(t)$  indistintamente.

Un primer paso para la identificación de un modelo matemático adecuado para representar las series temporales, se basa en el uso de los procesos estocásticos, que tal como Brockwell y Davis (2006) definen como una familia de variables aleatorias  $\{X_t, t \in T\}$  definido en un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ ; sí el lector se encuentra interesado en aprender más sobre los espacios de probabilidad, puede revisarlo con mayor detalle y profundidad en Blanco, Arunachalam y Dharmaraja (2012). Las realizaciones de dicho proceso, denotadas como  $\{X(\omega), \omega \in \Omega\}$  tal que  $X_t(\cdot)$  es una función en el conjunto  $\Omega$  serían efectivamente la serie temporal en cuestión  $x_t$ .

Este tipo de datos puede ser encontrado en muchas áreas de estudio, desde ingeniería y economía, hasta biología y química. Algunos ejemplos de este tipo de datos pueden encontrarse en series usadas para la medición de indicadores de un país, como el nivel de desempleo en

un país.



**Figura 3-1.:** Tasa de Desempleo Enero 2015 - Enero 2023. **Fuente:** DANE.

La figura **3-1** representa el nivel de desempleo desestacionalizado de la economía colombiana desde 2015. Es posible ver que en 2020 se disparó el nivel de desempleo, principalmente como consecuencia de la pandemia y ha habido un ajuste gradual a niveles prepandemia.

## 3.2. Estacionariedad

En una serie de tiempo, es de especial interés la función de autocovarianza para obtener información del proceso en cuestión. Para una serie de tiempo según Brockwell y Davis (2006), se hará uso de las definiciones de función de autocovarianza y estacionariedad. La función de autocovarianza indica que sí un proceso  $\{X_t, t \in T\}$  tiene  $\text{Var}(X_t) < \infty$ , entonces la función de autocovarianza  $\gamma_x(\cdot, \cdot)$  de  $\{X_t\}$  es definida por:

$$\gamma_x(r, s) = \text{Cov}(X_r, X_s) = \text{E}[(X_r - \text{E}(X_r))(X_s - \text{E}(X_s))],$$

para  $r, s \in T$ . Con base en la función de autocovarianza, la estacionariedad de una serie temporal  $\{X_t, t \in \mathbb{Z}\}$  con un conjunto de índices  $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$  se define como:

1.  $\text{E}|X_t|^2 < \infty$  para todo  $t \in \mathbb{Z}$

2.  $E(X_t) = m$  para todo  $t \in \mathbb{Z}$
3.  $\gamma_x(r, s) = \gamma_x(r + t, s + t)$  para todo  $r, s, t \in \mathbb{Z}$

En esta definición de estacionariedad se hace referencia a esta en el sentido débil, ya que no se hace hincapié en las distribuciones conjuntas de  $(X_{t_1}, \dots, X_{t_k})^\top$  y  $(X_{t_1+h}, \dots, X_{t_k+h})^\top$ .

Para evaluar si una serie de tiempo es estacionaria o no debido a la presencia de una raíz unitaria en el polinomio autoregresivo, se emplea la prueba de Dickey Fuller, tal como se describe en el libro de Lütkepohl y Krätzig (2004), donde se busca identificar si una serie de tiempo  $x_t$  es integrada, esto es, en notación en términos del operador de rezagos:  $\alpha(1) = 1 - \alpha_1 - \dots - \alpha_p = 0$  y cada  $\alpha$  es un coeficiente fijo. En esta prueba, se evalúa si una serie de tiempo dada tiene una raíz unitaria, y en términos matemáticos, un proceso  $\{X_t, t \in T\}$  tiene raíz unitaria, si en la ecuación:

$$\Delta y_t = \phi y_{t-1} + \sum_{j=1}^{p-1} \alpha_j^* \Delta y_{t-j} + u_t,$$

$\phi$  resulta ser estadísticamente significativo, donde  $\phi$  representa a  $\alpha(1)$ . Para ello, se testea la hipótesis nula  $H_0 : \phi = 0$  contra  $H_1 : \phi < 0$  contrastada con el estadístico  $t$  de dicho coeficiente de manera que, la hipótesis nula indica que la serie es integrada o contiene raíz unitaria, mientras que la alternativa indica estacionariedad. Cabe recordar que la distribución de la estadística  $t$  no es estándar y se ajusta de acuerdo a los valores de una tabla dada.

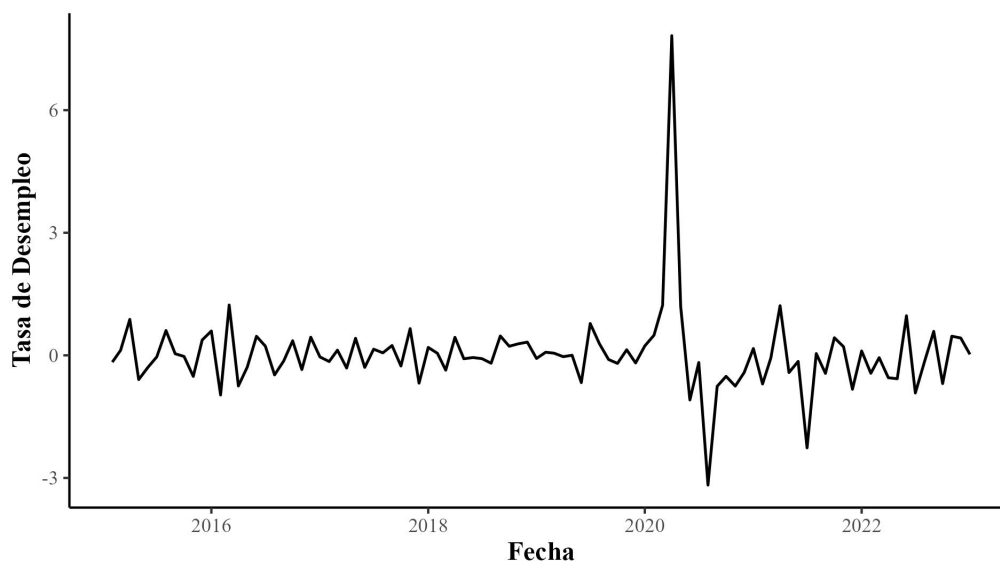
Para ejemplificarlo mejor, retomando con la serie de desempleo de Colombia, se evalúa si dicha serie estacionaria por medio del uso del test de Dickey-Fuller en el software R, particularmente en el paquete aTSA. Los resultados son presentados en el cuadro **3-1**.

Rezago	ADF	P-Valor
<b>0</b>	-0.776	0.401
<b>1</b>	-0.931	0.346
<b>2</b>	-0.986	0.327
<b>3</b>	-0.702	0.428
<b>4</b>	-0.773	0.402

**Tabla 3-1.:** Resultados ADF Test - Serie de desempleo. Elaboración propia.

Los resultados muestran que la serie no es estacionaria, es decir, tiene raíz unitaria ya que

no se rechaza la  $H_0$  a un nivel de confianza de 5 %. Sí diferenciamos la serie, y aplicamos el test de Dickey-Fuller se presentan en la figura y en el cuadro **3-2**.



**Figura 3-2.:** Tasa de Desempleo Enero 2015 - Enero 2023, diferenciada. **Fuente:** DANE.

Rezago	ADF	P-Valor
<b>0</b>	-8.29	0.01
<b>1</b>	-6.19	0.01
<b>2</b>	-5.19	0.01
<b>3</b>	-6.07	0.01

**Tabla 3-2.:** Resultados ADF Test - Serie de desempleo, diferenciada. Elaboración propia.

Cuando se realiza la prueba ADF sobre la serie diferenciada, se rechaza que exista raíz unitaria y por ende, se dice que no hay evidencia estadística de la presencia de una raíz unitaria, a pesar del pico de 2020. Para el trabajo que se realizará en este proyecto, es de especial interés las series que son no estacionarias en nivel pero que una vez diferenciadas, son estacionarias (conocidas como integradas de orden 1 o  $I(1)$ ).

### 3.3. Cointegración

Para el presente trabajo, se usarán series que no son estacionarias en nivel, pero que son estacionarias cuando se toma su primera diferencia, es decir, son variables  $I(1)$ . También se

buscará si existe una relación de equilibrio entre dichas variables. Esta aproximación se dio por vez primera en 1987 en Engle y Granger (1987), donde se relacionan variables económicas que usualmente no son estacionarias en nivel, evitando caer en relaciones espurias<sup>1</sup>. La definición propuesta se centra en los componentes de un vector  $x_t$  que se dice son cointegrados de orden  $d, b$ ; con  $d$  orden de integración y  $0 < b \leq d$ , indicando  $x_t \sim CI(d, b)$  si:

1. Todos los componentes de  $x_t$  son  $I(d)$  con  $x_t = (x_{1t}, \dots, x_{kt})^\top$ .
2. Existe un vector  $\alpha \neq 0$  tal que  $z_t = \alpha^\top x_t \sim I(d - b), b > 0$ .  $\alpha$  es el vector de cointegración.

Sí en un dado caso, tomamos como restricción  $d = b = 1$ , las relaciones de cointegración identificadas por  $z_t$  serían  $I(0)$  que resulta conveniente para centrarse en  $\alpha$  ya que contiene las relaciones de equilibrio de las variables analizadas.

En un caso donde se analicen  $n$  variables o componentes que pertenecen a  $x_t$ , puede existir más de una relación de cointegración que puede ser representada por medio de  $r$ . Entonces,  $r$  contiene a lo sumo  $n - 1$  relaciones de cointegración linealmente independientes, reunidas en el arreglo  $\alpha$  de tamaño  $n \times r$  cuyo rango sería efectivamente  $r$ .

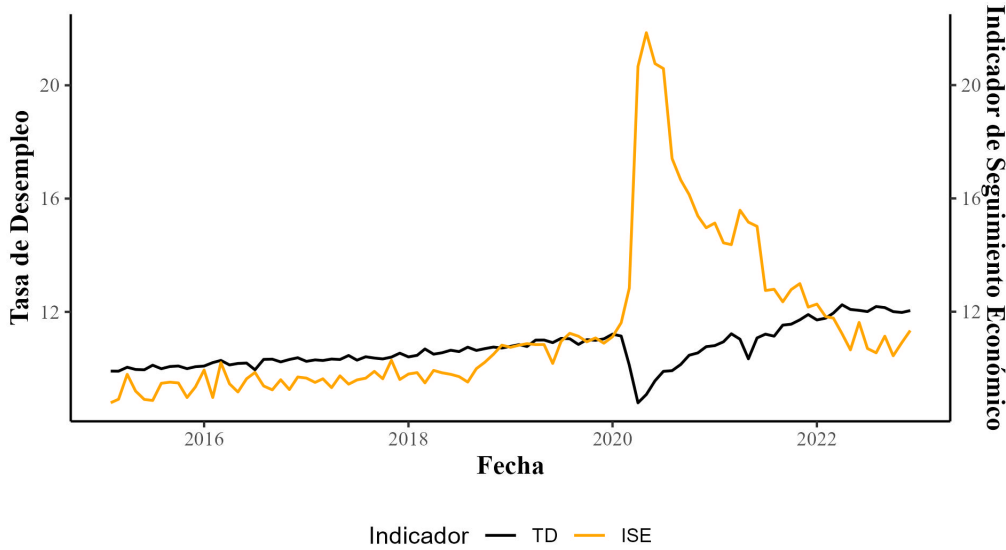
Para la identificación del orden de cointegración en el sistema, los autores proponen el uso de una regresión usual y una evaluación del término de error de manera que se pueda establecer sí el error es efectivamente estacionario o  $I(0)$ . Se proponen usualmente los test de Dickey Fuller vistos en el numeral anterior.

Para ejemplificar el concepto, se retoma el uso de la tasa de desempleo colombiana junto con el indicador de seguimiento económico (ISE) desestacionalizado, ya que este tiene la misma periodicidad de la publicación de desempleo y está conformado por una aproximación previa a la publicación del PIB trimestral. En la figura **3-3** se ilustra el comportamiento de ambas series.

Para identificar sí efectivamente las variables  $I(d)$  al ser relacionadas resultan en un proceso  $I(d)$  tal como se menciona en Engle y Granger (1987), procederemos a resolver por mínimos

---

<sup>1</sup>Se dice que dos variables tienen una relación espuria si existe una conexión aparente a priori pero una vez se examina más a fondo, la conexión sugiere que se trata de una casualidad.



**Figura 3-3.:** Tasa de Desempleo vs ISE Enero 2015 - Enero 2023. **Fuente:** DANE

cuadrados la relación entre ambas variables:

$$z_t = \alpha_1 TD_t + \alpha_2 ISE_t + \varepsilon_t,$$

dando como resultado  $\alpha_1 = 1, \alpha_2 = 1,058$  con un  $R^2$  de 93,6. Evaluando los residuales, el valor de la prueba de Dickey Fuller aumentada está consignada en el cuadro **3-3**:

Rezago	ADF	P-Valor
<b>0</b>	-1.99	0.046
<b>1</b>	-2.58	0.010
<b>2</b>	<b>-2,47</b>	<b>0,015</b>
<b>3</b>	-2.51	0.013

**Tabla 3-3.:** Resultados ADF Test - Residuales Tasa de Desempleo vs. ISE. Elaboración propia.

Cuando se examina el criterio de Tsay, el AIC sugiere que el mejor número de rezagos es 2 con un p-valor de 0.39 sugiriendo que el verdadero proceso es de tipo AR. Con esto, se muestra que existe una relación de cointegración o equilibrio entre la tasa de desempleo y el ISE, de manera que el error resultante de una combinación lineal entre ambas variables es estacionario.



### 3.4. Modelos multiecuacionales en series de tiempo

Cuando existe un interés particular en estudiar las relaciones entre varias variables dentro de un mismo sistema con sus interacciones y cambios, se usan modelos con varias variables dependientes. Lütkepohl (2005) en su libro, explica la notación básica de una serie temporal múltiple, que puede ser expresada por medio de la función de pronóstico como:

$$\hat{y}_{k,T} = f_k(y_{1,1}, \dots, y_{1,T}, \dots, y_{K,T}),$$

dónde la colección de series  $k = 1, \dots, K$  desde la observación  $t = 1, \dots, T$  comprende un conjunto de series de tiempo multivariadas. En este contexto, cada una de las  $k$  series se inter-relaciona con las demás  $k_{-i}$  variables del sistema teniendo en cuenta sus rezagos para la predicción de nuevos valores de las series.

Para una serie fijada  $k = \hat{k}$ , se tiene que su pronóstico  $h$  pasos adelante  $y_{1,T+h}$  sigue la forma:

$$\hat{y}_{k=\hat{k},T} = f_{k=\hat{k}}(y_{1,T}, \dots, y_{K,T}, y_{1,T-1}, \dots, y_{K,T-1}, y_{1,T-2}, \dots),$$

de manera que para identificar la mejor predicción  $h$  pasos adelante para un conjunto de series  $K$  dado, es necesario identificar el conjunto de funciones  $f_1, \dots, f_k$  teniendo como analogía el caso univariado, que a pesar de ser el caso más general, a veces no es el más viable.

Para estos modelos, las variables del conjunto deben ser estacionarias para poder aplicar supuestos sobre el término de error que aseguren la calidad del modelo y minimizar el riesgo de encontrar sólo relaciones espurias. Para incorporar variables no estacionarias, se incluye la variable que recoge las diferentes relaciones de cointegración  $z_t$  para la predicción de las primeras diferencias de las variables del sistema, asegurando así que todas las variables sean estacionarias.

### 3.5. Datos longitudinales

Los datos longitudinales son útiles en la medición de fenómenos a través del tiempo cuando se mantienen los mismos individuos a través de la ventana de observación; concretamente, Diggle et al. (2002), lo define como el estudio basado en individuos que son medidos repetidamente a lo largo del tiempo; y a diferencia con los datos de tipo transversal, los datos longitudinales permiten estudiar la evolución de los individuos a través del tiempo. Este tipo de datos es particularmente útil en las ciencias sociales y económicas, ya que varias variables de relativo interés se toman a partir de las muestras temporales de varios individuos. Cabe resaltar que el objetivo principal del análisis de este tipo de datos es evaluar tanto la variación entre los individuos como a través del tiempo en efectos que varíen entre individuos pero no a través del tiempo (efectos fijos), y variaciones a través del tiempo pero no entre individuos (efectos aleatorios).

A pesar de que algunos de los estudios pueden usarse para datos longitudinales y transversales, lo cierto es que existen diferencias importantes cómo la existencia de correlación entre individuos o temporal entre las observaciones del mismo individuo. Por lo que, de entrada, los errores de este tipo de modelos se asumen correlacionados o no independientes. Un primer modelo general para este tipo de datos, se basa en la siguiente ecuación donde se representa un fenómeno  $y_{i,t}$  con  $i = 1, \dots, N$  individuos evaluados durante el período  $t = 1, \dots, T$  con  $p$  variables explicativas:

$$Y_{i,t} = \beta_1 x_{i,t1} + \dots + \beta_p x_{i,tp} + \epsilon_{i,t},$$

con  $\epsilon_{i,j}$  una secuencia aleatoria de tamaño  $T$  asociada a  $N$  individuos. En un modelo normal, es posible asumir que  $\epsilon_{i,j}$  es mutuamente independiente entre individuos y períodos observados con distribución  $N(0, \sigma^2)$ ; pero en este enfoque, se entenderá que existe una estructura de correlación tal que pueda tener en cuenta las interdependencias entre individuos, de manera que:

$$Y \mid \boldsymbol{\beta}, \sigma^2 \sim N_{(NT) \times 1}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{V}),$$

donde  $\boldsymbol{\beta}$  refiere a los parámetros del modelo,  $\mathbf{X}$  refiere a la matriz de diseño del modelo y  $\mathbf{V}$  refiere a la estructura de covarianza de las observaciones  $Y$ , de tal manera que para la

estimación de los parámetros  $\beta$  se utiliza la metodología de mínimos cuadrados ponderados. Para la caracterización de datos longitudinales o bien llamados *panel* se puede hacer uso de modelos de efectos fijos, aleatorios o mixtos.

Para los modelos de efectos fijos, se asume que para cada individuo  $i = 1, \dots, N$  existen efectos inherentes a dicho individuo que son constantes a través del tiempo y se denotan por medio de un subíndice independiente de la unidad de tiempo en la que sea observada. Para los modelos de efectos aleatorios, se asume que para toda unidad temporal de medida  $t = 1, \dots, T$  existen fenómenos que son transversales a todos los individuos de la muestra y varían con respecto al momento en que son medidos. Finalmente, para los modelos de efectos mixtos, se usan simultáneamente efectos aleatorios y efectos fijos, donde se tiene sospecha de que un individuo  $i = 1, \dots, N$  tiene efectos inherentes a sí mismo y efectos compartidos con los demás individuos que varían a través de la ventana de observación  $t = 1, \dots, T$ .

## 3.6. Metodología Bayesiana

Para la estimación del modelo en el documento presentado, se hará uso de inferencia Bayesiana. Tal como está consignado en Gelman et al. (2004), un modelo Bayesiano se rige por la regla de Bayes, que consiste realizar inferencia sobre un conjunto de parámetros  $\theta = \{\theta_1, \dots, \theta_j\} \in \Theta$  dado un conjunto de datos o información  $\mathbf{y}$  que está determinado por  $\theta$ . Para ello, se hace uso de la función de verosimilitud de los datos o distribución muestral dada por  $p(\mathbf{y} | \theta)$  y una distribución previa de los parámetros dada por  $p(\theta)$ , tal que:

$$p(\theta | \mathbf{y}) = p(\mathbf{y} | \theta) p(\theta) \propto p(\mathbf{y} | \theta) p(\theta),$$

usualmente existe una dificultad en la proposición de  $p(\theta)$  ya que requiere de información previa de parte del investigador. La predicción en este tipo de modelos, se basa en estimar el valor esperado de ciertas funciones  $g(\cdot)$  del vector de parámetros  $\theta$  de manera que la expresión se obtiene usando:

$$E(g(\theta | \mathbf{y})) = \int_{\Theta} g(\theta) p(\theta | \mathbf{y}) d\theta,$$

bajo la función pérdida cuadrática  $E((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{Bayes})^2)$  y debido a que usualmente  $p(\boldsymbol{\theta} | \mathbf{y})$  no tiene una forma funcional definida se utilizan métodos de aproximación tales como cadenas de Markov, en particular, el algoritmo MCMC. De manera que sí se tiene un vector  $\boldsymbol{\theta}$  muestreado un número  $B$  (suficientemente grande) de veces  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(B)}$  de la distribución posterior  $p(\boldsymbol{\theta} | \mathbf{y})$ , la integral anterior o valor esperado de  $g(\cdot)$  se puede aproximar como:

$$E(g(\boldsymbol{\theta} | \mathbf{y})) \approx \frac{1}{B} \sum_{b=1}^B g(\boldsymbol{\theta}^{(b)}),$$

la ventaja de esta metodología es que permite de manera sencilla calcular diferentes funciones  $g(\cdot)$  según se necesite de la distribución posterior, tales como la mediana, varianza, cuartiles, etc. Esto puede consultarse con mayor detalle y profundidad en Hoff (2009). La manera en la que se suele estimar la distribución posterior es usando algoritmos de muestreo. El primero usado en este trabajo es el algoritmo de muestreo de Gibbs (*Gibbs Sampler* en inglés), el cual consiste en muestrear alternadamente subvectores de  $\boldsymbol{\theta}$  tal que  $(\theta_1, \dots, \theta_d)$ , como lo explican a detalle Casella y George (1992). Cada iteración del algoritmo se basa en tomar algún valor  $d$  y muestrearlo con base en los demás  $-d$  valores en un total de  $B$  iteraciones. Entonces por cada iteración  $t = 1, \dots, B$  se tienen  $d$  iteraciones. Las iteraciones se ejecutan de manera que:

$$p(\theta_j | \boldsymbol{\theta}_{-j}^{t-1}, \mathbf{y}),$$

donde  $\boldsymbol{\theta}_{-j}^{t-1}$  viene a representar todos los componentes de  $\boldsymbol{\theta}$  menos  $\theta_j$  en sus valores actuales con  $j = 1, \dots, d$ :

$$\boldsymbol{\theta}_{-j}^{t-1} = (\theta_1^t, \dots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \dots, \theta_d^{t-1}),$$

de manera que cada subvector  $\theta_j$  es actualizado condicional con base en los últimos valores de los demás componentes de  $\boldsymbol{\theta}$ . La limitación principal del algoritmo se basa en que sí  $p(\theta_j | \boldsymbol{\theta}_{-j}^{t-1})$  no tiene una forma funcional cerrada, no es posible muestrear los valores.

El segundo método utilizado es el algoritmo de Metropolis - Hastings que se basa en el algoritmo de Metropolis cambiando la tasa de aceptación  $r$ . Este se encarga de solventar el problema de que  $p(\theta_j | \boldsymbol{\theta}_{-j}^{t-1})$  no tenga una forma funcional cerrada, ya que pondera por los valores de una distribución arbitraria  $J_t$  dada.

El algoritmo se desarrolla en los siguientes pasos:

1. Muestrar  $\theta^0$  de una distribución inicial  $p_0(\theta)$
2. Para cada  $t = 1, \dots, B$ :
  - a) Muestrear una propuesta,  $\theta^*$  de una distribución propuesta en el paso  $t$ ,  $J_t(\theta^* | \theta^{t-1})$ .
  - b) Calcular la razón de las densidades

$$r = \frac{p(\theta^* | \mathbf{y})/J_t(\theta^* | \theta^{t-1})}{p(\theta^{t-1} | \mathbf{y})/J_t(\theta^{t-1} | \theta^*)}$$

- c) Establecer

$$\theta^t = \begin{cases} \theta^* & \text{Con probabilidad } \min\{r, 1\} \\ \theta^{t-1} & \text{En otro caso} \end{cases}$$

Este algoritmo permite que  $J_t(a | b)$  no deba ser estrictamente simétrica, aumentando el rango de posibilidades para la escogencia de  $J_t$ . Para la predicción de valores no observados, se usa la lógica interna de la regla de Bayes, tal que para una variable no observada  $\tilde{y}$  se tiene:

$$p(\tilde{y} | \mathbf{y}) = \int p(\tilde{y} | \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta},$$

de la misma manera, la forma de evaluar la calidad de ajuste del modelo dado un arreglo de parámetros  $(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(B)})$  se genera el vector  $(\hat{\mathbf{y}}^{(1)}, \dots, \hat{\mathbf{y}}^{(B)})$  con  $\hat{\mathbf{y}}^{(b)} \sim p(\hat{\mathbf{y}}^{(b)} | \mathbf{y})$  y

$$p(\hat{\mathbf{y}} | \mathbf{y}) = \int p(\hat{\mathbf{y}} | \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta},$$

de manera que sí se evalúa un estadístico  $T(\mathbf{y})$  de los datos observados para cada uno de los  $\hat{\mathbf{y}}^{(b)}$  tal que se obtenga un arreglo  $(T(\hat{\mathbf{y}}^{(1)}), \dots, T(\hat{\mathbf{y}}^{(B)}))$  y se compare con los valores de  $T(\mathbf{y})$  de los datos observados se evalúa sí estos últimos son valores extremos de la distribución de  $(T(\hat{\mathbf{y}}^{(1)}), \dots, T(\hat{\mathbf{y}}^{(B)}))$  ya que sí ocurre esto, el modelo no es capaz de replicar la característica deseada de los datos observados satisfactoriamente.

Finalmente, para evaluar el ajuste de un modelo Bayesiano, se evalúa qué tan extrema puede ser una *réplica* de los datos observados. Tal como es consignado en Gelman et al. (2004), para el análisis, se define entonces  $y^{rep}$  como datos que “pudo” ser observada con los mismos valores  $\theta$  que produjeron a  $y$  en un principio, de manera que se establece:

$$p(y^{rep} | y) = \int p(y^{rep} | \theta) p(\theta | y) d\theta,$$

como la distribución de  $y^{rep}$  dado el estado actual de conocimiento o distribución predictiva posterior. Con estos valores, es posible computar los p-valores predictivos posteriores o valores PPP, comparando los datos observados con la distribución predictiva posterior usando el valor-p Bayesiano, definido como la probabilidad que los datos replicados sean más extremos que los datos observados medida por el valor:

$$p_B = p(T(y^{rep}, \theta) \geq T(y, \theta) | y),$$

haciendo uso de la distribución posterior de  $\theta$  y la distribución predictiva posterior de  $y^{rep}$  de manera que se unen los conceptos vistos anteriormente para poder estimar la probabilidad de que un valor replicado con los valores de la distribución posterior de  $\theta$  sean mayores que los datos observados. Naturalmente se espera que  $p_B$  sea un valor cercano a 0.5 si el modelo puede predecir correctamente la datos observados  $y$ .

En el presente trabajo se usarán cuatro criterios de información con el fin de seleccionar el número adecuado de rezagos del modelo aplicado en los datos de *marketing*, en el ejercicio aplicado a datos reales con el fin de encontrar cuál es el modelo que mejor representa a los datos. El primero en revisarse será el AIC o *Akaike Information Criterion*, definido en su versión simple como:

$$AIC = 2k - 2p(y | \theta),$$

donde  $k$  es el número de parámetros en el modelo y  $p(y | \theta)$  es el valor de la verosimilitud del modelo estimado. El siguiente criterio de información es el BIC o *Bayesian Information*

*Criterion*, también conocido como *Schwarz Information Criterion* y se define como sigue:

$$\text{BIC} = k \log(n) - 2 \log p(\mathbf{y} | \boldsymbol{\theta}),$$

y nuevamente se hace uso de la verosimilitud del modelo  $p(y | \boldsymbol{\theta})$  y el número de parámetros  $k$  añadiendo un conteo de datos de los datos observados  $n$ . Estos criterios y variaciones de estos criterios pueden ser encontrados en Lütkepohl y Krätzig (2004). Para continuar, se hará uso de otros dos criterios de información más Bayesianos en su construcción.

El primero de estos criterios es el DIC o *Deviance Information Criterion*, que tal como se define en Gelman et al. (2004) se construye de la siguiente manera:

$$\text{DIC} = \hat{D}_{avg}^{pred}(\mathbf{y}) = 2\hat{D}_{avg}(\mathbf{y}) - D_{\boldsymbol{\theta}}(\mathbf{y}),$$

donde  $D$  hace referencia a una medida de discrepancia, y las partes que componen la ecuación del DIC son:

$$\hat{D}_{avg}(\mathbf{y}) = \frac{1}{B} \sum_{b=1}^B D(\mathbf{y}, \boldsymbol{\theta}^b),$$

y  $B$  representa la  $b = 1, \dots, B$  simulación, por lo que esta primera parte puede interpretarse cómo el promedio de la medida de discrepancia luego de las  $B$  simulaciones, y seguidamente, la segunda parte del cálculo de DIC se define como:

$$D_{\boldsymbol{\theta}}(\mathbf{y}) = D(\mathbf{y}, \hat{\boldsymbol{\theta}}(\mathbf{y})),$$

y se tiene que  $\hat{\boldsymbol{\theta}}(\mathbf{y})$  es un estimador puntual de  $\boldsymbol{\theta}$  usando las simulaciones obtenidas a partir de la distribución posterior, que puede ser el promedio, media o mediana dependiendo de la función de pérdida escogida; en el presente trabajo se hace uso del promedio debido a la escogencia de la función de pérdida cuadrática y a su facilidad de cálculo en las matrices de las cadenas de Markov resultantes. Así mismo, la función que se utiliza como medidora de la discrepancia es la desviación, definida como:

$$D(\mathbf{y}, \boldsymbol{\theta}) = -2 \log p(\mathbf{y} | \boldsymbol{\theta}).$$

Finalmente se hará uso del WAIC o *Widely Applicable Information Criterion* también conocido como *Watanabe-Akaike Information Criterion* y que tal como define su autor en Watanabe (2013), se tiene que:

$$\text{WAIC} = T_n + V_n/n,$$

que guarda ciertas similitudes con el DIC, que se verán a continuación cuando se estudia más a fondo que significan sus componentes. En primer lugar, se tiene  $T_n$  que se define como:

$$T_n = -\frac{1}{n} \sum_{i=1}^n \log p(\mathbf{y}_i | \boldsymbol{\theta}),$$

donde es posible ver que guarda una estrecha relación con el término  $\hat{D}_{avg}(\mathbf{y})$  del DIC, ya que se calculan prácticamente de la misma manera, salvo que las  $\mathbf{y}_i$  son de diferentes muestras a comparar, siendo usado para validación cruzada. Por otro lado, se define  $V_n$  como:

$$V_n = \sum_{i=1}^n \{E_{\boldsymbol{\theta}}[(\log p(\mathbf{y}_i | \boldsymbol{\theta}))^2] - E_{\boldsymbol{\theta}}[\log p(\mathbf{y}_i | \boldsymbol{\theta})]^2\},$$

y es importante aclarar que  $E_{\boldsymbol{\theta}}$  puede tomarse como una función de un parámetro, que para efectos de este documento, se usó como el promedio.

Sí el lector se encuentra interesado en ver más estudios como este o en profundizar en más temas de estadística Bayesiana, se recomiendan los textos de Gelman et al. (2004), Hoff (2009), Gamerman y Lopes (2006) o un texto de inferencia Bayesiana de su preferencia.



## 4. Metodología

Dado que la metodología se aplicará en series de tiempo, vale la pena tener en cuenta ciertas distinciones con respecto a los modelos lineales usuales. El foco del trabajo se basará en encontrar un modelo probabilístico que permita relacionar de manera efectiva las variables indexadas por el tiempo contenidas en el grupo  $\mathbf{Y}$  a través de los individuos por medio de la combinación de un modelo VEC y metodologías panel. Debido al tratamiento especial que debe ser llevado a cabo para las relaciones de cointegración presentes en el modelo, desde la identificación hasta el muestreo posterior de las distribuciones condicionales para cada uno de los parámetros, es necesario hacer distinciones entre los procesos de desarrollo y muestreo de cada conjunto de parámetros. Esta metodología puede encontrarse con más detalle en Koop, Leon-Gonzalez y R. Strachan (2008). En este trabajo, en particular, se modifican las ecuaciones (4-11) y (4-12) debido a que el trabajo original usa la misma previa a lo largo de toda la estimación, mientras que en esta modificación se permite actualizar la previa con cada iteración, lo que le proporciona al modelo mayor flexibilidad en la estimación de los parámetros de cointegración. Esta modificación tiene dos consecuencias fundamentales: la primera, plantear previas no informativas (difusas) tanto para los parámetros de cointegración como para los de corto plazo; y la segunda, desarrollar una nueva estrategia computacional con el fin de llevar a cabo todos los procesos de estimación e inferencia.

Cabe resaltar que a pesar de que las matrices de varianzas y covarianzas se actualizan en las ecuaciones (4-11) y (4-12), no se habla de un método de Monte Carlo adaptativo, ya que es una propuesta hecha con base en el establecimiento de previas jerárquicas multinivel, tal como se verá en el DAG más adelante.

## 4.1. Modelo Panel Cointegrado Vectorial

El modelo contemplado en este trabajo, se encarga de combinar las metodologías de series de tiempo cointegradas multivariantes y datos de tipo panel, donde para un cierto individuo  $1, \dots, N$  se tiene una serie de variables  $1, \dots, n$  que componen un sistema particular para dicho individuo que describe sus relaciones de cointegración.

El planteamiento del modelo es el siguiente:

$$\Delta y_{i,t} = \Pi_i y_{i,t-1} + \sum_{h=1}^L \Gamma_{i,h} \Delta y_{i,t-h} + \Phi_i d_{i,t} + \epsilon_{i,t}, \quad (4-1)$$

donde se definen:

- $\Delta y_{i,t}$  la primera diferencia de cada una de las  $n$  variables dependientes del sistema, de cada uno de los  $i = 1, \dots, N$  individuos variando desde  $t = 1, \dots, T$ , dando como resultado un vector de tamaño  $(TNn \times 1)$ . El vector  $y_{i,t-1}$  tiene la misma dimensión pero contiene las variables en nivel del sistema, es decir, sin primeras diferencias. Los vectores pueden ser escritos como  $\Delta y_{i,t} = (\Delta y_{1,i,t}, \Delta y_{2,i,t}, \dots, \Delta y_{n,i,t})^T$  y  $y_{i,t-1} = (y_{1,i,t-1}, y_{2,i,t-1}, \dots, y_{n,i,t-1})^T$ .
- $\Pi_i$  es el resultado de la multiplicación de dos matrices de rango completo tal que  $\Pi_i = \alpha_i \times \beta_i^T$ :
  - $\beta_i$  que contiene los coeficientes de las relaciones de cointegración para cada individuo, de tamaño  $(n \times r_i)$  con  $r_i$  rango de cointegración para cada individuo  $i$ . Esto está íntimamente relacionado con los espacios de cointegración que serán denotados por  $\mathfrak{p}_i = sp(\beta)$ . Un ejemplo sería donde se tienen  $i = 1, \dots, N$  individuos con  $n = 3$  variables independientes  $(a, b, c)$  con  $r = 2$  relaciones de cointegración de manera que el espacio de cointegración para los individuos es:

$$\mathfrak{p}_i = sp \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix},$$

que denota las posibles combinaciones lineales de  $(a, b, c)$ .

- $\alpha_i$  que contiene los coeficientes que concatenan las relaciones de largo plazo con las relaciones de corto plazo  $(n \times r_i)$ .

De manera que  $\Pi_i$  es de tamaño  $(n \times n)$ . Otra manera de conceptualizar esta matriz es a través de las relaciones de largo plazo de las variables o de cointegración. Se observa que cada  $r_i$  no hace parte del conjunto de parámetros por estimar, y por lo tanto se considera como un hiperparámetro adicional que debe ser establecido de antemano por el investigador, basado en las pruebas que considere necesarias para hallar el rango de cointegración de cada individuo.

- $\Gamma_{i,h}$  matriz de tamaño  $(n \times n)$  que contiene las relaciones de corto plazo de las variables del modelo, y con estas matrices pueden asociarse los efectos de causalidad de Granger. El número de matrices depende de la cantidad de rezagos utilizados en el proceso de modelado. Estas matrices cambian con los individuos y el rezago asociados, sin embargo, se asume el número de rezagos común para todos los individuos.
- $\Phi_i$  matriz de tamaño  $(n \times k)$  con  $k$  número de términos determinísticos usados, en este caso hacen referencia al uso de tendencias, ciclos o variables dicotómicas que se requieran en el modelo.
- $\epsilon_{i,t}$  el término de error asociado.

Este modelo permite que cada uno de los individuos tenga un número diferenciado de relaciones de cointegración con tal de que este sea menor o igual que el rango de cointegración total del modelo  $r_i \leq r$ , permitiendo capturar efectos diferenciadores entre los individuos del panel.

Para el error, un supuesto para las covarianzas es que están presentes sólo a través de individuos de manera contemporánea, es decir:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \Sigma_{ij} & \text{para } t = s; \\ 0 & \text{en otro caso} \end{cases},$$

sin embargo, es posible probar otras estructuras de autocorrelación en los errores dado que se trabaja con series temporales pero este supuesto facilita notablemente el proceso de estimación ya que permite simplificar la matriz de varianzas y covarianzas. Para este trabajo, se usará la versión simplificada de la matriz de covarianzas, debido a que es una primera aproximación pero se propone para siguientes trabajos, implementar una estructura de covarianzas que refleje las autocorrelaciones presentes en las series de tiempo individuales. Los parámetros que se estiman en este modelo son  $\mathbf{\Pi}_i$ ,  $\mathbf{\Gamma}_{i,h}$ ,  $\mathbf{\Phi}_i$ , mientras que los valores que se deben establecer a priori son los  $r_i$ .

## 4.2. Representaciones de la verosimilitud

Debido a los problemas mencionados previamente para la identificación de la matriz de multiplicadores de largo plazo en trabajos como el de Kleibergen y Paap (2002), la matriz  $\mathbf{\Pi}_i$  debe identificarse de alguna manera práctica, por lo que en esta sección se dan las pautas principales para identificar las matrices importantes con las cuales se trabajará para la estimación de los parámetros del modelo.

### 4.2.1. Reparametrizaciones para la matriz de multiplicadores de largo plazo.

Existen varias maneras de representar la verosimilitud para este modelo, pero principalmente se mostrará cómo reparametrizar la matriz de multiplicadores de largo plazo,  $\mathbf{\Pi}_i$  en términos de  $\beta_i$ ,  $\alpha_i$  para encontrar representaciones que permiten muestrear de manera más sencilla.

La matriz  $\mathbf{\Pi}_i$  puede ser escrita cómo:

$$\mathbf{\Pi}_i = \beta_i \alpha_i^\top = [\beta_i \kappa_i] [\alpha_i \kappa_i^{-1}]^\top = \beta_i^* \mathbf{A}_i^\top, \quad (4-2)$$

donde  $\beta_i$  está restringida a ser semi-ortogonal y  $\kappa_i$  debe ser definida positiva de manera que  $\alpha_i \kappa_i^{-1}$  sea definida positiva. Para establecer un punto de partida que satisfaga las condiciones sobre  $\kappa_i$  se pueden escoger:

$$\boldsymbol{\kappa}_i = (\boldsymbol{\alpha}_i^\top \boldsymbol{\alpha}_i)^{1/2} = (\boldsymbol{\beta}_i^{*\top} \boldsymbol{\beta}_i^*)^{1/2},$$

y a medida que avance el documento, se deberán cambiar las representaciones de (4-2) por conveniencia en el muestreo.

### 4.2.2. Re-escribiendo el término de error

Los errores, por su parte, deben ser re planteados totalmente para una mayor conveniencia en su muestreo. Partiendo de los  $\epsilon_{i,t}$  procedemos:

- Convertir  $\epsilon_{i,t}$  en matrices  $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \dots, \epsilon_{i,T})^\top$  de tamaño  $(T \times n)$  con  $i = 1, \dots, N$ .
- Convertir las  $N$  matrices  $\boldsymbol{\epsilon}_i$  en una matriz  $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N)^\top$  de tamaño  $(T \times Nn)$
- Aplicando vectorización a la matriz  $\boldsymbol{\epsilon}$  tenemos  $\mathbf{e} = \text{vec}(\boldsymbol{\epsilon})$

El vector  $\mathbf{e}$  tiene cómo matriz de covarianza:

$$E(\mathbf{e}\mathbf{e}^\top) = \mathbf{V}_e = (\boldsymbol{\Sigma} \otimes \mathbf{I}_T). \quad (4-3)$$

Sí el lector está interesado en relajar el supuesto sobre  $\boldsymbol{\Sigma}$ , se puede incorporar un nuevo parámetro  $\rho \in [-1, 1]$  que identifique adecuadamente la estructura de autocovarianza de primer orden de las series de tiempo modificando la estructura de  $\mathbf{I}_T$  de manera que no se trate de una matriz identidad, sino de una matriz con la siguiente estructura:

$$\mathbf{F}_\rho = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}$$

En este caso, con fines de facilitar la estimación del modelo, se mantendrá el supuesto sobre  $\boldsymbol{\Sigma}$ .

### 4.2.3. Primera aproximación a la verosimilitud

Dadas las bases de las matrices más relevantes,  $\mathbf{\Pi}_i, \boldsymbol{\epsilon}_i$ , se procede a generar una proporcionalidad a la primera función de densidad para la caracterización de los errores:

$$|\boldsymbol{\Sigma}|^{-T/2} \exp \left\{ -\frac{1}{2} \mathbf{e}^\top (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T) \mathbf{e} \right\} = |\boldsymbol{\Sigma}|^{-T/2} \exp \left\{ -\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}) \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} \right\}.$$

### 4.2.4. Segunda aproximación a la verosimilitud

Sí se reparametriza la ecuación (4-1) de manera que se defina  $z_{i,t} = \boldsymbol{\beta}_i^\top y_{i,t-1}$ , el vector  $1 \times (k + r_i)$ ,  $X_{i,t} = (z_{i,t}^\top, \Delta y_{i,t-1}^\top, \dots, \Delta y_{i,t-l}^\top, d_t^\top)$  con  $k$  número de términos determinísticos (dimensión de  $d_t$ ) más  $n$  veces el número de rezagos  $l$  (común para todos los individuos), y la matriz  $\mathbf{B}_i = (\boldsymbol{\alpha}_i, \boldsymbol{\Gamma}_{i,1}, \dots, \boldsymbol{\Gamma}_{i,l}, \boldsymbol{\Phi}_i)^\top$  de tamaño  $(k + r_i) \times n$  se tiene:

$$\Delta y_{i,t}^\top = X_{i,t} \mathbf{B}_i + \boldsymbol{\epsilon}_{i,t}^\top, \quad (4-4)$$

y sí se vectoriza la ecuación (4-4) sobre  $t$  tal que  $\Delta y_i = (\Delta y_{i,1}, \dots, \Delta y_{i,T})^\top$  lo cual resulta en una matriz de tamaño  $(T \times n)$  y  $X_i = (X_{i,1}^\top, \dots, X_{i,T}^\top)^\top$  se puede escribir  $\Delta y_i = X_i \mathbf{B}_i + \boldsymbol{\epsilon}_i$ . Aplicando el operador  $\text{vec}$  se obtiene la forma:

$$\text{vec}(\Delta y_i) = (\mathbf{I}_n \otimes X_i) \text{vec}(\mathbf{B}_i) + \mathbf{e}_i$$

$$y_i = \mathbf{x}_i \mathbf{b}_i + \mathbf{e}_i,$$

donde la matriz de covarianzas de  $\mathbf{e}_i$  es  $E(\mathbf{e}_i \mathbf{e}_i^\top) = \boldsymbol{\Sigma}_{ij} \otimes \mathbf{I}_T$ . Tomando los vectores  $y_i, b_i$  y apilándolos de manera que  $y = (y_1^\top, \dots, y_N^\top)^\top$  y  $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_N^\top)^\top$ , y definiendo la matriz  $\mathbf{x}$  de tamaño  $TNn \times Nn(k + \bar{r})$  con  $\bar{r} = \sum_{i=1}^N r_i / N$  que es diagonal en bloques con diagonal igual a  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ , así que se puede escribir el sistema completo como:

$$y - \mathbf{x} \mathbf{b} = \mathbf{e},$$

donde  $y, \mathbf{b}, \mathbf{e}$  son las versiones vectorizadas y apiladas sobre  $i$  y su verosimilitud correspon-

diente es:

$$\begin{aligned} L(b, \Sigma | \beta) &\propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} (y - \mathbf{x}\mathbf{b})^\top \mathbf{V}_e^{-1} (y - \mathbf{x}\mathbf{b}) \right\} \\ &\propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} \left[ s^2 + (\mathbf{b} - \hat{\mathbf{b}})^\top \mathbf{V}_e^{-1} (\mathbf{b} - \hat{\mathbf{b}}) \right] \right\}. \end{aligned} \quad (4-5)$$

Los nuevos términos que aparecen, son descritos como:

- $s^2 = y^\top \mathbf{M}_V y$ ,
- $\mathbf{M}_V = \mathbf{V}_e^{-1} - \mathbf{V}_e^{-1} \mathbf{x} \mathbf{V}_x^\top \mathbf{V}_e^{-1}$
- $\hat{\mathbf{b}} = \mathbf{V}_x^\top \mathbf{V}_e^{-1} y$
- $\mathbf{V}_e = (\Sigma \otimes \mathbf{I}_T)$
- $\mathbf{V} = (\mathbf{x}^\top \mathbf{V}_e^{-1} \mathbf{x})^{-1}$ ,

y es posible ver que la base de esta representación se encuentra en  $\mathbf{V}_e$ , calculada en (4-3). Así mismo, esta representación permite ver que la forma posterior de  $b | \beta_i, \Sigma$  es normal si la previa en  $b$  es normal.

### 4.2.5. Tercera representación de la verosimilitud

Re-escribiendo nuevamente (4-1) con el fin de identificar  $\beta^*$ , y definiendo el vector de tamaño  $(1 \times k)$ ,  $w_{i,t} = (\Delta y_{i,t-1}^\top, \dots, \Delta y_{i,t-l}^\top, d_t^\top)$ , y la matriz  $\mathbf{C}_i = (\Gamma_{i,1}, \dots, \Gamma_{i,l}, \Phi_i)^\top$ , de manera que:

$$\Delta y_{i,t}^\top = y_{i,t-1}^\top \beta_i^* \mathbf{A}_i^\top + w_{i,t} \mathbf{C}_i + \epsilon_{i,t}^\top, \quad (4-6)$$

de nuevo, apilando los vectores a través de  $t$  tal que  $\Delta y_i = (\Delta y_{i,1}, \dots, \Delta y_{i,T})^\top$  al igual que para  $y_{i,-1} = (y_{i,0}, \dots, y_{i,T-1})^\top$  y  $w_i = (w_{i,1}^\top, \dots, w_{i,T}^\top)^\top$  se puede escribir

$$\begin{aligned}\Delta y_i &= y_{i,-1} \boldsymbol{\beta}_i^* \mathbf{A}_i^\top + w_i \mathbf{C}_i + \epsilon_i \\ \text{vec}(\Delta y_i - w_i \mathbf{C}_i) &= (\mathbf{A}_i \otimes y_{i,-1}) \text{vec}(\boldsymbol{\beta}_i^*) + \text{vec}(\epsilon_i) \\ \hat{y}_i &= \hat{\mathbf{x}}_i \mathbf{b}_{\beta^*,i} + \mathbf{e}_i,\end{aligned}$$

y apilando nuevamente sobre  $i$  se tiene entonces  $\hat{\mathbf{y}} = (\hat{y}_1^\top, \dots, \hat{y}_N^\top)^\top$  y definiendo  $\hat{\mathbf{x}}$  cómo la matriz de tamaño  $TNn \times Nn\bar{r}$  con estructura diagonal por bloques con diagonal igual a  $(\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N)$  se puede expresar el sistema de ecuaciones cómo  $\hat{\mathbf{y}} - \hat{\mathbf{x}} \mathbf{b}_{\beta^*} = \mathbf{e}$ , con  $\mathbf{b}_{\beta^*} = (\mathbf{b}_{\beta^*,1}^\top, \dots, \mathbf{b}_{\beta^*,N}^\top)^\top$  donde  $\mathbf{b}_{\beta^*,i} = \text{vec}(\boldsymbol{\beta}_i^*)$ . El lector debe notar que  $\hat{\mathbf{y}}, \mathbf{b}_{\beta^*}, \mathbf{e}$  son vectores y se usa esta notación para facilitar la lectura de las ecuaciones y evitar notaciones adicionales. La verosimilitud correspondiente queda expresada como sigue:

$$\begin{aligned}L(b, \boldsymbol{\Sigma} | \beta) &\propto |\boldsymbol{\Sigma}|^{-T/2} \exp \left\{ -\frac{1}{2} (\hat{\mathbf{y}} - \hat{\mathbf{x}} \mathbf{b}_{\beta^*})^\top \mathbf{V}_e^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{x}} \mathbf{b}_{\beta^*}) \right\} \\ &\propto |\boldsymbol{\Sigma}|^{-T/2} \exp \left\{ -\frac{1}{2} \left[ s_{\beta^*}^2 + (\mathbf{b}_{\beta^*} - \hat{\mathbf{b}}_{\beta^*})^\top \mathbf{V}_{\beta^*}^{-1} (\mathbf{b}_{\beta^*} - \hat{\mathbf{b}}_{\beta^*}) \right] \right\}.\end{aligned}\quad (4-7)$$

Nuevamente, los nuevos términos son descritos a continuación:

- $s_{\beta^*}^2 = \hat{\mathbf{y}}^\top M_{\mathbf{V}_{\beta^*}} \hat{\mathbf{y}}$
- $M_{\mathbf{V}_{\beta^*}} = \mathbf{V}_e^{-1} - \mathbf{V}_e^{-1} \hat{\mathbf{x}} \mathbf{V}_{\beta^*} \hat{\mathbf{x}}^\top \mathbf{V}_e^{-1}$
- $\hat{\mathbf{b}}_{\beta^*} = \mathbf{V}_{\beta^*} \hat{\mathbf{x}}^\top \mathbf{V}_e^{-1} \hat{\mathbf{y}}$
- $\mathbf{V}_{\beta^*} = (\hat{\mathbf{x}}^\top \mathbf{V}_e^{-1} \hat{\mathbf{x}})^{-1}$ ,

y de nuevo, esta forma de la verosimilitud permite ver que la distribución posterior para  $b_{\beta^*} | \mathbf{C}_i, \boldsymbol{\Sigma}$  es normal sí la previa condicional para  $b_{\beta^*}$  es normal.



### 4.3. Distribuciones previas

En la literatura se ha discutido que este tipo de modelos puede tener problemas de identificación local, por lo que en algunos trabajos sugieren el uso de restricciones lineales para no tener problemas con la identificación de los parámetros, sin embargo, esto puede tener problemas en la estimación de las distribuciones posteriores del modelo.

En el artículo de Kleibergen & van Dijk Kleibergen y Dijk (1994) sí en un dado caso, la matriz  $\alpha = \mathbf{0}$ , el resultado es que las distribuciones posteriores serán impropias ya que las relaciones de cointegración recogidas en  $\beta$  no aparecerían en el modelo y, por tanto, en la región paramétrica del modelo, las probabilidades marginales asociadas a esta matriz tendrían comportamientos asintóticos o límites en el interior del espacio de parámetros.

Para la especificación de una previa no informativa sobre el parámetro  $\beta$ , uno de los coautores del documento en coautoría con otro investigador, R. W. Strachan e Inder (2004) sugieren el uso de una distribución que sea similar al principio de las previas de Jeffreys de manera que sea convexa en todos los puntos de  $\beta$  y cero donde hayan puntos de no identificación (por ejemplo, cuando  $\alpha = 0$ ), de manera que no sea necesario usar restricciones de identificación lineales, es decir, intervenir lo menos posible en el modelo una vez establecidas las distribuciones previas de los parámetros. En este trabajo, para la previa no informativa, se hará uso de la previa planteada por los autores originales de la metodología Koop, Leon-Gonzalez y R. Strachan (2008) que resulta ser la misma planteada por R. W. Strachan e Inder (2004).

#### 4.3.1. Previa no informativa

Sí se define  $\mathbf{b}_\beta = (\mathbf{b}_{\beta,1}^\top, \dots, \mathbf{b}_{\beta,N}^\top)^\top$  donde  $\mathbf{b}_{\beta,i} = \text{vec}(\beta_i)$  determina los espacios de cointegración, y los parámetros restantes del modelo son  $\Sigma, \mathbf{b}$  con  $\mathbf{b}$  definido entre las ecuaciones (4-4) y (4-5). Sí se condiciona sobre  $\mathbf{b}_\beta$  el modelo se reduce a un modelo lineal como en (4-5) y la previa estándar para modelos lineales multivariados es entonces:

$$p(\mathbf{b}, \Sigma \mid \mathbf{b}_\beta) \propto |\Sigma|^{-(Nn+1)/2}, \quad (4-8)$$

donde la restricción sobre  $\beta_i$  es que sea semi ortogonal tal que  $\beta_i^T \beta_i = \mathbf{I}_{r_i}$ .

El problema principal radica en que una vez computada una previa sobre  $b_\beta$  la distribución posterior puede no ser propia, llevando a problemas de convergencia del algoritmo de muestreo. Por ello, la previa informativa incluye algunas homogeneidades tales que permitan sortear el problema de la distribución posterior impropia.

### 4.3.2. Previa informativa

Desde la ecuación (4-2), es posible ver que la restricción sobre  $\beta_i$  puede tomarse de manera que el mismo  $\beta_i$  sea semi ortogonal mientras que no existan restricciones sobre  $\alpha_i$ , o similarmente que  $\beta_i^*$  sea irrestricta mientras que  $\mathbf{A}_i$  sea semi ortogonal. Esto con el fin de que todas las distribuciones previas desarrolladas provengan de distribuciones normales y se pueda manejar mejor el tiempo de computación.

Para la distribución de  $\Sigma$  se mantendrá la previa planteada en (4-8) o se puede tomar directamente de una Wishart Inversa. Para el vector de parámetros  $\mathbf{b}$  se tiene una previa independiente:

$$\mathbf{b} \mid \Sigma, \nu \sim \mathbf{N} \left( 0, \frac{1}{\nu} \tilde{\mathbf{V}} \right), \quad (4-9)$$

donde  $\nu$  controla los grados de precisión de la previa, que puede interpretarse como un parámetro de contracción y se tratará como un parámetro del modelo. La matriz de covarianza previa, identificada por  $\tilde{\mathbf{V}}$  puede ser escogida arbitrariamente, pero puede relacionarse tanto con  $\alpha_i$  como con  $\mathbf{A}_i$ .

La matriz  $\tilde{\mathbf{V}}$  puede ser escogida con una estructura que implique o no, correlación entre los mismos parámetros para diferentes individuos del panel. La matriz de tamaño  $Nn(k + \bar{r}) \times Nn(k + \bar{r})$  puede ser particionada en  $n(k + r_i) \times n(k + r_i)$  bloques en la diagonal  $\tilde{\mathbf{V}}_{ii}$  y para relacionarlo con los  $\alpha_i$  se usa la ecuación (4-11). Por fuera de la diagonal, no es incorrecto asumir que existen ceros, sugiriendo no correlación *a priori* entre los  $\alpha_i$ ; sin embargo, puede ser importante asumir cierta correlación positiva entre los  $\text{vec}(\mathbf{C}_i)$  y  $\text{vec}(\mathbf{C}_j)$  (definidos en la ecuación (4-6)) y esto puede ser notado especificando la matriz  $nk \times nk$  de covarianzas previas de manera que  $\underline{\rho}_b \mathbf{I}_{nk}$  con  $0 < \underline{\rho}_b < 1$ .

Las relaciones planteadas en (4-2) permiten relacionar previas para  $\boldsymbol{\alpha}_i$  y  $\mathbf{A}_i$ . En principio, para las matrices de tamaño  $n \times r_i$ , semi ortogonales  $(\mathbf{A}_1, \dots, \mathbf{A}_N)$  son independientes y:

$$p(\mathbf{A}_1, \dots, \mathbf{A}_N \mid \tau, \nu, b_{\beta^*}) \propto 1, \quad (4-10)$$

implicando una densidad uniforme propia para cada uno de los subespacios definidos por  $\mathbf{A}_i, i = 1, \dots, N$ . Así mismo, la ecuación (4-10) implica que:

$$\text{vec}(\boldsymbol{\alpha}_i) \mid \tau, \boldsymbol{\beta}_i, \nu \sim \mathbf{N}_{(Nnr) \times 1} \left( 0, \frac{1}{\nu} (\boldsymbol{\beta}_i^\top \mathbf{P}_\tau^{-1} \boldsymbol{\beta}_i)^{-1} \otimes \mathbf{I}_n \right), \quad (4-11)$$

de manera que los bloques diagonales de  $\tilde{\mathbf{V}}$  que correspondan a los  $\boldsymbol{\alpha}_i$  son iguales a la estructura de covarianza  $(\boldsymbol{\beta}_i^\top \mathbf{P}_\tau^{-1} \boldsymbol{\beta}_i)^{-1} \otimes \mathbf{I}_n$ . Los elementos restantes de  $\tilde{\mathbf{V}}$  correspondientes a los parámetros  $(\mathbf{C}_1, \dots, \mathbf{C}_N)$  pueden ser especificados de manera informativa o no informativa, y serán discutidos más adelante al igual que el valor de  $\mathbf{P}_\tau$ .

Para los espacios de cointegración  $\mathbf{p}_i$ , en algunos momentos el investigador desearía establecer unos coeficientes iniciales donde las masas de probabilidad se centrarían, esto se puede notar con una matriz  $\mathbf{H}$  que contenga los coeficientes deseados en las relaciones de cointegración y que tenga la misma dimensión que el espacio de cointegración. Siguiendo el ejemplo presentado previamente,  $\mathbf{H}$  puede ser definida a partir de  $\mathbf{p} = \mathbf{H}^g$  de manera que  $\mathbf{H} = \mathbf{H}^g (\mathbf{H}^{g^\top} \mathbf{H}^g)^{-1/2}$  y:

$$\mathbf{H}^g = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix},$$

teniendo la propiedad que a partir de  $\mathbf{H}^g$ ,  $\mathbf{H}$  será semi ortogonal. Se podría afirmar que el investigador desea que los espacios de cointegración iniciales propuestos sean lo más parecidos una vez terminado el proceso de muestreo, por lo que se buscaría que el resultado final esté más cerca de  $\mathbf{p}^{\mathbf{H}} = sp(\mathbf{H})$  y más lejos de  $\mathbf{p}^{\mathbf{H}_\perp} = sp(\mathbf{H}_\perp)$  donde  $\mathbf{H}_\perp$  es el complemento ortogonal de  $\mathbf{H}$ .

Sí se define  $\mathbf{Z}_i$  de tamaño  $n \times r_i$  con todos sus elementos i.i.d.<sup>1</sup>  $\mathbf{N}(0, \nu^{-1})$ , puede decirse en un caso no informativo que  $\beta_i^* = \mathbf{Z}_i$  es una previa válida para  $\beta_i^*$ . Para el caso informativo, se puede proyectar  $\mathbf{Z}_i$  en  $\mathbf{p}^{\mathbf{H}}$  y un resultado útil en álgebra matricial indicaría que  $\beta_i^* = \mathbf{H}\mathbf{H}^{\mathbf{T}}\mathbf{Z}_i$ ; mientras que en el otro extremo se tendría que una proyección en  $\mathbf{p}^{\mathbf{H}\perp}$  y así mismo, se tendría que  $\beta_i^* = \mathbf{H}_{\perp}\mathbf{H}_{\perp}^{\mathbf{T}}\mathbf{Z}_i$ . Sí se incorpora  $\eta$ , variable aleatoria centrada en 0, que centra la previa alrededor de  $\mathbf{p}^{\mathbf{H}}$  de manera que

$$\begin{aligned}\beta_i^* &= \mathbf{H}\mathbf{H}^{\mathbf{T}}\mathbf{Z}_i + \eta\mathbf{H}_{\perp}\mathbf{H}_{\perp}^{\mathbf{T}}\mathbf{Z}_i \\ &= \mathbf{P}_{\eta}\mathbf{Z}_i,\end{aligned}$$

con  $\mathbf{P}_{\eta} = \mathbf{H}\mathbf{H}^{\mathbf{T}} + \eta\mathbf{H}_{\perp}\mathbf{H}_{\perp}^{\mathbf{T}}$ . Usando las propiedades de la distribución normal, se tiene:

$$\text{vec}(\beta_i^*) \mid \eta, \nu \sim \mathbf{N}_{(Nnr) \times 1} \left( 0, \mathbf{I}_{r_i} \otimes \frac{1}{\nu} \mathbf{P}_{\eta} \mathbf{P}_{\eta}^{\mathbf{T}} \right),$$

y por la estructura de  $\mathbf{P}_{\eta}$ , se tiene que  $\mathbf{P}_{\eta}\mathbf{P}_{\eta}^{\mathbf{T}} = \mathbf{H}\mathbf{H}^{\mathbf{T}} + \eta^2\mathbf{H}_{\perp}\mathbf{H}_{\perp}^{\mathbf{T}} = \mathbf{P}_{\eta^2}$  pero esta es la misma estructura que se podía ver en la ecuación (4-11) teniendo en cuenta que  $\tau = \eta^2$  dando como resultado:

$$\text{vec}(\beta_i^*) \mid \tau, \nu \sim \mathbf{N}_{(Nnr) \times 1} \left( 0, \mathbf{I}_{r_i} \otimes \frac{1}{\nu} \mathbf{P}_{\tau} \right),$$

equivalentemente:

$$\mathbf{b}_{\beta^*} \mid \tau, \nu \sim \mathbf{N}_{(Nnr) \times 1} \left( 0, \frac{1}{\nu} \tilde{\mathbf{V}}_{\beta^*} \right), \quad (4-12)$$

con  $\tilde{\mathbf{V}}_{\beta^*} = \text{diag}(\mathbf{I}_{r_i} \otimes \mathbf{P}_{\tau})$ . Para la previa de  $\nu$ , se selecciona como:

$$\nu \sim \mathbf{G}(\underline{\mu}_{\nu}, \underline{\nu}_{\nu} - nN\bar{r}), \quad (4-13)$$

donde  $\underline{\mu}_{\nu}, \underline{\nu}_{\nu}$  son hiperparámetros. Así mismo, la previa para  $\tau$  se selecciona:

$$\tau^{-1} \sim \mathbf{G}(\underline{\mu}_{\tau}, \underline{\nu}_{\tau}), \quad (4-14)$$

<sup>1</sup>Independientes e idénticamente distribuidas

donde  $\underline{\mu}_\tau, \underline{\nu}_\tau$  son hiperparámetros. Los autores sugieren previas que permitan que  $p(\tau < 1) \approx 1$  de manera que escogen los hiperparámetros con valores de 5, 15 respectivamente.

## 4.4. Computación

Usando las previas especificadas y las verosimilitudes en (4-5) y (4-7), se pueden derivar las distribuciones condicionales completas de cada uno de los parámetros del modelo con el fin implementar un muestreador de Gibbs. Para  $\Sigma$  la condicional posterior es una Wishart inversa con  $T$  grados de libertad y matriz de escala  $\epsilon^T \epsilon$  (definido en la ecuación (4-3)).

Para la distribución posterior de  $\mathbf{b}$  dado  $(\Sigma, \mathbf{b}_\beta)$ , se tiene que es normal con media  $\bar{\mathbf{b}}_b = \bar{\mathbf{V}}_b \mathbf{V}^{-1} \hat{\mathbf{b}}$  y matriz de covarianzas  $\bar{\mathbf{V}}_b = [\mathbf{V}^{-1} + \nu \tilde{\mathbf{V}}^{-1}]^{-1}$ , donde  $\hat{\mathbf{b}}, \mathbf{V}$  se definen en (4-5) y  $\tilde{\mathbf{V}}$  se escoge con base en las condiciones expuestas previamente en (4-9).

Para la distribución de los vectores de cointegración, se toma  $\mathbf{b}_\beta = (\mathbf{b}_{\beta,1}^T, \dots, \mathbf{b}_{\beta,N}^T)^T$  con  $\mathbf{b}_{\beta,i} = \text{vec}(\beta_i)$ . Sin embargo, esta distribución no puede ser tomada directamente de una manera sencilla, pero tal como se definió en (4-12), la distribución condicional posterior para  $\mathbf{b}_{\beta^*} = (\mathbf{b}_{\beta^*,1}^T, \dots, \mathbf{b}_{\beta^*,N}^T)^T$  con  $\mathbf{b}_{\beta^*,i} = \text{vec}(\beta_i^*)$  es normal. Recordando la definición de  $\mathbf{C}_i$  en la ecuación (4-6) y definiendo  $\mathbf{c}_i = (\text{vec}(\mathbf{A}_i)^T, \text{vec}(\mathbf{C}_i)^T)^T$  y  $\mathbf{c} = (\mathbf{c}_1^T, \dots, \mathbf{c}_N^T)^T$ , la distribución posterior para  $\mathbf{b}_{\beta^*}$  condicional en  $(\Sigma, \mathbf{c})$  es normal con media  $\bar{\mathbf{b}}_{\beta^*} = \bar{\mathbf{V}}_{\beta^*} \mathbf{V}_{\beta^*}^{-1} \hat{\mathbf{b}}_{\beta^*}$  y covarianza  $\bar{\mathbf{V}}_{\beta^*} = [\mathbf{V}_{\beta^*}^{-1} + \nu \tilde{\mathbf{V}}_{\beta^*}^{-1}]^{-1}$  definidos posterior a la ecuación (4-7), recordemos al lector que tanto  $\mathbf{c}_i$  como  $\mathbf{c}$  son vectores.

Para la distribución condicional completa de  $\nu$ , se tiene que es una distribución gamma con media

$$\bar{\mu}_\nu = \bar{\nu}_\nu \left[ (\underline{\nu}_\nu - nN\bar{r}) / \mu_\nu + \mathbf{b}^T \tilde{\mathbf{V}}^{-1} \mathbf{b} \right]^{-1},$$

y grados de libertad  $\bar{\nu}_\nu = Nnk + \underline{\nu}_\nu$ .

Para la distribución condicional completa de  $\tau^{-1}$  se tiene una gamma con media

$$\bar{\mu}_\tau = \bar{\tau}_\tau \left[ \underline{\nu}_\tau + \nu \sum_{i=1}^N \text{tr}(\beta_i^{*T} \mathbf{H}_\perp \mathbf{H}_\perp^T \beta_i^*) \right]^{-1},$$

y grados de libertad  $\bar{\nu}_\tau = \underline{\nu}_\tau + N(n\bar{r} - \sum_{i=1}^N r_i^2 / N)$ .

La distribución posterior completa se resume en las siguientes expresiones:

$$\begin{aligned}
p(\Theta \mid \mathbf{y}, \mathbf{x}) &\propto p(\mathbf{y} \mid \mathbf{x}, \mathbf{b}, \Sigma) \times p(\mathbf{b} \mid \nu, \Sigma) \times p(\Sigma) \times p(\nu) \\
&\propto \prod_{i=1}^N \mathbf{N}(\mathbf{x}_i \mathbf{b}_i, \Sigma) \times \prod_{j=1}^{Nn(k+\bar{r})} \mathbf{N}(\hat{\mathbf{b}}_j, \nu \mathbf{V}) \times \text{IW}(\boldsymbol{\epsilon}_0^\top \boldsymbol{\epsilon}_0, T) \times \mathbf{G}(\mu_\nu, \nu_\nu) \\
p(\Theta_{\beta^*} \mid \hat{\mathbf{y}}, \hat{\mathbf{x}}) &\propto p(\hat{\mathbf{y}} \mid \hat{\mathbf{x}}, \mathbf{b}_{\beta^*}, \Sigma) \times p(\mathbf{b}_{\beta^*} \mid \nu, \Sigma) \times p(\Sigma) \times p(\nu) \times p(\tau^{-1}) \\
&\propto \prod_{i=1}^N \mathbf{N}(\hat{\mathbf{x}}_i \mathbf{b}_{\beta^*,i}, \Sigma) \times \prod_{j=1}^{Nn\bar{r}} \mathbf{N}(\hat{\mathbf{b}}_{\beta^*,j}, \nu \mathbf{V}_{\beta^*}) \times \text{IW}(\boldsymbol{\epsilon}_0^\top \boldsymbol{\epsilon}_0, T) \times \mathbf{G}(\mu_\nu, \nu_\nu) \\
&\quad \times \mathbf{G}(\mu_\tau, \nu_\tau),
\end{aligned}$$

donde la primera identifica a la representación de la verosimilitud presentada en (4-5) que facilita la estimación de los parámetros de corto plazo; y la segunda identifica a la representación de la verosimilitud presentada en (4-7), que facilita la identificación de los parámetros de los multiplicadores de largo plazo. En resumen, las distribuciones condicionales completas de los parámetros resultan en las expresiones siguientes:

- La distribución condicional completa de  $\Sigma$  es:

$$\Sigma \mid \text{resto} \sim \text{IW}(\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}, T).$$

- La distribución condicional completa de  $\mathbf{b}$  es:

$$\mathbf{b} \mid \text{resto} \sim \mathbf{N}_{Nn(k+\bar{r})}([\mathbf{V}_e^{-1} \mathbf{x} + \nu \tilde{\mathbf{V}}^{-1}]^{-1} \mathbf{x}^\top \mathbf{V}_e^{-1} \mathbf{y}, [\mathbf{V}_e^{-1} \mathbf{x} + \nu \tilde{\mathbf{V}}^{-1}]^{-1}).$$

- La distribución condicional completa de  $\mathbf{b}_\beta$  es:

$$\mathbf{b}_\beta \mid \text{resto} \sim \mathbf{N}_{Nn\bar{r}}([\hat{\mathbf{x}}^\top \mathbf{V}_e^{-1} \hat{\mathbf{x}} + \nu \tilde{\mathbf{V}}_{\beta^*}^{-1}]^{-1} \hat{\mathbf{x}}^\top \mathbf{V}_e^{-1} \hat{\mathbf{y}}, [\hat{\mathbf{x}}^\top \mathbf{V}_e^{-1} \hat{\mathbf{x}} + \nu \tilde{\mathbf{V}}_{\beta^*}^{-1}]^{-1}).$$

- La distribución condicional completa de  $\nu$  es:

$$\nu \mid \text{resto} \sim \mathbf{G}((Nnk + \underline{\nu}_\nu) \left[ (\underline{\nu}_\nu - nN\bar{r})/\mu_\nu + \mathbf{b}^\top \tilde{\mathbf{V}}^{-1} \mathbf{b} \right]^{-1}, Nnk + \underline{\nu}_\nu).$$

- La distribución condicional completa de  $\tau^{-1}$  es:

$$\tau^{-1} \mid \text{resto} \sim \mathbf{G} \left( \left( \underline{\nu}_\tau + N \left( n\bar{r} - \sum_{i=1}^N r_i^2/N \right) \right) \left[ \underline{\nu}_\tau + \nu \sum_{i=1}^N \text{tr}(\boldsymbol{\beta}_i^{*\top} \mathbf{H}_\perp \mathbf{H}_\perp^\top \boldsymbol{\beta}_i^*) \right]^{-1}, \right. \\ \left. \underline{\nu}_\tau + N \left( n\bar{r} - \sum_{i=1}^N r_i^2/N \right) \right).$$

Vale la pena resaltar que la metodología presentada para identificar las distribuciones posteriores de los parámetros se basa principalmente en las previas informativas, esto es directamente observable en la manera en la que se define  $\mathbf{V}$ , ya que los bloques que corresponden a los  $\boldsymbol{\alpha}_i$  deben ser actualizados en cada iteración con los valores de  $\nu, \boldsymbol{\beta}_i, \tau$  respectivamente tal como se ve en la ecuación (4-11), así mismo debe actualizarse constantemente  $\mathbf{V}_{\boldsymbol{\beta}^*}$  de la ecuación (4-12). A pesar de que en el documento esto no se menciona, puede ahorrarle mucho tiempo al investigador a la hora de escribir su propio algoritmo de muestreo así como entender mejor la estructura de la covarianza en  $\bar{\mathbf{V}}_{\mathbf{b}}$  ya que normalmente en modelos normales multivariados, se le da más peso al valor nuevo de covarianza calculado  $\mathbf{V}$  y para ser consistente con la intención del modelo de calcular de manera correcta la matriz  $\boldsymbol{\Pi}_i$ , se le da más peso a la previa informativa.

#### 4.4.1. Algoritmo

Teniendo en cuenta que el conjunto de parámetros e hiperparámetros del modelo son respectivamente  $(\mathbf{b}, \boldsymbol{\Sigma}, \mathbf{b}_\beta, \nu, \tau)$  y  $(\underline{\mu}_\nu, \underline{\nu}_\nu, \underline{\mu}_\tau, \underline{\nu}_\tau, r_i, \tilde{\mathbf{V}}, \tilde{\mathbf{V}}_{\boldsymbol{\beta}^*})$ , el algoritmo consta de un muestreador de Gibbs (Gamerman y Lopes (2006)) con los siguientes pasos:

1. Inicializar  $(\mathbf{b}, \boldsymbol{\Sigma}, \mathbf{b}_\beta, \nu, \tau) = (\mathbf{b}^{(0)}, \boldsymbol{\Sigma}^{(0)}, \mathbf{b}_\beta^{(0)}, \nu^{(0)}, \tau^{(0)})$ .
2. Muestrear  $\boldsymbol{\Sigma} \mid \mathbf{b}^{(0)}, \mathbf{b}_\beta^{(0)}, \nu^{(0)}, \tau^{(0)}$  de  $\text{IW}(\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}, T)$ , de tamaño  $Nn \times Nn$ , esto puede ser hecho al principio o al final del muestreo.
3. Muestrear  $\mathbf{b} \mid \boldsymbol{\Sigma}^{(b)}, \mathbf{b}_\beta^{(0)}, \nu^{(0)}, \tau^{(0)}$  de  $\mathbf{N}(\bar{\mathbf{b}}_b, \bar{\mathbf{V}}_b)$  de tamaño  $Nn(k + r_i) \times 1$ .
4. Calcular  $\mathbf{A}_i = (\boldsymbol{\alpha}_i^\top \boldsymbol{\alpha}_i)^{-1/2} \boldsymbol{\alpha}_i$  y crear  $\mathbf{c}$ . Recordando que  $\boldsymbol{\alpha}_i$  está contenido en  $\mathbf{b}$ , tal

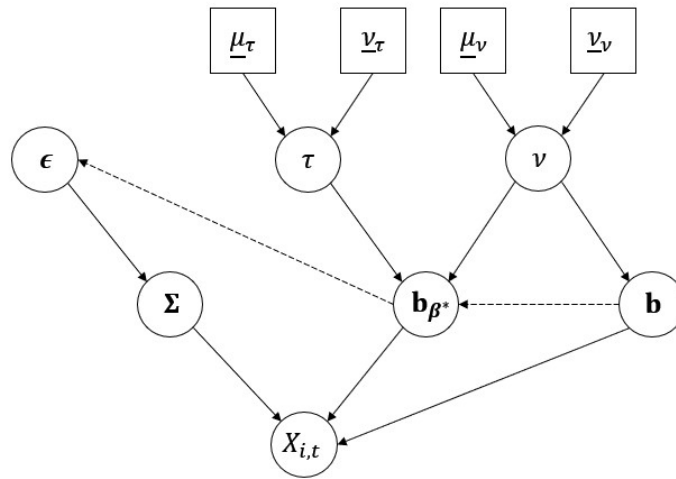
como se ilustró previo a la ecuación (4-4). Así mismo,  $\mathbf{A}_i, \mathbf{c}$  sirven para calcular  $\hat{\mathbf{x}}_i$  y con base en éstos, calcular  $\mathbf{V}_{\beta^*}$  y  $\hat{\mathbf{b}}_{\beta^*}$  como se ilustra en la ecuación (4-7).

5. Muestrear  $\mathbf{b}_{\beta^*} \mid \mathbf{c}^{(b)}, \boldsymbol{\Sigma}^{(b)}, \nu^{(0)}, \tau^{(0)}$  de  $\mathbf{N}(\bar{\mathbf{b}}_{\beta^*}, \bar{\mathbf{V}}_{\beta^*})$  de tamaño  $Nn(k + \bar{r}) \times 1$ . Nótese que  $\mathbf{b}_{\beta^*}$  tiene elementos similares a  $\mathbf{b}$ , por lo que en gran medida la distribución de  $\mathbf{b}$  determina a  $\mathbf{b}_{\beta^*}$ , y por ello aparece  $\nu$  con el fin de comportarse como un parámetro de “encogimiento”.
6. Descomponer cada  $\beta_i^* = \beta_i \kappa_i$  con  $\kappa_i = (\beta_i^{*\top} \beta_i^*)^{1/2}$  y  $\beta_i = \beta_i^* \kappa_i^{-1}$ . Posteriormente, construir  $\boldsymbol{\alpha}_i = \mathbf{A}_i \kappa_i$
7. Muestrear  $\nu \mid \mathbf{b}^{(b)}, \mathbf{b}_{\beta}^{(b)}, \boldsymbol{\Sigma}^{(b)}, \tau^{(0)}$  de  $\mathbf{G}(\bar{\mu}_\nu, \bar{\nu}_\nu)$ , escalar resultado de la ecuación (4-13).
8. Muestrear  $\tau^{-1} \mid \mathbf{b}^{(b)}, \mathbf{b}_{\beta}^{(b)}, \boldsymbol{\Sigma}^{(b)}, \nu^{(b)}$  de  $\mathbf{G}(\bar{\mu}_\tau, \bar{\nu}_\tau)$ , escalar resultado de la ecuación (4-14).
9. Repetir pasos 2 – 8 por un número  $B$  de réplicas suficientes hasta convergencia a la distribución estacionaria.

Al final, se tiene un total de  $Nn(k + 2\bar{r} + 1) + 2$  parámetros totales del modelo. El grafo acíclico dirigido, DAG por sus siglas en inglés, se encuentra en la figura 4-1. El DAG muestra que los parámetros muestreados en  $\mathbf{b}$  determinan a los parámetros donde están los multiplicadores de largo plazo  $\mathbf{b}_{\beta^*}$  y estos a su vez, una vez calculados, permiten encontrar los parámetros que le dan forma a la distribución Wishart Inversa de  $\boldsymbol{\Sigma}$ . El parámetro  $\tau$  determina los espacios de cointegración en  $\mathbf{b}_{\beta^*}$  y el parámetro  $\nu$  actúa como un factor de corrección en las varianzas de  $\mathbf{b}$  y  $\mathbf{b}_{\beta^*}$ . Los hiperparámetros del modelo se encuentran en la definición de  $\tau$  y  $\nu$  que tal como se indicó en su sección, conviene que  $p(\tau < 1) \approx 1$ . En resumen, se deben muestrear los vectores correspondientes a  $\mathbf{b}$  y  $\mathbf{b}_{\beta^*}$ , la matriz  $\boldsymbol{\Sigma}$  que depende de los errores  $\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}$ , y los escalares  $\nu$  y  $\tau$ .

Es importante tener en cuenta que el modelo se especializa en estimar de manera correcta la matriz de parámetros  $\mathbf{\Pi}_i$ , es decir, el producto de las matrices  $\boldsymbol{\alpha}_i, \beta_i$  de rango completo. El modelo no puede estimar de manera consistente estos parámetros de manera individual, ya que el cambio en las representaciones de la verosimilitud, sí bien solucionan el problema de identificación de  $\beta_i$  cuando  $\boldsymbol{\alpha}_i = \mathbf{0}$ , no permiten la estimación directa de estos parámetros.





**Figura 4-1.:** DAG modelo, elaboración propia.

Sí el investigador se encuentra en una situación donde requiera conocer de manera directa el efecto de corrección de error, o las relaciones específicas de largo plazo entre las variables dependientes, esta metodología no es recomendable.

En caso de que el lector desee estimar  $\rho$  en el modelo para caracterizar mejor la estructura de las series temporales usadas, basta con añadir un paso adicional en el muestreador de Gibbs presentado anteriormente, haciendo uso de un algoritmo de Metropolis-Hastings simulándolo de una distribución simétrica y con el cálculo de su respectiva tasa de aceptación.

## 5. Simulación

En este capítulo se muestra un ejercicio de simulación que pone a prueba la capacidad del modelo de identificar los valores reales de los parámetros de las matrices  $\mathbf{\Pi}_i$  y  $\mathbf{\Gamma}_i$ , con el fin de evaluar el desempeño del modelo panel cointegrado a la hora de estimar los parámetros reales del proceso simulado.

En este estudio de simulación, se consideran tres procesos con diferente longitud pero con algunas similitudes para lograr estresar el modelo al máximo con el fin de evaluar su precisión. Principalmente, dada la naturaleza de los datos temporales, se estresa a través del uso de diferente número de observaciones por individuo  $T$ , manteniendo constante el número de relaciones de cointegración  $r$ , el número de individuos  $N$  y el número de variables dependientes del sistema  $n$ . Esto debido a la naturaleza del modelo, ya que la precisión posterior de los estimadores depende de la longitud de observaciones de cada individuo y no de su cantidad de variables dependientes o de sus relaciones de cointegración.

Para cada uno de los parámetros del modelo, se evalúan las medidas de convergencia de las cadenas, con el fin de revisar si el modelo efectivamente convergió o si hubo problemas de autocorrelación en las cadenas de Markov. Para esto, se evalúa tanto el tamaño efectivo de muestra como los errores de Monte Carlo para cada uno de los parámetros del modelo.

En cada uno de los casos, para medir la precisión de la estimación, se tomarán 5 medidas diferentes:

- Proporción de intervalos de credibilidad al 95 % de las cadenas posteriores que contie-

nen al parámetro simulado, esto es  $\frac{P}{n}$  donde:

$$P = \begin{cases} 1, & \text{si el intervalo de credibilidad contiene al parámetro real;} \\ 0, & \text{en otro caso.} \end{cases}$$

- Error cuadrático medio, RMSE por sus siglas en inglés, calculado como:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\gamma_i - \bar{\gamma}_i)^2}.$$

- Error medio absoluto, MAE por sus siglas en inglés, calculado como:

$$\frac{1}{n} \sum_{i=1}^n |\gamma_i - \bar{\gamma}_i|.$$

- Longitud media del intervalo de credibilidad al 95 %.
- Sesgo, calculado como:

$$\frac{1}{n} \sum_{i=1}^n (\gamma_i - \bar{\gamma}_i).$$

En cada una de las ecuaciones, se simboliza  $\gamma_i$  al valor real que se usó en la simulación de cada caso, mientras que  $\bar{\gamma}_i$  representa el valor promedio posterior de cada parámetro. En cada caso, se calcula la medida en los dos subgrupos de parámetros  $\mathbf{\Pi}_i, \mathbf{\Gamma}_i$ .

Todos los ejercicios de simulación cuentan con un período de calentamiento de 1.000 iteraciones y una cadena de 10.000 iteraciones, para un total de 11.000 iteraciones. Los ejercicios se realizan en R Studio, contando con un procesador Ryzen 7 3700x con una frecuencia máxima de 4.3 GHz y 32 GB de RAM.

## 5.1. Caso 1: Longitud mínima

Para este caso de simulación, se usan 30 datos de longitud generados por un proceso con un sólo rezago de las variables diferenciadas y las siguientes características:

$$\mathbf{\Pi}_i = \begin{pmatrix} -0,048 & 0,006 & -0,069 & 0,038 \\ -0,060 & 0,045 & -0,054 & 0,058 \\ -0,010 & -0,009 & -0,020 & 0,004 \\ 0,009 & 0,057 & -0,049 & 0,052 \end{pmatrix}, \mathbf{\Gamma}_{i,1} = \begin{pmatrix} 0,3 & 0,01 & -0,13 & 0 \\ 0 & 0,4 & -0,3 & 0 \\ 0,3 & -0,24 & 0,1 & -0,3 \\ -0,2 & 0,45 & -0,09 & -0,12 \end{pmatrix}$$

Con  $i = 1, 2, 3$ , y cada individuo del panel tiene el mismo proceso generador de datos. Adicionalmente, para el término de perturbación, se tiene la matriz de covarianzas:

$$\mathbf{\Sigma}_{i,i} = \begin{pmatrix} 1,80 & 0,45 & 1,60 & 0,70 \\ 0,45 & 1,10 & 0,70 & 0,20 \\ 1,60 & 0,70 & 5 & 1,30 \\ 0,70 & 0,20 & 1,30 & 1,90 \end{pmatrix}, \mathbf{\Sigma}_{i,j} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

En otros términos, entre individuos no existe covarianza, por lo que el modelo los trata como independientes. Esto se aplica principalmente para evaluar sí el modelo tiene diferencias entre individuos que sean iguales a la hora de estimar las distribuciones posteriores.

Los resultados de la simulación pueden verse en la figura **A-1** de los anexos, donde se muestra con un triángulo rojo el valor real del parámetro, con un cuadrado naranja el valor estimado del parámetro y con una línea azul el intervalo de credibilidad correspondiente al 95 %, esto con el fin de establecer directamente los resultados de los procedimientos de estimación para cada individuo y para cada conjunto de parámetros. Es posible ver que las estimaciones de las matrices de coeficientes de corto plazo identificadas por  $\mathbf{\Gamma}$  en su mayoría están centradas alrededor de 0, mostrando que el modelo falla en predecir correctamente las entradas de la matriz. Los únicos valores que logra predecir correctamente son aquellos que están muy cerca de 0.

Sin embargo, es notable que los resultados de la matriz  $\mathbf{\Pi}$  se encuentran considerablemente cerca de los valores reales simulados, con intervalos de confianza considerablemente cortos en comparación a los demás ejercicios de simulación, sin embargo cuando no es capaz de predecir correctamente un valor simulado, su error a primera vista parece ser mayor en comparación

a los fallos de los demás ejercicios de simulación.

El tiempo de ejecución por cada 100 iteraciones de este modelo es de 38.18 segundos totalizando un tiempo de 1.16 horas para las 11.000 iteraciones, realizado completamente en R Studio.

## 5.2. Caso 2: Longitud moderada

En el segundo ejercicio se evalúa un proceso similar en cuanto a la estructura de individuos y variables dependientes, contando con una longitud de observación de 100 períodos. Las características de este proceso son las siguientes:

$$\mathbf{\Pi}_i = \begin{pmatrix} -0,260 & 0,220 & -0,150 & 0,250 \\ -0,150 & 0,120 & -0,150 & 0,320 \\ -0,100 & 0,200 & -0,200 & 0,500 \\ 0,230 & -0,160 & 0,180 & -0,350 \end{pmatrix}, \mathbf{\Gamma}_{i,1} = \begin{pmatrix} 0,3 & 0,01 & -0,5 & 0 \\ 0 & 0,4 & -0,3 & 0 \\ 0,3 & -0,2 & 0,1 & -0,3 \\ -0,2 & 0,45 & -0,2 & -0,12 \end{pmatrix}$$

Con  $i = 1, 2, 3$ , y cada individuo del panel tiene el mismo proceso generador de datos. Adicionalmente, para el término de perturbación, se tiene la matriz de covarianzas:

$$\mathbf{\Sigma}_{i,i} = \begin{pmatrix} 1,30 & 0,40 & 1,60 & 0,80 \\ 0,40 & 1,10 & 0,70 & 0,20 \\ 1,60 & 0,70 & 5 & 1,30 \\ 0,80 & 0,20 & 1,30 & 1,90 \end{pmatrix}, \mathbf{\Sigma}_{i,j} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Teniendo el mismo caso que la simulación anterior donde se asume que el individuo es el mismo y no hay covarianza entre individuos. Los resultados de la simulación pueden encontrarse en la figura **A-2**.

Es posible notar que para las entradas de la matriz  $\mathbf{\Gamma}$  tiene un desempeño mucho mejor que la simulación anterior, ya que varios de los intervalos de credibilidad no sólo contienen el valor simulado real, sino que también en varios casos estos intervalos están centrados alrededor de este, lo cual cabría de esperarse ya que se tiene 3 veces el tamaño de muestra de la simulación anterior.

En el caso de las entradas de la matriz  $\mathbf{\Pi}$  pasa algo curioso, ya que es posible notar que la mayoría de los intervalos están centrados en el valor simulado, pero la amplitud de los mismos es considerablemente mayor a los que se pueden ver en la matriz de la primera simulación. Es posible que esto se deba a que los valores de la matriz  $\mathbf{\Pi}$  simulada sean mayores en tamaño. El valor promedio posterior de cada valor de la entrada está más cerca del valor simulado que en el caso de la primera simulación.

El tiempo de ejecución por cada 100 iteraciones de este modelo es de 70.90 segundos totalizando 2.16 horas para las 11.000 iteraciones, realizado en R Studio salvo el cálculo de la matriz  $\mathbf{V}_e$  que se realiza importando un método de python para la inversión de matrices, que recordándole al lector, esta matriz depende del tamaño de muestra usado  $T$  ya que se calcula como  $(\mathbf{\Sigma} \otimes \mathbf{I}_T)$ .

### 5.3. Caso 3: Longitud máxima

En el último ejercicio, se quiso evaluar la cantidad más grande posible de longitud observada con 300 observaciones, con el fin de estresar el tiempo de ejecución. Las características de este proceso son las siguientes:

$$\mathbf{\Pi}_i = \begin{pmatrix} -0,105 & 0,001 & -0,111 & 0,050 \\ -0,100 & 0,002 & -0,069 & 0,048 \\ -0,010 & -0,020 & -0,020 & 0,004 \\ 0,078 & -0,018 & 0,064 & -0,038 \end{pmatrix}, \mathbf{\Gamma}_{i,1} = \begin{pmatrix} 0,3 & 0,01 & -0,15 & 0 \\ 0 & 0,4 & -0,3 & 0 \\ 0,3 & -0,2 & 0,1 & -0,3 \\ -0,2 & 0,45 & -0,2 & -0,12 \end{pmatrix}$$

Con  $i = 1, 2, 3$ , y cada individuo del panel tiene el mismo proceso generador de datos. Adicionalmente, para el término de perturbación, se tiene la matriz de covarianzas:

$$\mathbf{\Sigma}_{i,i} = \begin{pmatrix} 1,80 & 0,45 & 1,60 & 0,70 \\ 0,45 & 1,10 & 0,70 & 0,20 \\ 1,60 & 0,70 & 5 & 1,30 \\ 0,70 & 0,20 & 1,30 & 1,90 \end{pmatrix}, \mathbf{\Sigma}_{i,j} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Los resultados pueden verse en la figura **A-3**. En este caso pueden verse los mejores resultados de las matrices de coeficientes, lo cual es esperado dada la cantidad de datos que se tienen y la información que puede ser recogida de los mismos. En la matriz  $\mathbf{\Gamma}$  la mayoría de intervalos de credibilidad están centrados en el valor simulado real y la longitud de intervalos de credibilidad es considerablemente pequeña teniendo en cuenta los dos escenarios anteriores. En el caso de la matriz  $\mathbf{\Pi}$  también es posible ver que la precisión de estimación de sus entradas es alta, pero es interesante ver que a través de los 3 individuos evaluados, en algunos casos, los intervalos de credibilidad varían de un individuo a otro aunque eso no afecte el valor promedio posterior que suele coincidir con el valor simulado real.

El tiempo de ejecución por cada 100 iteraciones de este modelo es de 693.81 segundos totalizando 21.2 horas para las 11.000 iteraciones, realizado en R Studio salvo el cálculo de la inversa de la matriz  $\mathbf{V}_e$  que nuevamente se hace por medio de Python. Esta es una clara limitación de este modelo ya que a medida que se incluyan más observaciones, más lenta será la estimación. Una estrategia útil para la estimación de este modelo es usar un lenguaje de programación más veloz o que permita paralelizar con GPU el cálculo de la inversa de  $\mathbf{V}_e$ .

## 5.4. Evaluación de la convergencia de las cadenas

Antes de revisar la precisión más allá de la manera gráfica, vale la pena revisar los tamaños efectivos de muestra y los errores de montecarlo de las cadenas de cada uno de los parámetros usados para la estimación de cada escenario de simulación.

En la figura **A-4** se presentan los tamaños efectivos de muestra para cada uno de los escenarios de simulación. Lo más notable es que en la simulación más corta de 30 observaciones, la distribución de los tamaños efectivos es mucho más dispersa que en los otros dos ejercicios de simulación. La línea horizontal que divide los gráficos marca el tamaño de las cadenas finales (10.000 observaciones).

Los tamaños de muestra que consistentemente sobrepasan el tamaño de las cadenas son los de la simulación moderada de 100 observaciones, con excepción de las entradas de la matriz  $\mathbf{\Sigma}$ . Los tamaños efectivos de muestra más pequeños se encuentran principalmente en la simulación extrema, mostrando indicios de existencia de autocorrelación en sus cadenas.

Es notable que  $\tau$  consistentemente tiene un tamaño efectivo de muestra mucho mayor al tamaño de las cadenas, esto es positivo ya que este parámetro se encarga de identificar el espacio de cointegración y la correcta estimación de la matriz  $\mathbf{\Pi}$ .

En la figura **A-5** se presentan los errores estándar de Monte Carlo, y como es posible ver en el eje vertical, todos los errores de los parámetros son muy bajos, confirmando una vez más que las cadenas han convergido de manera satisfactoria. En la representación gráfica se omiten los errores de Monte Carlo de  $\nu$  para efectos de una mejor visualización porque este parámetro tiene los errores más grandes en todo el modelo, que son: 123 para la simulación de 30 observaciones, 32 para la simulación de 100 observaciones y 57 para la simulación de 300 observaciones.

Este resultado va en línea con la dificultad de muestrear de  $\nu$  teniendo en cuenta que gran parte de la información proviene de  $\mathbf{V}$  donde tiene que tenerse en cuenta que incluye la información de las covarianzas de los parámetros de corto plazo en  $\mathbf{\Gamma}$  y los  $\alpha$ , por lo que la gran dificultad del modelo radica en gran parte sobre la definición de esta matriz. Para el caso evaluado en este trabajo, se toman valores pequeños en los bloques diagonales correspondientes a los parámetros de corto plazo sin información previa. Pero cabe resaltar que es el único parámetro que tiene un error de Monte Carlo alto, comparado contra los más de 200 parámetros del modelo.

## 5.5. Evaluación de la precisión del modelo

La tabla **A-1** muestra las medidas de precisión de las simulaciones, divididas por parámetro y simulación con el fin de evaluar la diferenciación entre los procesos de muestreo.

En el caso de la estimación de  $\mathbf{\Gamma}$ , la simulación corta contiene apenas un 35.41% de los valores simulados dentro de los intervalos de credibilidad al 95%. La simulación moderada tiene el menor error cuadrático medio para este parámetro con 0.01. El error absoluto medio es mínimo para la simulación extrema con 0.05. La longitud promedio es mínima para la simulación corta, pero dado que no contiene a muchos de los valores simulados, se tomará la longitud promedio de la simulación extrema como la menor medida para este ítem. Finalmente, el sesgo es menor en la simulación moderada.



Es posible notar que tanto en la simulación moderada como en la simulación extrema, el modelo tiene buenos resultados a la hora de estimar las entradas de la matriz  $\mathbf{\Gamma}$ , por lo que se recomienda que este modelo no se aplique en series cortas de observación, y que sean preferiblemente de 100 observaciones para no comprometer el tiempo de estimación.

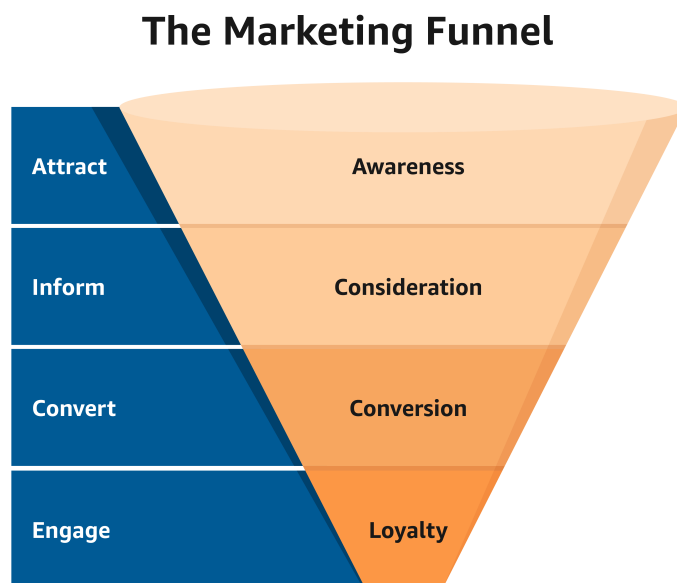
Para la estimación de  $\mathbf{\Pi}$ , las tres simulaciones tienen resultados sobresalientes, donde tanto la simulación moderada como la simulación extrema contienen en un 97.91 % a los valores simulados dentro del intervalo de credibilidad del 95 %. El error cuadrático medio es menor en la simulación extrema con un valor de 0.026. El error medio absoluto y la longitud promedio también son menores en la simulación extrema con 0.019 y 0.103 respectivamente. El sesgo también es menor en la simulación extrema con -0.003.

En la estimación de  $\mathbf{\Pi}$  la simulación extrema tiene un notable mejor desempeño que las demás, lo cual va en línea con lo esperado debido a la información adicional que se puede obtener a la hora de contar con más puntos de data. Sin embargo, la simulación corta tiene un desempeño notable con respecto a la simulación moderada ya que tiene mejores métricas en todos los items salvo en la proporción de intervalos que contienen al valor simulado.

## 6. Marketing Digital

### 6.1. Ilustración y análisis de los datos

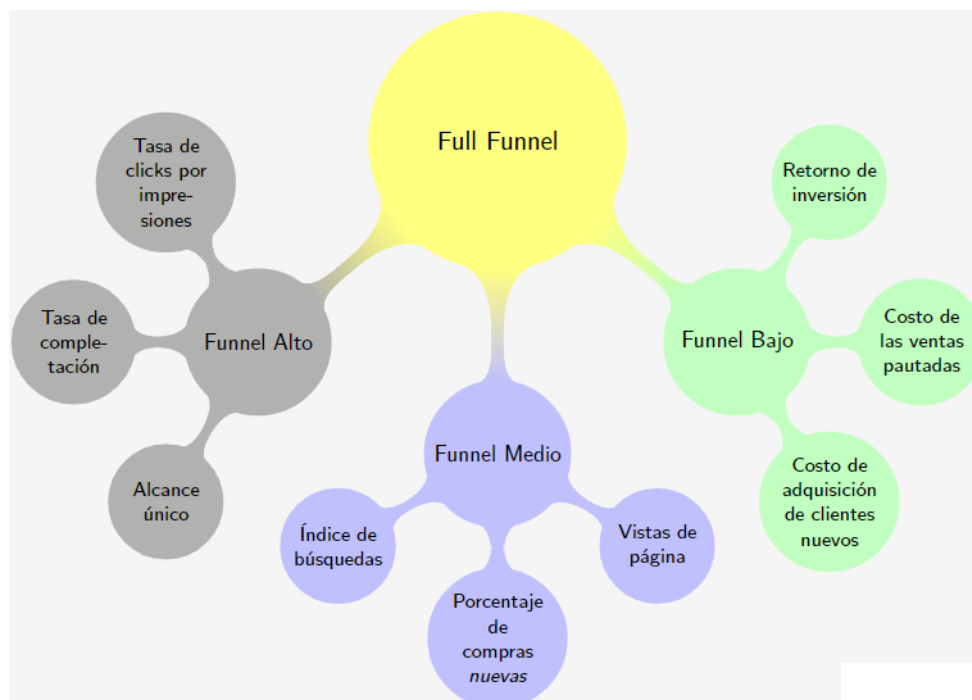
En el contexto de márketing se hace ampliamente necesario el análisis de variables simultáneas, ya que representan diferentes etapas de compra de un usuario promedio. En particular, se identifica la necesidad de hacer un seguimiento minucioso a las variables que son parte integral de una pauta publicitaria y así mismo, se hace necesario identificar la etapa de consumo en la que un cliente se encuentra. Se presenta entonces el concepto de *marketing funnel* con diferentes etapas identificables tal como se presenta en el gráfico **6-1**:



**Figura 6-1.:** Full funnel, tomado de Amazon Ads.

En este gráfico, se identifican cuatro etapas principales en el proceso de conversión de una

pauta publicitaria: **Atracción**, relacionada con el conocimiento de un potencial cliente de la marca; **información**, que es la consideración que tienen los clientes a la hora de elegir la marca pautaada u otras disponibles en la misma categoría; **conversión**, el KPI estratégico según la necesidad del cliente relacionado con las conversiones sean ventas o alguna métrica en particular como sesiones de página web; y **engage** que refiere a la recurrencia de conversión de los clientes. Dependiendo de la etapa en la que esté el cliente, se trabajarán diferentes estrategias para lograr que se lleve a cabo una conversión o algún objetivo de la marca en cuestión.



**Figura 6-2.:** Métricas en las etapas del *marketing funnel*, elaboración propia.

Este análisis, según la figura 6-2, estará centrado en la parte alta y media del *funnel* ya que se tienen disponibles las medidas de clicks, impresiones, inversión y sesiones nuevas por página. En cuanto a la utilidad de este análisis práctico realizado, permite plantear una alternativa a los modelos por etapas realizados usualmente en la industria del *marketing*, ya que el problema es que estos modelos usualmente se realizan usando mínimos cuadrados por etapas sobre variables que no son estacionarias, incurriendo en muchas ocasiones en problemas de relaciones espurias. Luego de dar un contexto apropiado al problema, se procede a explicar

el procedimiento realizado.

El análisis se realizará sobre un conjunto de datos el cual corresponde a una marca de alimentos latinoamericana, conjunto que consiste en la evaluación de las métricas relevantes de la inversión en medios digitales, esto es, inversión, clicks, impresiones y sesiones de una página web donde se aloja el servicio que el cliente está interesado en promocionar. La estrategia que se siguió fue recomendarle al cliente que invirtiera en tres campañas en un período de tiempo posterior al período de pandemia de poco más de año y medio. Los datos se escalaron de 1 a 100 con el fin de proteger la información real del cliente. En este contexto particular, una campaña hace referencia a algún tipo de pauta publicitaria en un formato específico, llámese digital o televisión por nombrar alguno, que se hace con el fin de dar a conocer una marca, promocionar un producto o impulsar las ventas de un producto ya existente.

A modo de contexto, cada una de las variables pueden ser descritas de la siguiente forma:

- **Inversión:** Es la variable sobre la cual el cliente tiene más control, así mismo, esta variable suele tener ciertos procesos de optimización que buscan maximizar el retorno de inversión dado un objetivo. En el proyecto original de este caso particular, se buscaba maximizar la conversión de ventas en página web y de manera presencial. La inversión está medida en la unidad monetaria del país latinoamericano de referencia y el cliente directamente envía esta información.
- **Clicks:** Mide la cantidad de interacciones que tienen las personas que ven el anuncio en cualquier plataforma donde se aplique el formato, por ejemplo, Meta contiene la información de redes sociales como Instagram y Facebook. Esto depende no sólo del formato del anuncio, sino también, de la población a la que va dirigida. Proviene de tres diferentes fuentes: La información de la página propia recopilada por ellos mismos, los datos de Google Ads y el proveedor de espacios de publicidad.
- **Impresiones:** Mide la cantidad total de veces que un anuncio es mostrado a una o a varias personas. Suelen depender de la estrategia de subasta que tenga el cliente y la prioridad que tenga en la plataforma de ventas, ya que usualmente las impresiones

pueden comprarse siguiendo una estrategia de expansión (audiencias nuevas que no conocen el producto) o de fidelización (audiencias que son familiares con el producto o audiencias con altas probabilidades de comprar el producto). Proviene de cinco diferentes fuentes y plataformas entre las que destacan Google Ads, Teads y Facebook Ads.

- **Sesiones:** Las sesiones son registradas por la página web y mide la cantidad de usuarios nuevos que son registrados y que provienen de alguna publicidad de los canales correspondientes, esto quiere decir que la sesión se originó gracias a un anuncio y se identifica a través del historial de cookies. Estas sesiones también se les conoce como sesiones pagas.

Los datos son recogidos directamente desde Google Ads, Facebook Ads, y el gestor de información de la inversión e impresiones de los anuncios se mantiene anónimo con el fin de proteger los datos y la marca. A pesar de que los datos de inversión y sesiones pueden ser extraídas diariamente, el proveedor de la información de medios maneja las estrategias de campaña de forma semanal, por lo cual, el análisis se hará de dicha manera, contando con 70 observaciones por canal desde principios de 2021 hasta mediados de 2022 considerando observaciones semanalmente.

En las figuras se pueden apreciar los comportamientos de cada una de las campañas con sus respectivas variables. En la figura **6-3** se tiene la campaña 1 que tiene la particularidad de que todas sus variables tienen ciertas similitudes entre 2021-Q3 y 2022-Q1, salvo sesiones, y la estrategia del cliente se suele concentrar más en los resultados de esta estrategia. En la figura **6-4** se ve a primera vista que las impresiones y la inversión son las variables que tienen mayor similitud entre sí, al igual que los clicks y las sesiones. Finalmente en la figura **6-5** se observan los comportamientos de la campaña 3; la cual es una campaña experimental con la que el cliente estuvo probando una estrategia innovadora con sus compradores y que se centra en formatos de anuncios más interactivos.

Al analizar la participación relativa de cada uno de los componentes de la marca mediante la figura **6-6** se puede notar que para la campaña 3 hubo un aumento sustancial de inversión en la primera mitad de 2021, notable también en impresiones y clicks. pero que no tuvo

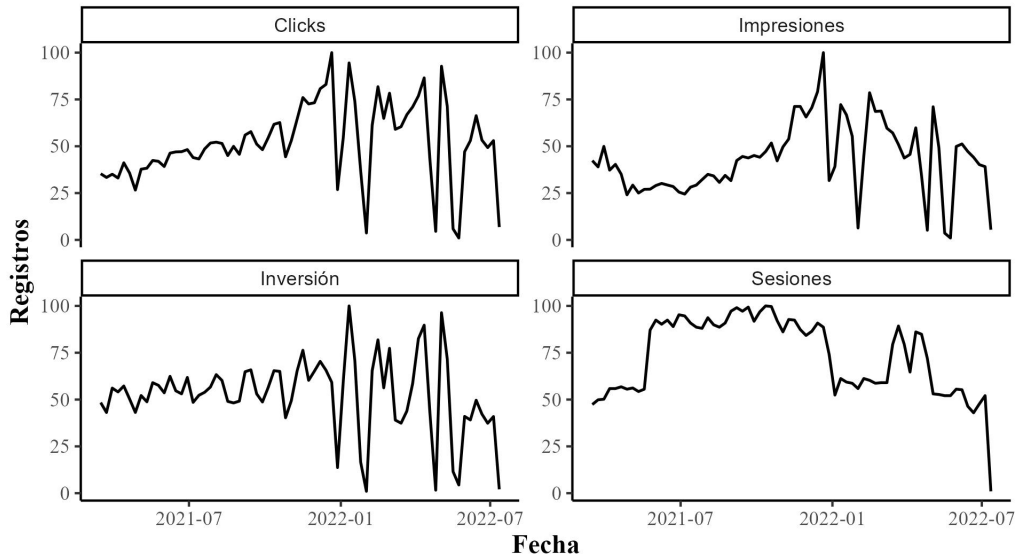


Figura 6-3.: Campaña 1, variables del sistema.

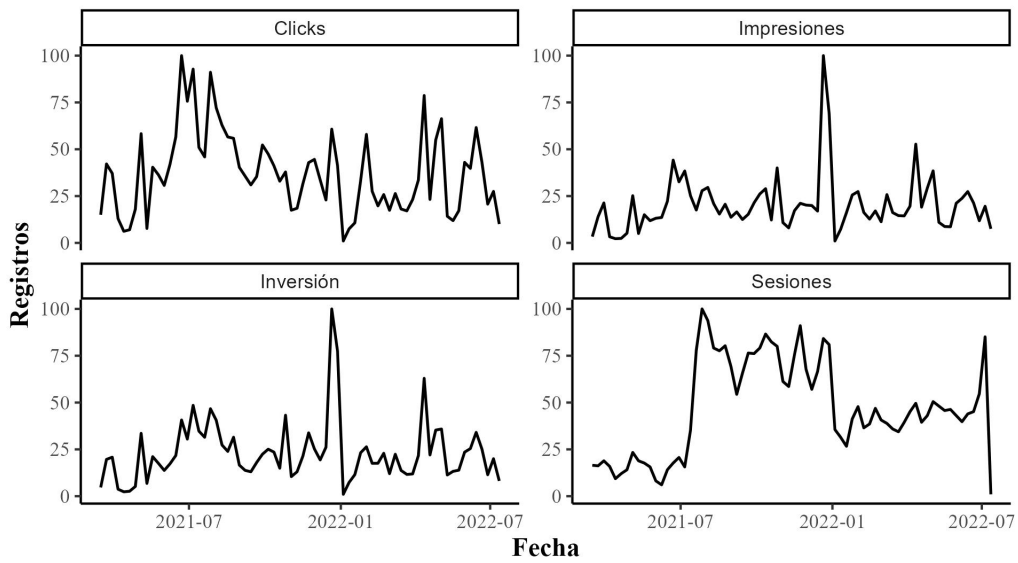


Figura 6-4.: Campaña 2, variables del sistema.

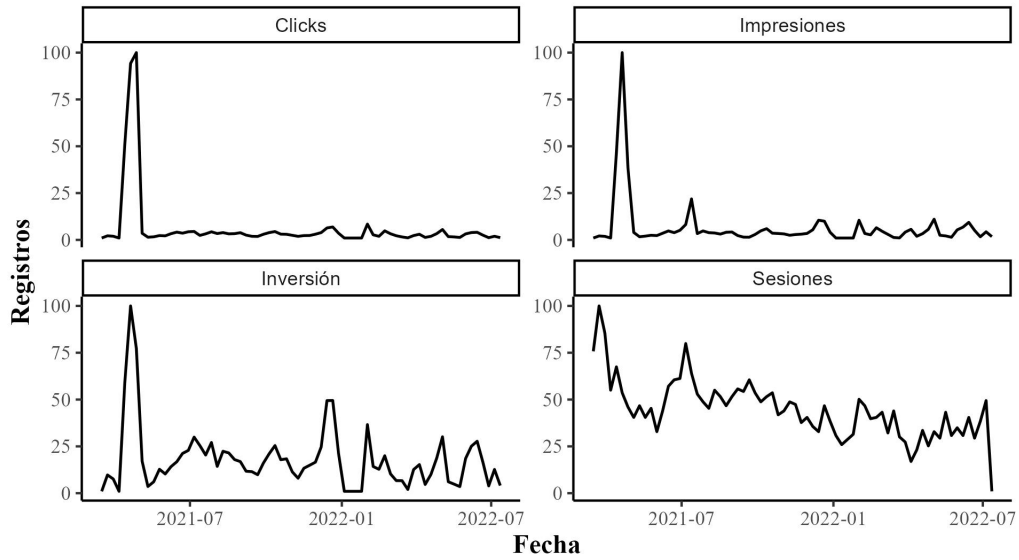


Figura 6-5.: Campaña 3, variables del sistema.

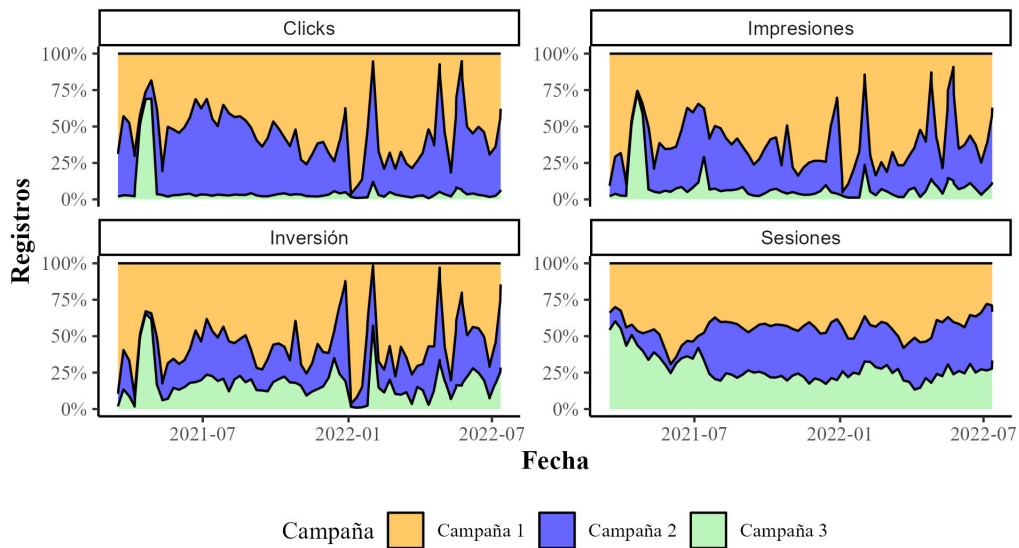


Figura 6-6.: Montos relativos de cada componente en el total.

gran repercusión en las sesiones. Así mismo, a pesar de que la campaña 1 mantiene la mayor contribución en inversión y clicks, parece estar rezagada contra la campaña 2 en impresiones, y en sesiones la campaña 2 incrementa su porcentaje de participación.

Al evaluar los estadísticos de prueba de las series sobre si son o no estacionarias, haciendo uso del estadístico Dickey-Fuller sin deriva y sin tendencia lineal, usando 3 rezagos para la prueba usados principalmente para evaluar un número de rezagos estándar a través de las series, se encuentran los p-valores consignados en la tabla **6-1**. A primera vista, es notable que las series de la campaña 3 no pueden ser tomadas como series no estacionarias con las condiciones que se han planteado para la prueba ADF. Las series correspondientes a la campaña 2, apenas sí aceptan la hipótesis nula de no estacionariedad, en particular los valores de inversión e impresiones.

<b>Campaña</b>	<b>Clicks</b>	<b>Inversión</b>	<b>Impresiones</b>	<b>Sesiones</b>
<b>Campaña 1</b>	0.371	0.298	0.301	0.447
<b>Campaña 2</b>	0.190	0.083	0.069	0.414
<b>Campaña 3</b>	0.010	0.016	0.010	0.045

**Tabla 6-1.:** P-valores para prueba ADF tipo 1 con tres rezagos.

Usando la metodología de Tsay con el fin de identificar el número de rezagos adecuado para cada prueba, cuyos resultados se encuentran consignados en la tabla **6-2**, se encuentra evidencia de que para varias series el número idóneo de rezagos está por encima de los rezagos usados para la prueba ADF, pero los resultados son igualmente válidos en este caso.

Sí se revisa un poco más a fondo la estructura de cada serie temporal, cuando las series son diferenciadas, apenas el primer rezago resulta tener un valor de autocorrelación alto, como se ve principalmente en las variables de clicks, impresiones e inversión de la campaña 1 y la

<b>Campaña</b>	<b>Clicks</b>		<b>Impresiones</b>		<b>Inversion</b>		<b>Sesiones</b>	
	<b>p-valor</b>	<b>lags</b>	<b>p-valor</b>	<b>lags</b>	<b>p-valor</b>	<b>lags</b>	<b>p-valor</b>	<b>lags</b>
<b>Campaña 1</b>	0.02	5	0.02	4	0.04	2	0.53	2
<b>Campaña 2</b>	0.67	1	0.12	2	0.89	1	0.02	1
<b>Campaña 3</b>	0.00	2	0.00	2	0.07	2	0.83	1

**Tabla 6-2.:** P-valores encontrados por medio de la metodología de Tsay, con sus respectivos rezagos.



campaña 3, sin embargo en el rezago 2 sólo los clicks e inversión en la campaña 3 consiguen tener valores altos de autocorrelación. La primera diferencia entonces soluciona el problema presente de altas correlaciones en rezagos de las series temporales del ejercicio práctico que se plantea en esta sección.

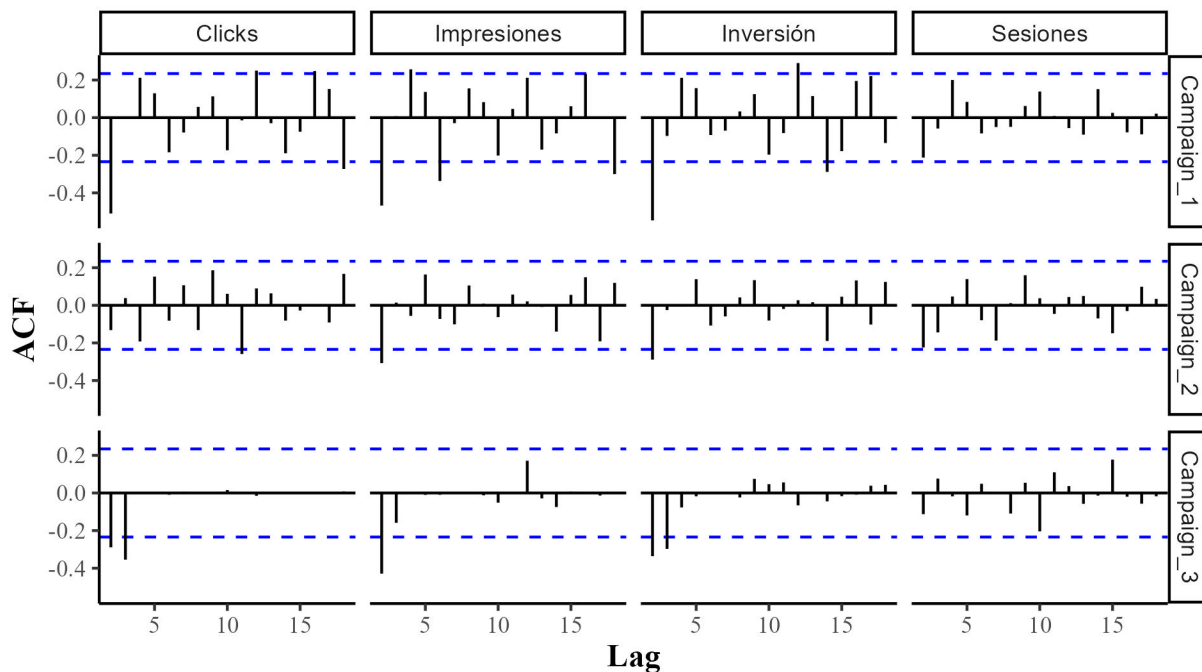


Figura 6-7.: ACF de cada serie en primeras diferencias, por campaña.

## 6.2. Modelamiento

Para el proceso de modelación, se ajustó el formato de los datos para tener un arreglo de acuerdo a lo consignado en Koop, Leon-Gonzalez y R. Strachan (2008), de manera que se agrupen los datos por canal, variable y fecha. Debido a que se trata de un modelo de datos longitudinales, mantener el orden es clave para el ajuste final de los parámetros y la inferencia posterior.

Cada una de las matrices relevantes de las representaciones de la verosimilitud en sus dos formas tal cómo se plantean en 4-5 y 4-7, y las matrices relevantes para el cálculo de las distribuciones que hacen parte del proceso de muestreo de la sección 4.4 se calcularon por

medio de funciones de elaboración propia, consignadas en el GitHub correspondiente del proyecto. Se buscó optimizar al máximo el proceso de estimación de las matrices inversas haciendo uso de referencias y evitando al máximo prácticas poco óptimas como la repetición de código y la redundancia de procesos similares.

Para el establecimiento de las distribuciones previas, se usaron las sugeridas por los autores en las distribuciones de  $\nu$  y  $\tau$ , respectivamente serían:  $\underline{\mu}_\tau = 5, \underline{\nu}_\tau = 15, \underline{\mu}_\nu = 21, \underline{\nu}_\nu = 42$ . Para los  $\mathbf{b}^*$ , se estableció una previa no informativa para cada uno de los parámetros que conforma el vector, de manera que cada  $b_i^* \sim \mathbf{N}(0, 1000)$ . Se estableció de igual manera para cada uno de los elementos del vector  $\boldsymbol{\beta}^*$  de manera que  $\beta_i^* \sim \mathbf{N}(0, 1000)$ , y se aplica el mismo principio para los elementos del vector  $\boldsymbol{\alpha}$ .

Para la matriz de covarianzas de los  $\boldsymbol{\beta}^*$ , simplemente se establece como aparece en la ecuación 4-12. Se encuentra el mayor reto de este proyecto al momento de establecer la previa para la matriz  $\underline{\mathbf{V}}$ , que contiene las covarianzas de los  $\mathbf{b}$ , ya que en las pruebas preliminares debía tenerse especial cuidado con dos aspectos de esta matriz:

- En esta matriz, las filas y columnas que correspondan, en orden, a los valores del vector  $\boldsymbol{\alpha}$  deberán ser actualizados en cada iteración por medio de los valores computados en la ecuación 4-11. En el documento original no se hace una clara mención a este proceso en particular salvo el hecho de mencionarlo exclusivamente para el establecimiento de la previa.
- Se debe tener en cuenta que la inversa de esta matriz aparece en la distribución condicional completa de  $\nu$ , por tanto, sí por algún motivo el valor de  $\mathbf{b}^T \underline{\mathbf{V}}^{-1} \mathbf{b}$  resulta muy grande, el valor de  $\nu$  va a tender a 0 y los valores de varianza de  $\boldsymbol{\alpha}$  en  $\underline{\mathbf{V}}$  van a tender a 0 y generar un error en el iterador.

Teniendo en cuenta lo anterior, la matriz  $\underline{\mathbf{V}}$  se establece de manera no informativa por medio de  $\underline{\mathbf{V}} \sim \text{IW}(\mathbf{V}_V, T)$  definiendo  $\mathbf{V}_V$  como una matriz de tamaño  $Nn(k + r_i) \times Nn(k + r_i)$  con  $Nn(k + r_i)$  el número de elementos del vector  $\mathbf{b}$  cuyas entradas son iguales a 0.1, salvo en las filas y columnas correspondientes a los elementos de  $\boldsymbol{\alpha}$  definidos por la ecuación 4-11.

Para el espacio de cointegración,  $\mathbf{H}^g$ , se usó la matriz:

$$H^g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix}$$

Por simplicidad del modelo y el cálculo de los parámetros, se asume que para los 3 individuos en la muestra, existen 3 relaciones de cointegración únicas e independientes que relacionan las 4 variables del sistema. Esto también es motivado por medio del uso de la metodología de Johansen para identificar el número idóneo de relaciones de cointegración, donde adicionalmente se compara la log-verosimilitud y se aconseja hacer uso de 3 relaciones de cointegración.

# Relaciones de cointegración	Log-Verosimilitud
<b>0</b>	-2943.62
<b>1</b>	-2928.06
<b>2</b>	-2919.60
<b>3</b>	<b>-2916.96</b>

**Tabla 6-3.:** Resultados de la log-verosimilitud de acuerdo con el número de relaciones de cointegración usadas.

Teniendo en cuenta el establecimiento de los valores de los hiperparámetros y las distribuciones previas usadas para los conjuntos de parámetros, se establecieron 4 diferentes modelos donde se varía sólo el número de rezagos de cada modelo. Los resultados pueden observarse en la tabla **6-4** donde se calculan: Deviance Information Criteria (DIC), Watanabe - Akaike Information Criteria (WAIC), Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC) con el fin de identificar el mejor modelo entre los 4 calculados. Para mayor información con respecto al cálculo de estos criterios, se recomienda remitirse al libro de Gelman et al. (2004).

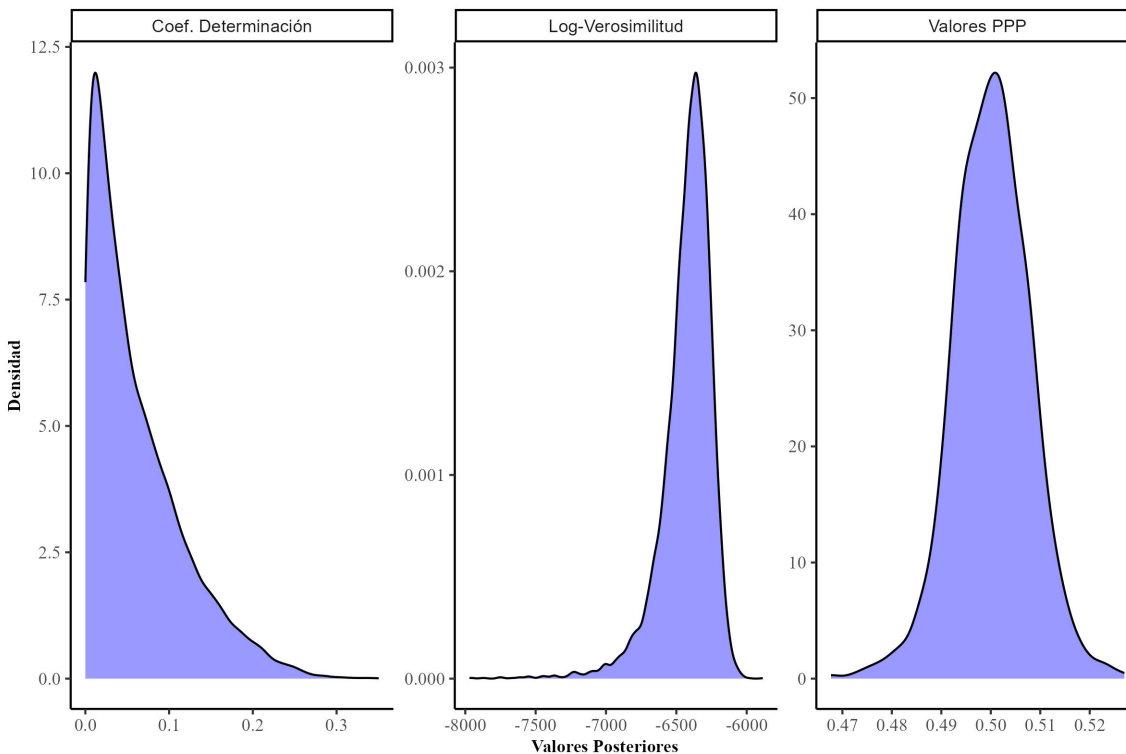
Según los resultados observados, los criterios de información DIC, WAIC y AIC recomiendan el uso del modelo con 4 rezagos, mientras que el BIC recomienda el uso de 1 rezago. De acuerdo con este resultado, se hará inferencia sobre el modelo con 4 rezagos en la siguiente sección.

Criterio de Información	1 rezago	2 rezagos	3 rezagos	4 rezagos
DIC	13968.98	13904.54	13924.19	<b>13903.30</b>
WAIC	12553.57	12337.99	12163.61	<b>11908.18</b>
BIC	<b>13210.60</b>	13314.67	13458.47	13519.73
AIC	12749.57	12629.99	12551.61	<b>12392.18</b>

**Tabla 6-4.:** Criterios de información para evaluar el número de rezagos adecuados para el modelo. Elaboración propia.

### 6.3. Resultados Finales

A la hora de evaluar los resultados posteriores del modelo, lo primero que es posible calcular son los estadísticos posteriores del modelo, cómo los valores PPP, coeficiente de determinación y log-verosimilitud del modelo.



**Figura 6-8.:** Estadísticos posteriores del modelo.

La ventaja de la estimación Bayesiana, es que es posible tener una distribución de los estadísticos que son calculados en función de otros. En este caso, en el panel izquierdo es posible ver la distribución de los coeficientes de determinación del modelo, calculados a par-

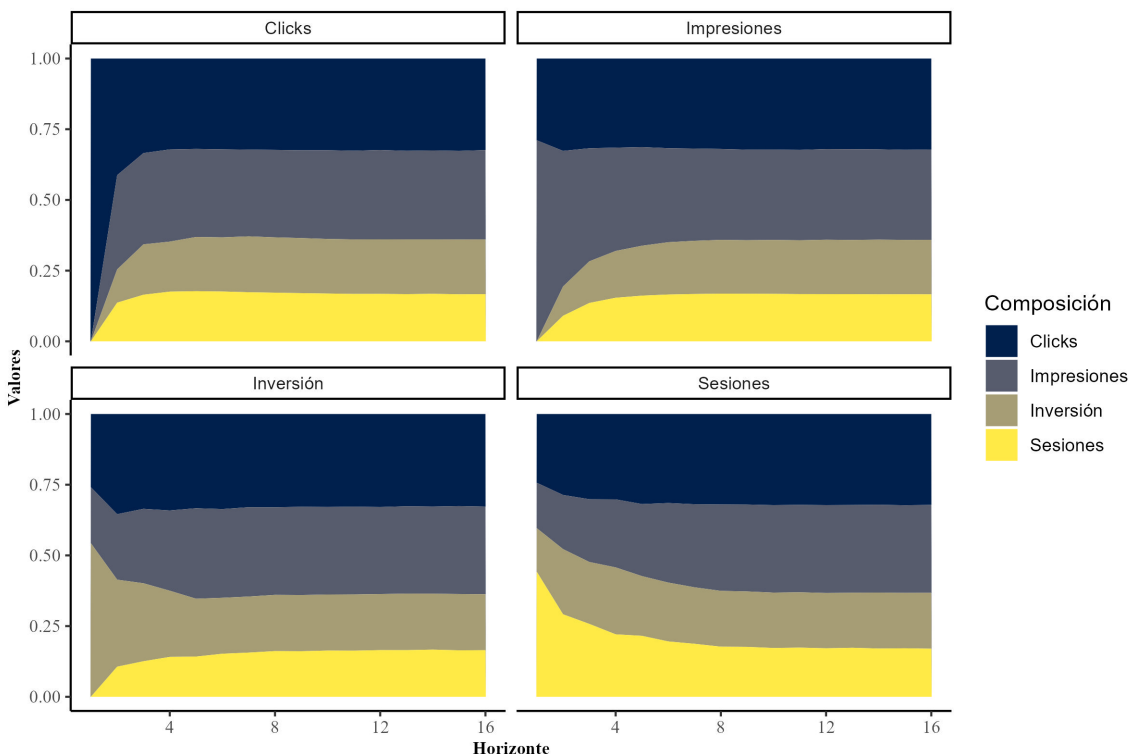
tir de los errores del modelo. Es posible ver que el modelo tiene un desempeño pobre a la hora de capturar la variabilidad de la variable de interés  $\Delta y_t$  ya que la mayor parte de los valores calculados con los parámetros se concentran principalmente por debajo de 0.2.

Con respecto a los valores posteriores de la log-verosimilitud, se centran principalmente alrededor de -6500, con excepción de 4 valores cercanos a -10000, lo que significa que existen algunos valores atípicos en esta distribución. Sin embargo, la gran parte de la masa está concentrada en valores superiores a -8000 e inferiores a -6000.

Por último, se tienen los valores PPP o valores predictivos posteriores, y estos se encuentran siempre alrededor de 0.5, esto quiere decir que en el modelo, los datos reales son valores típicos de la distribución predictiva posterior. Esto se interpreta cómo la probabilidad de que los datos replicados sean más extremos que la data observada, y dado que son valores alrededor de 0.5, implica directamente que el modelo es consistente con los datos.

Un análisis popular que se suele hacer a la hora de evaluar los resultados de un modelo tipo VEC son las funciones de impulso respuesta y la descomposición del error de predicción, IRF y FEVD por sus siglas en inglés, respectivamente. En este caso, por ser un panel de tres individuos, se presenta este análisis por cada uno de los individuos, fijando el foco del análisis en los FEVD. Esto es principalmente debido a que los datos proporcionados por el cliente interesado en el análisis, no superan el umbral de 100 datos para garantizar una varianza baja de los parámetros, por lo cual las funciones de impulso respuesta no convergen y no es posible analizar el modelo de esta manera. Para calcularlos, se usó el paquete *bvartools* de R, que contiene funciones que facilitan el cálculo posterior de IRF y FEVD de modelos VEC Bayesianos, desarrollado por medio de una metodología planteada por los autores del documento que sirvió de inspiración para la realización de este proyecto.

Es importante resaltar que esta metodología permite la estimación de la matriz de parámetros  $\mathbf{\Pi}_i$  pero no la estimación de las matrices  $\beta_i, \alpha_i$  separadamente, impidiendo el análisis de las relaciones de largo plazo. Así mismo, no es posible llevar a cabo el análisis de IRF en este conjunto de datos debido a la baja longitud de los datos que favorece varianzas muy altas de los parámetros implicando que los intervalos de credibilidad en varios casos superen el umbral de  $-1$  y  $1$ .



**Figura 6-9.:** Descomposición de la varianza del error de predicción, canal 1.

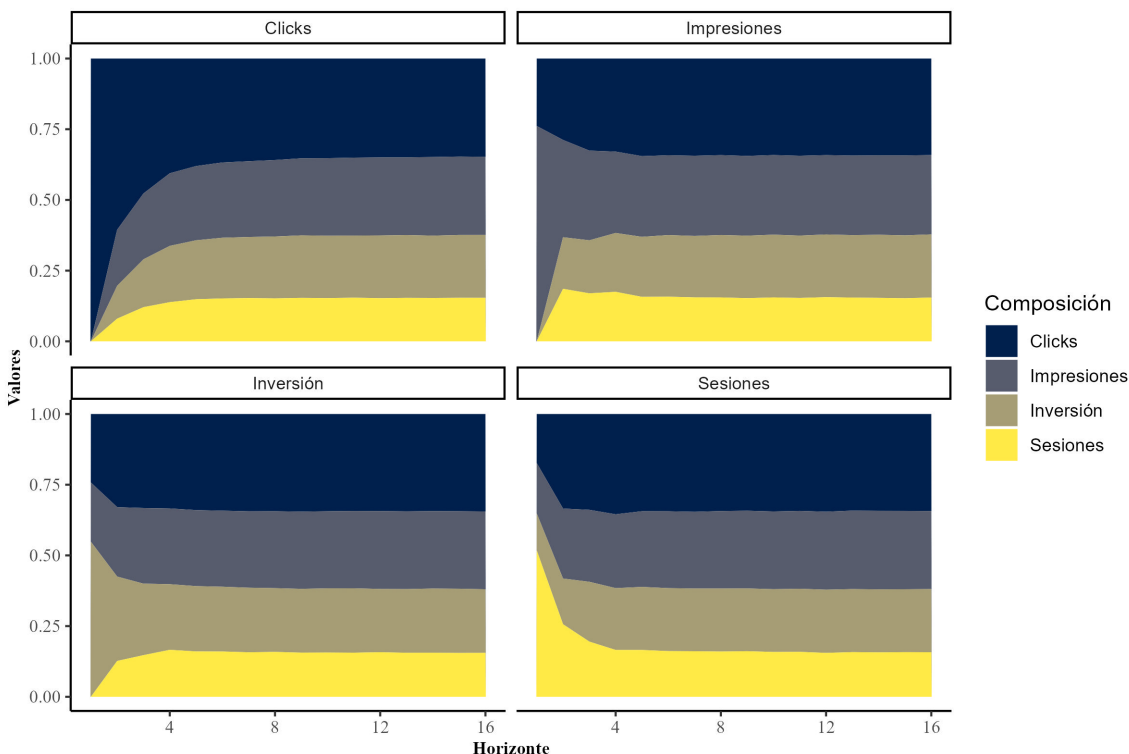
Para interpretar estos gráficos, se tiene que tener en cuenta que para cada uno de los 4 subgráficos se afecta una variable del sistema en una unidad; la variable afectada es la variable que titula al subgráfico. En ese orden de ideas, en la figura 6-9, es posible ver la FEVD del canal 1, donde los clicks y las impresiones predominan a lo largo de la descomposición de las 4 variables principales del sistema, sin embargo, los clicks mantienen su valor estable en el 25% de la explicación de la varianza de todas las variables, mientras que las impresiones en los primeros 4 períodos suelen tener una participación fluctuante entre 15% y 20%. Las sesiones de página web a través de las 4 variables del sistema suelen tener una participación inferior al 16%, salvo cuando esta es directamente afectada, e incluso, en este escenario converge a un valor de participación inferior al 21%.

Desde el punto de vista de negocio, tiene sentido que una vez se afecte la variable de inversión pase al menos un período de tiempo antes de que se vean las aportaciones en clicks e impresiones, ya que desde teoría de *marketing*, las inversiones suelen tener un efecto retardado que perdura en el tiempo, no obstante, es relevante entender cómo retroalimentan dichas

variables sobre la inversión, ya que una vez comienza una campaña de medios, el efecto bucle que se presenta al mantener un nivel estable de clicks e inversiones afectan las decisiones de *flighting* del mánager de la campaña. Respecto a las sesiones, al ser acciones correspondientes y llevadas a cabo por el usuario final que recibe la publicidad, son el objetivo final de las campañas; por lo cual, se halla el sentido a que todas las demás variables afecten la variable de más interés para el cliente, debido a que y tiene sentido que todas las demás variables del sistema terminen afectando la variable de más interés para el cliente debido a que se considera que para alcanzar una sesión adicional, el público objetivo debe pasar por el proceso de tener una impresión, hacer click y finalmente tomar una acción respecto al producto promocionado. Así mismo, el mayor número de sesiones a través de determinado tiempo, retroalimenta positivamente, ya que el cliente mantiene un nivel de inversión e incluso puede llegar a aumentarla para ver reflejado su efecto en las sesiones. Para el canal 1, la participación de sesiones si bien es baja, se estabiliza luego de 10 períodos o 2 meses según la temporalidad usada en este trabajo, siendo consistente con el período de adstock usado usualmente en modelos semanales de 10 a 12 períodos.

En la figura **6-10** se ven los FEVD del canal 2, en este canal dada la madurez de la campaña, es muy similar al primer canal con ligeras diferencias importantes. En el subgráfico de impresiones, la participación de la inversión a partir del período 2 crece notablemente, dando indicios de que las impresiones en este caso tienen una respuesta mucho más rápida frente a la inversión que en el canal 1, además de estabilizarse más rápidamente a partir del período 3. En el subgráfico de las sesiones, los porcentajes de contribución también se estabilizan más rápido en comparación con el canal 1, teniendo en cuenta también que las impresiones tienen menos participación en la generación de sesiones.

Desde la perspectiva de negocio, este canal tiene una generación de sesiones más de corto plazo, ya que a partir del período 4 se estabiliza la contribución de varianza de todas las 4 variables, por lo que el total del efecto sobre las sesiones es visible antes de que pase el mes posterior al período de inversión. Es notable que este canal tiene las contribuciones más equilibradas de los 3 analizados en este documento, teniendo en cuenta que nuevamente los clicks tienen la mayor participación a través de las diferentes variables, y además tanto



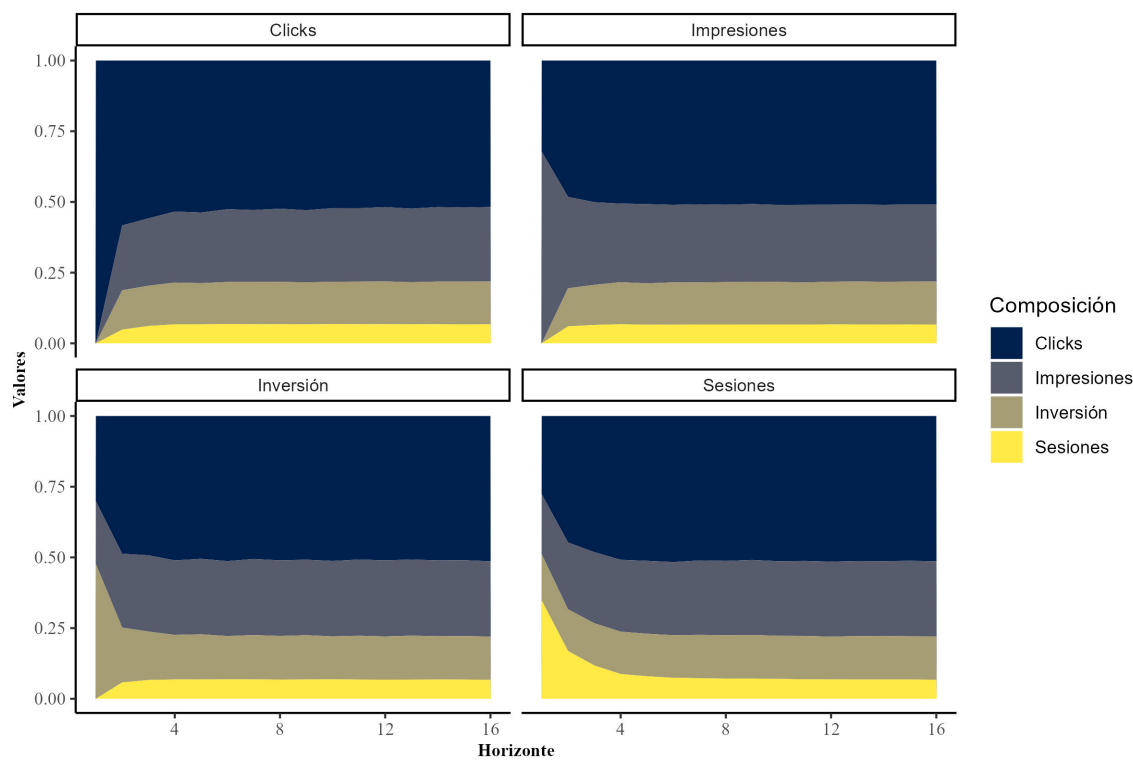
**Figura 6-10.:** Descomposición de la varianza del error de predicción, canal 2.

sesiones como inversión pesan poco menos del 50% de participación en la contribución, mientras que en el canal 1 pesan menos del 40% del total de la contribución total.

En la figura 6-11 se consolidan los resultados del canal 3, en este canal los clicks pesan casi el 50% de la descomposición de la varianza a lo largo de las 4 variables del sistema. En este caso, las sesiones pesan menos del 10% consistentemente a lo largo de todo el período analizado salvo en el caso donde estas mismas son impactadas teniendo en cuenta que posterior al período 7 se estabilizan en valores inferiores al 10%. Las impresiones alcanzan a pesar un poco menos del 27% una vez estabilizada la contribución, con la inversión pesando poco más del 15%. Es claro que para este canal, los clicks son vitales para la explicación de la variación de las demás variables del sistema, diferenciándolo claramente de los dos canales evaluados previamente.

Analizándolo desde una perspectiva de negocio, es consistente con la experiencia la contribución que se tiene por variable ya este canal es el más joven de los 3, debido a que se trata de campañas de inversión experimental que usan estrategias de interacción menos convencio-





**Figura 6-11.:** Descomposición de la varianza del error de predicción, canal 3.

nales. En primer lugar, se esperaba desde un principio que las sesiones en este canal fueran las menores en términos de aportación entre los 3 canales. Así mismo, en el subgráfico de sesiones la contribución se estabiliza a partir del período 7, mostrando que en términos de convergencia tiene un plazo intermedio entre el individuo 1 (que presenta el mayor período de estabilización) y el individuo 2 (con el menor período de estabilización). En el caso de la inversión, la retroalimentación se ve más fuertemente por el lado de clicks, por lo que la estrategia de campaña a pesar de tener metas de conversiones por parte de sesiones, tiene más metas de alcanzar un público que interactúe, con el fin de dar a conocer la marca, que un público que realice conversiones en forma de sesiones y esto se confirma también en la gran participación de las impresiones en la dinámica del modelo.

Avanzando hacia un análisis de convergencia de las cadenas, se muestran los tamaños efectivos de muestra en el gráfico **6-12** donde se puede ver que la mayoría de parámetros están por encima de la línea horizontal en 10.000 que representa el total de simulaciones, con excepción de  $\nu$ ,  $\tau$  y algunas muestras de la matriz de varianzas y covarianzas. Es notable que

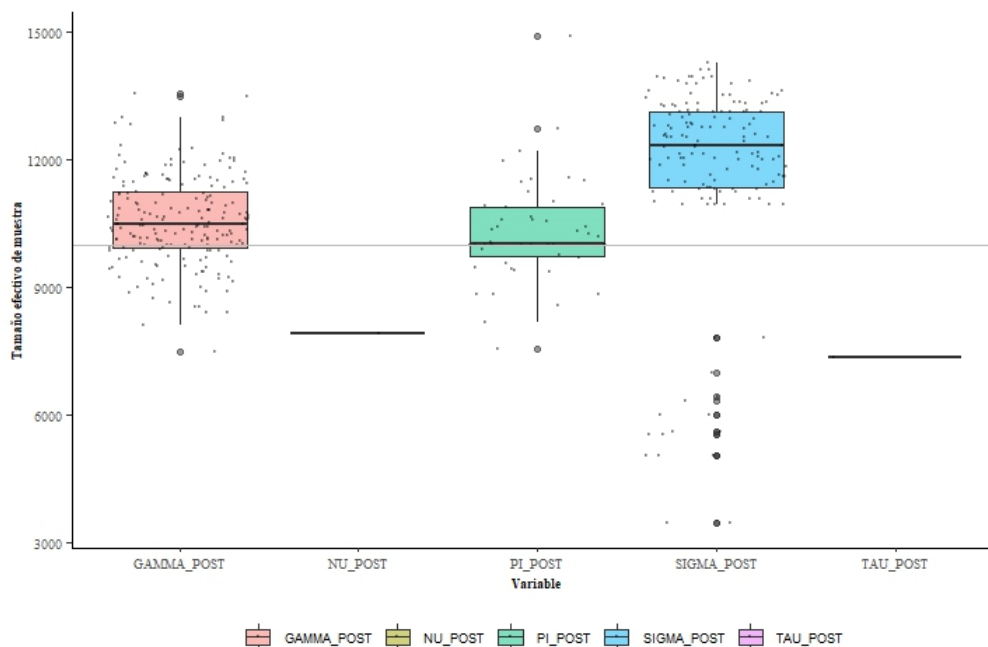


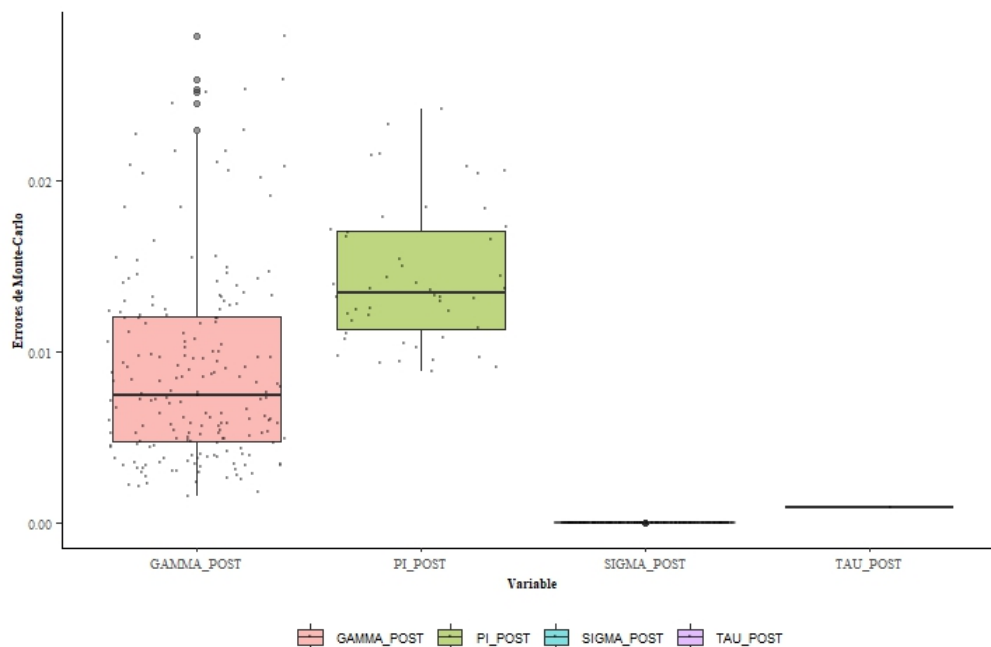
Figura 6-12.: Tamaños efectivos de muestra de los parámetros.

los tamaños de muestra son mejores en los datos reales para los muestreos de  $\Gamma$  y  $\Sigma$ , pero  $\nu$  tiene el mismo comportamiento que el mostrado en la simulación corta.

Analizando los errores de Monte Carlo consignados en la figura 6-13, los valores están en su mayoría por debajo de 0.05, a diferencia del valor de  $\nu$  que tiene un comportamiento similar en las simulaciones donde destaca con un valor elevado, que en este caso es de 3.12. Nuevamente los valores de  $\Sigma$  tienen los errores de Monte Carlo más bajos de las simulaciones y los valores de  $\Gamma$  tienen en general un error inferior a los valores de  $\Pi$ .

En este caso, el algoritmo converge sin problemas a la distribución estacionaria posterior, por lo que no es necesario aplicar técnicas de muestreo sistemático como el *thinning* o adelgazamiento de las cadenas debido a que no se encuentra evidencia de una correlación entre las muestras del algoritmo.

El modelo también se puede comparar con la estimación frecuentista de un modelo de tipo VEC. En este caso particular, se hizo uso del software estadístico E-Views para calcular los coeficientes del modelo y compararlos con el modelo ajustado en esta metodología. Los resultados están consignados en la tabla 6-5 y se puede ver que los errores calculados entre los valores de los parámetros del modelo frecuentista y los valores promedio de las cadenas



**Figura 6-13.:** Errores de Monte Carlo de los parámetros, omitiendo  $\nu = 3.12$  para efectos de visualización.

Medida	Modelo Bayesiano vs. Frecuentista
<b>RMSE</b>	0.009
<b>MAE</b>	0.016
<b>MAPE</b>	0.027
<b>Sesgo</b>	$-5e^{-4}$

**Tabla 6-5.:** Diferencia entre el modelo estimado por medios frecuentistas vs. metodología Bayesiana.

posteriores de los parámetros del modelo Bayesiano son bajos, destacando el RMSE con un valor de 0.009. Esto muestra que los parámetros del modelo Bayesiano están muy cerca del modelo frecuentista, contando con los beneficios de la metodología Bayesiana como la cuantificación de la incertidumbre asociada con cualquier función de los parámetros.

## 7. Conclusiones y recomendaciones

En el presente documento se realizó y comprobó la efectividad del modelo panel Bayesiano cointegrado presentado en Koop, Leon-Gonzalez y R. Strachan (2008), analizando la metodología en detalle, donde se realizaron diferentes aclaraciones sobre la actualización de parámetros del modelo y algunas aproximaciones y pruebas haciendo uso de tres diferentes simulaciones que buscaban medir la efectividad, precisión y eficiencia de la ejecución del modelo. En particular cabe resaltar que el modelo debido a su tamaño y estructura, puede tomar varios minutos de ejecución a varias horas dependiendo de la longitud de los paneles variando principalmente dependiendo de la cantidad de períodos observados  $t$ , sin embargo, se mostró que con menos de 100 observaciones, el modelo tiende a tener varianzas muy altas en las distribuciones posteriores de los parámetros. Es notable que la mayoría de parámetros del modelo (salvo  $\nu$ ) convergen y tienen errores de Monte Carlo muy pequeños a lo largo de las simulaciones presentadas.

En el estudio de caso realizado con la marca de alimentos latinoamericana, de acuerdo con los resultados observados en la sección 6 y contrastando con los ejercicios de simulación de la sección 5, se tiene una valiosa oportunidad para mostrarle al cliente que suministró los datos los efectos que tiene sobre este modelo el no contar con observaciones suficientes, ya que, si bien fue posible graficar e interpretar los FEVD, hubiese sido conveniente también poder contar con las IRF que no fueron posibles de hacer debido a la alta varianza de los parámetros provocada por la baja cantidad de observaciones suministrada. Así mismo, para otros clientes que se encuentren interesados en realizar estudios similares para sus marcas a través de canales digitales o geografías donde tengan presentes sus productos, se recomendará contar con al menos 120 observaciones con el fin de tener los mejores entregables posibles.

Parte de una investigación futura, es plantear un estudio que integre los conocimientos de márketing sobre la saturación de medios de inversión, esto es: incorporar las nociones de saturación de canales por campañas de inversión demasiado agresivas; tener en cuenta además de las impresiones, la cantidad de público objetivo a la que llega la pauta y la frecuencia promedio con la que un público ve o tiene exposición a la campaña o al anuncio en cuestión; y finalmente tener en cuenta las audiencias de alto valor proporcionadas por Google para poder segmentar los clientes y hacer análisis por poblaciones para agregar granularidad al estudio. Otra línea de estudio valiosa es hacer análisis de causalidad por individuo del panel, que en este caso particular se dificulta por las altas varianzas de los parámetros y es conveniente tener más observaciones, así mismo, probar diferentes estructuras de correlación que representen mejor la naturaleza de los datos temporales para poder solventar el supuesto de la existencia únicamente de covarianza 0 en tiempos diferentes. Además, vale la pena explorar la robustez del modelo a la inclusión de un mayor número de variables por medio de un ejercicio exhaustivo de simulación.

Otra pauta de investigación relevante es poder: calcular la saturación por individuo de diferentes métricas de inversión, evaluando el costo por clicks, el costo por mil impresiones o directamente evaluando el coste agregado de tener un público base o clicks e impresiones orgánicas. Desde un punto de vista teórico el reto sería plantear funciones de saturación que sean acordes con la estructura e ideal de las campañas de inversión.

## A. Anexo: Resultados gráficos de las simulaciones

Simulación	Parámetro	Proporción	RMSE	MAE	Longitud Promedio	Sesgo
Simulación Corta	Gamma	35.41 %	0.221	0.190	<b>0.278</b>	-0.032
Simulación Moderada		91.66 %	<b>0.010</b>	0.121	0.561	<b>-0.001</b>
Simulación Extrema		<b>93.75 %</b>	0.020	<b>0.050</b>	0.297	-0.002
Simulación Corta	Pi	89.58 %	0.041	0.025	0.128	-0.005
Simulación Moderada		<b>97.91 %</b>	0.075	0.087	0.684	0.010
Simulación Extrema		<b>97.91 %</b>	<b>0.026</b>	<b>0.019</b>	<b>0.103</b>	<b>-0.003</b>

Tabla A-1.: Evaluación de la precisión de las simulaciones.

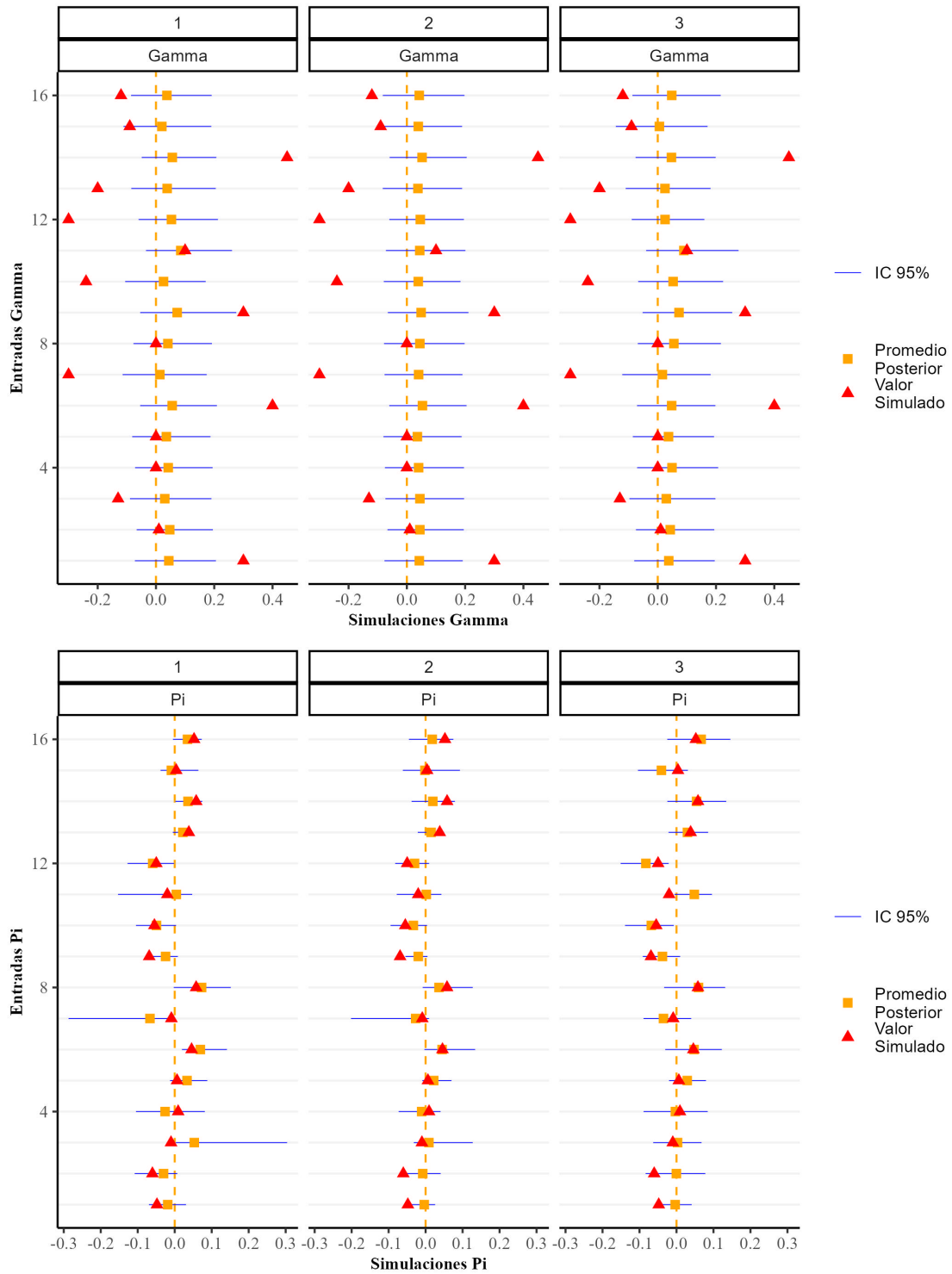


Figura A-1.: Resultados Caso 1 Simulación: Comparación posteriores con valores simulados.

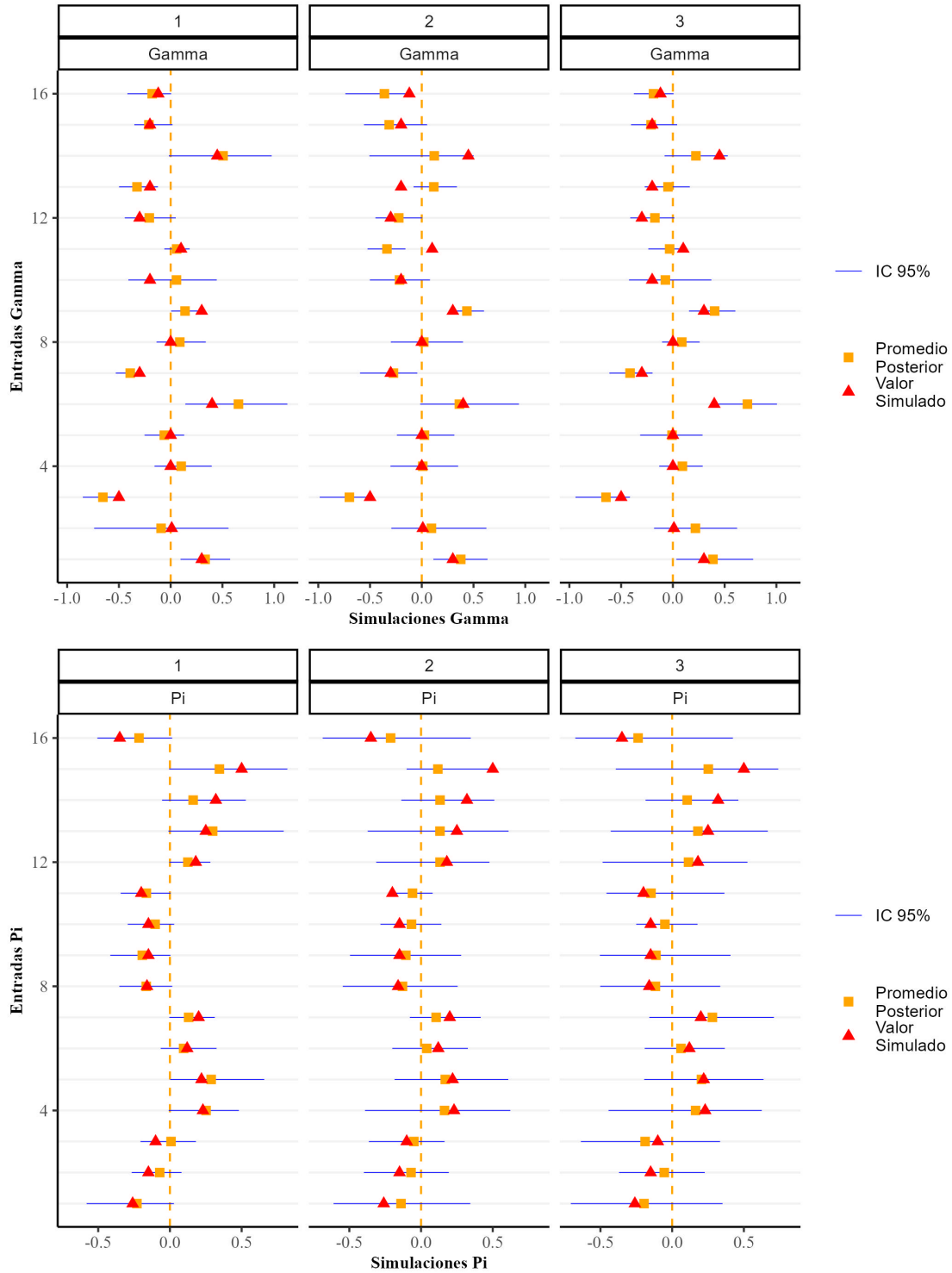


Figura A-2.: Resultados Caso 2 Simulación: Comparación posteriores con valores simulados.



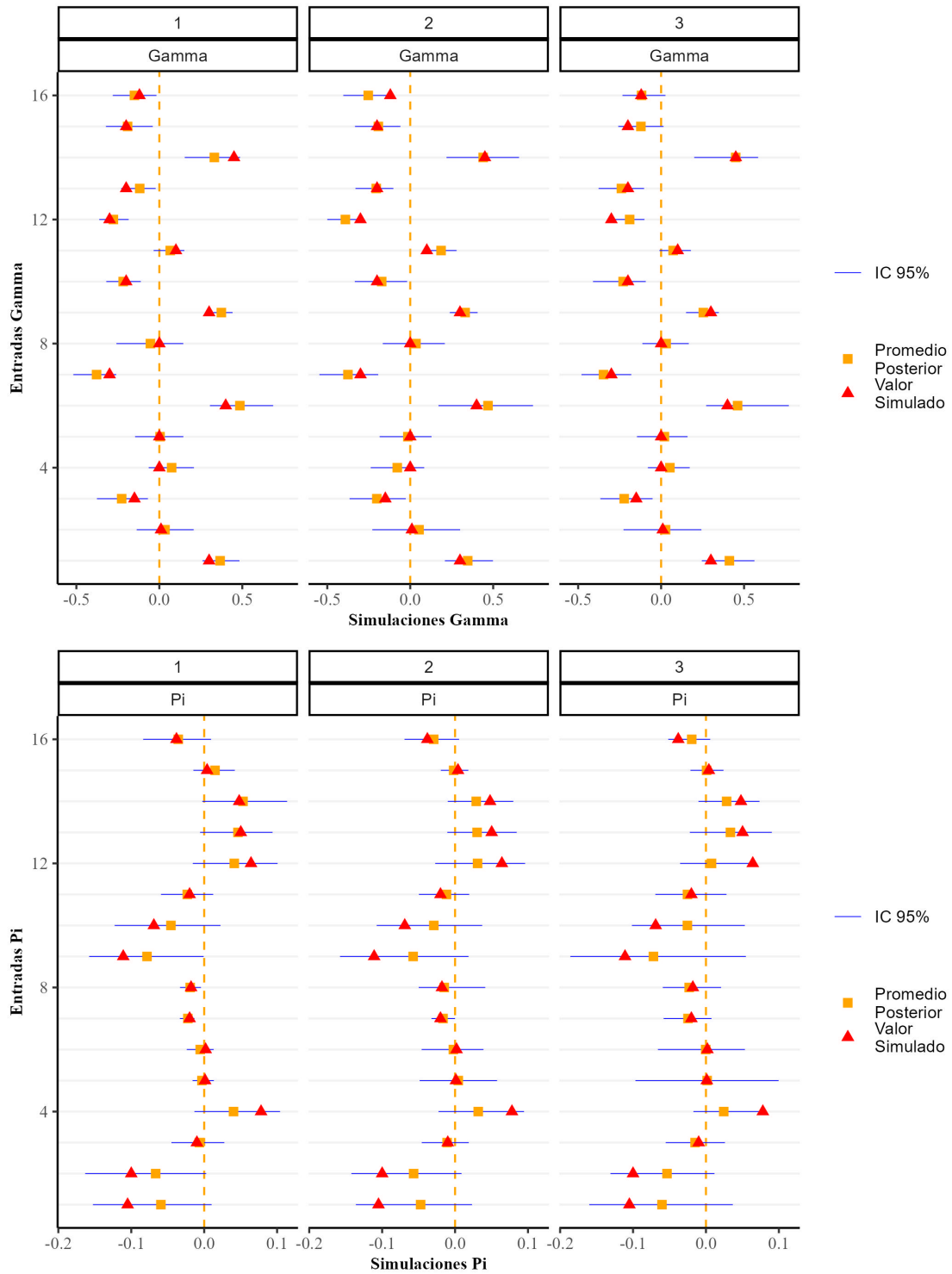


Figura A-3.: Resultados Caso 3 Simulación: Comparación posteriores con valores simulados.

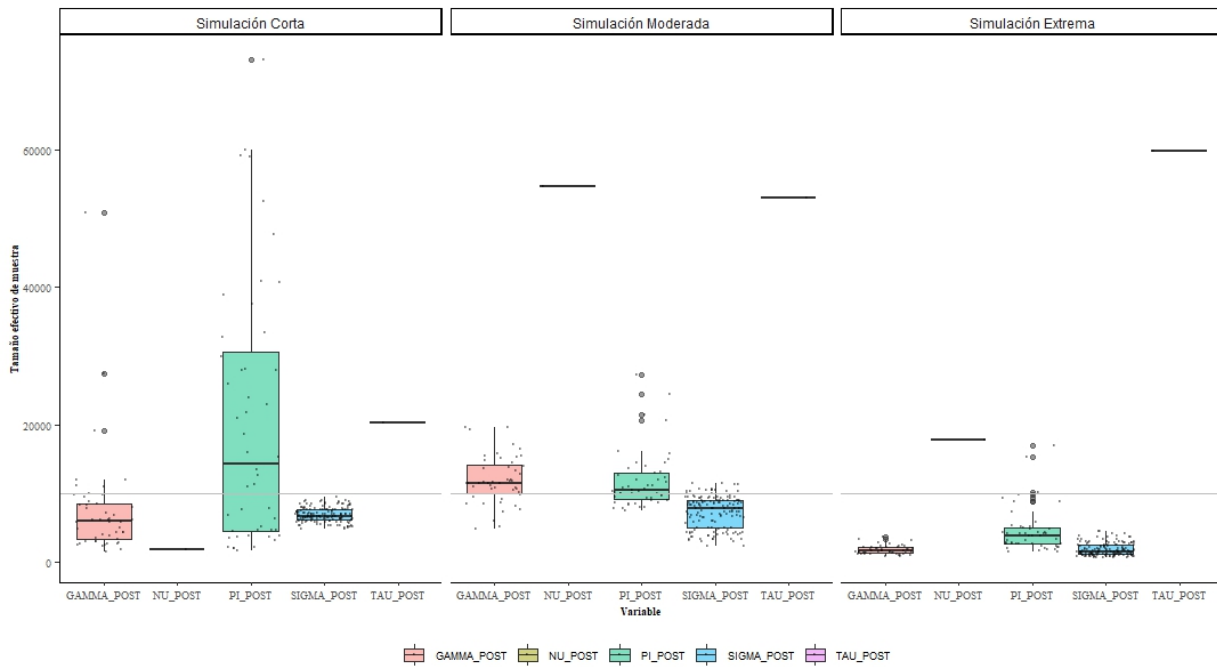


Figura A-4.: Tamaños Efectivos de Muestra, todas las simulaciones.

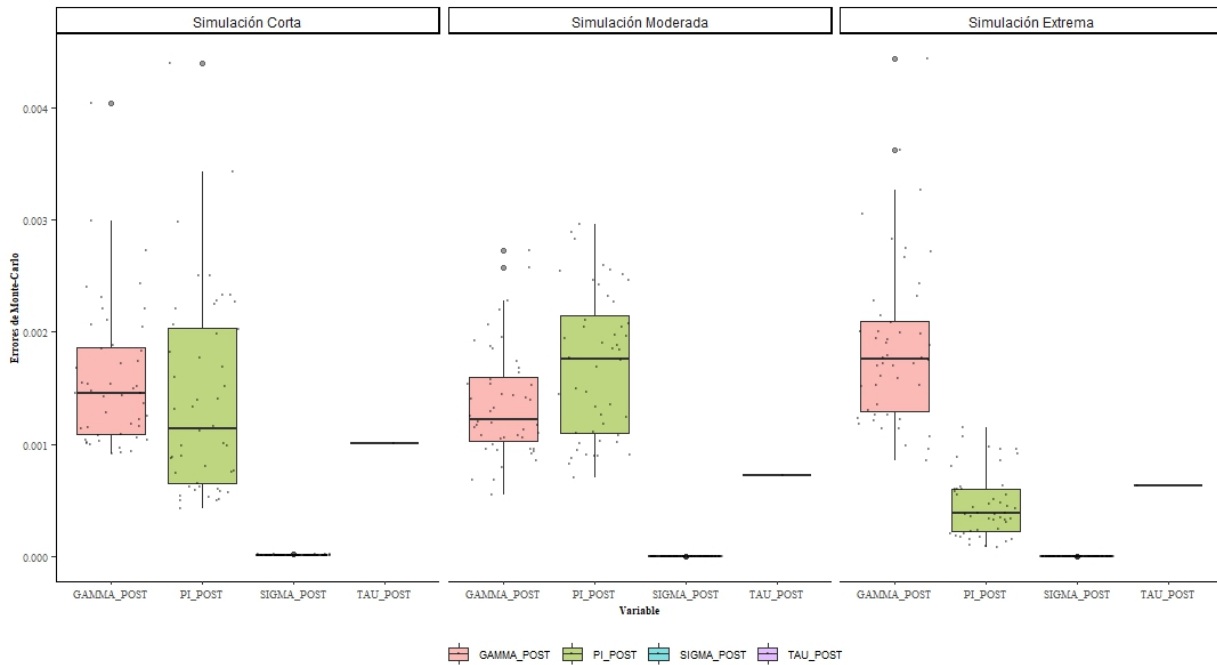


Figura A-5.: Errores de Monte Carlo, todas las simulaciones. . **NOTA:**  $\nu$  descartado por efectos de visualización, EMC = 123, 32, 57 respectivamente

# Bibliografía

- Blanco, Liliana, Viswanathan Arunachalam y Delvamuthu Dharmaraja (2012). *Introduction to probability and stochastic processes with applications*. 1.<sup>a</sup> ed. John Wiley: New Delhi.
- Brockwell, Peter J. y Richard A. Davis (2006). *Time Series: Theory and Methods*. 2.<sup>a</sup> ed. Springer: New York.
- Casella, George y Edward I. George (1992). «Explaining the Gibbs Sampler». En: *American Statistical Association* 46, págs. 164-174. DOI: 10.2307/2685208.
- Chigozie Nkalu, et. al (2020). «Financial Development and Energy Consumption in Sub-Saharan Africa: Evidence from Panel Vector Error Correction Model». En: *SAGE*, págs. 1-12.
- Dekimpe, Marnik G. y Dominique M. Hanssens (1995). «Empirical Generalizations about Market Evolution and Stationarity». En: *Marketing Science* 14, G109-G121. DOI: 10.2307/184153.
- Diggle, Peter J. et al. (2002). *Analysis of Longitudinal Data*. 2.<sup>a</sup> ed. Oxford Statistical Science Series: Great Britain.
- Engle, Robert F. y C. W. J. Granger (mar. de 1987). «Co-Integration and Error Correction: Representation, Estimation and Testing». En: *Econometrica* 55, págs. 251-276. DOI: 10.2307/1913236.
- Gamerman, Dani y Hedibert Freitas Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC.
- Gelman, Andrew et al. (2004). *Bayesian Data Analysis*. 2.<sup>a</sup> ed. Chapman & Hall CRC: United States.

- Hoff, Peter D. (2009). *A first course in bayesian statistical methods*. 1.<sup>a</sup> ed. Springer: New York.
- Keller, Kevin Lane (1993). «Conceptualizing, Measuring, and Managing Customer-Based Brand Equity». En: *Journal of Marketing* 57, págs. 1-22. DOI: 10.1177/002224299305700101.
- Kleibergen, Frank y Herman K. van Dijk (oct. de 1994). «On the shape of the likelihood/posterior in cointegration models». En: *Econometric Theory* 10, págs. 514-551.
- Kleibergen, Frank y Richard Paap (2002). «Priors, posteriors and bayes factors for a Bayesian analysis of cointegration». En: *Journal of Econometrics* 111, págs. 223-249. DOI: 10.1016/s0304-4076(02)00105-7.
- Koop, Gary, Roberto Leon-Gonzalez y Rodney Strachan (2008). «Bayesian inference in a cointegrating panel data model». En: *Bayesian Econometrics* 23(3), págs. 433-469.
- Koop, Gary, Roberto Leon-Gonzalez y Rodney Strachan (2011). «Bayesian inference in a time varying cointegration model». En: *Journal of Econometrics* 165, págs. 210-220. DOI: 10.1016/j.jeconom.2011.07.007.
- Kumar, Sabuj y S. Madheswaran (2010). «Causality between energy consumption and output growth in the Indian cement industry: An application of the panel vector error correction model (VECM)». En: *Energy Policy* 38, págs. 6560-6565.
- Lütkepohl, Helmut (2005). *New Introduction to multiple Time Series Analysis*. 1.<sup>a</sup> ed. Springer Verlag: Berlin, Germany.
- Lütkepohl, Helmut y Markus Krätzig (2004). *Applied Time Series Econometrics*. 1.<sup>a</sup> ed. Cambridge University Press: Cambridge, UK.
- Mahadevan, Renuka y John Asafu-Adjaye (2007). «Energy consumption, economic growth and prices: A reassessment using panel VECM for developed and developing countries». En: *Energy Policy* 35, págs. 2481-2490.
- Máté, Andor, Sarolta Somosi y Beata Farkas (sep. de 2022). «Co-movement between the convergence of economic and health status in Central and Eastern Europe». En: *Acta Oeconomica* 72, págs. 351-365. DOI: 10.1556/032.2022.00026.
- Pedroni, Peter (1997). «Panel cointegration. Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis - New Results». En:

*Cambridge University Press*. URL: <https://web.williams.edu/Economics/pedroni/pc-revc.pdf>.

Slotegraaf, Rebecca J. y Koen Pauwels (2008). «The impact of brand Equity and Innovation on the Long-Term Effectiveness of Promotions». En: *Journal of Marketing Research* 45, págs. 293-306. DOI: 10.2307/30162532.

Srinivasan, Shuba y Frank M. Bass (2000). «Cointegration analysis of brand and category sales: Stationarity and long-run equilibrium in market shares». En: *Applied stochastic models in bussiness and industry* 16, págs. 159-177. DOI: [https://doi.org/10.1002/1526-4025\(200007/09\)16:3%3C159::AID-ASMB417%3E3.0.CO;2-R](https://doi.org/10.1002/1526-4025(200007/09)16:3%3C159::AID-ASMB417%3E3.0.CO;2-R).

Strachan, Rodney W. y Brett Inder (2004). «Bayesian analysis of the error correction model». En: *Journal of Econometrics* 123, págs. 307-325. DOI: 10.1016/j.jeconom.2003.12.004.

Watanabe, Sumio (2013). «A Widely Applicable Bayesian Information Criterion». En: *Journal of Machine Learning Research*.