



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

PREDICCIÓN DE LA ACTIVIDAD ECONÓMICA EN
MÉXICO MEDIANTE INFORMACIÓN DE LOS SISTEMAS
DE PAGOS

T E S I S

PARA OBTENER EL TÍTULO DE:

ACTUARIO

PRESENTA:

JULIO CÉSAR MARENTES MOSQUEDA

DIRECTORA DE TESIS:

DRA. RUTH SELENE FUENTES GARCÍA

Ciudad de México, México. 2022





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

1. Datos del alumno

Marentes
Mosqueda
Julio César
(55) 65 26 97 75
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
417098364

2. Datos del tutor

Dra.
Ruth Selene
Fuentes
García

3. Datos del sinodal 1

Dra.
Lizbeth
Naranjo
Albarrán

4. Datos del sinodal 2

Dr.
Othón Martino
Moreno
González

5. Datos del sinodal 3

M. en C.
Jaime
Vázquez
Alamilla

6. Datos del sinodal 4

Act.
Orlando
Luisillo
Ramírez

7. Datos del trabajo escrito.

Predicción de la actividad económica en México mediante información de los sistemas de pagos
87 p
2022

Para mis padres, para Anna y para Bruno

Agradecimientos

A mi mamá, que has dedicado toda su vida a mis hermanos y a mí. Nos has brindado tu amor, tu comprensión, tu tiempo, tu esfuerzo, tu perseverancia y tu ejemplo. Siempre me has apoyado escuchándome, dándome excelentes consejos, brindándome tu cariño incondicional. No hubiese logrado nada sin tu apoyo, sin diariamente apoyarme en todas las cosas que necesitaba. Perdón por mis fallos, tú siempre serás mi mayor juez. Nunca te va a faltar nada, tienes a 3 ángeles que te cuidamos aquí en vida. Todo mi agradecimiento, obediencia, respeto y amor para ti, Aita.

A mi papá, Aito, que donde quiera que estés, siempre te llevo en mi mente. Sé que estarías muy feliz de poder verme terminar esta etapa. Siempre nos fomentaste la sed de investigar, de descubrir, de aprender cosas nuevas; el gusto por la ciencia y la tecnología siempre fue debido a ti. Gracias por dejar bien preparada esta familia para que pudieras partir en paz; sé que siempre intentaste dar lo mejor por nosotros, las cosas siempre pasan para algo y nos dejas miles de enseñanzas. Mi eterno amor para ti.

A mis hermanos, que siempre me han apoyado, amado, escuchado, divertido y aconsejado. Mi hermana Anna siempre con sus palabras exactas y su visión clara y realista de las situaciones. Mi hermano Bruno, mi eterno compañero, dándome grandes enseñanzas de vida aún con tu corta edad. Gracias a los dos por venir a este mundo a acompañarme. Prometo nunca dejarlos solos, somos un equipo indestructible; dicen por ahí: “el trío maravilla”. Los amo por siempre.

A Andrea, mi amada compañera de vida. Sólo tú y yo sabemos la montaña rusa de emociones y sentimientos, historias, sueños, aprendizajes y experiencias que conforman nuestra historia. Gracias por tantas pláticas interesantes, datos curiosos, debates, deconstrucciones de obras de arte y sueños. Gracias por soportarme y apoyarme en todo este viaje. Todos los días me inspiras a ser una mejor persona, personal y profesionalmente. Siempre querré lo mejor para ti. *Ojalá.*

A Magnucito, siempre cuidándonos y brindándonos su amor, fidelidad y alegrías. Nos has acompañado en las buenas, en las malas y en las peores. Me has enseñado demasiado.

A mis queridos profesores. Se que no es fácil lidiar con un alumno como yo: todo el tiempo preguntando y deteniendo la clase, platicando con el de a lado, o a veces no asistiendo a la sesión. Mi más sincera muestra de agradecimiento por el tiempo que dedican a cada uno de sus alumnos. Por mencionar a algunos: Dra. Ruth, que usted sabe que es mi segunda madre, Dra. Clarita Fittipaldi, sus clases me encantan; aun cuando se hace una ensalada mental, Dr. Manuel Domínguez, me adentró en este magnífico mundo de la probabilidad, Lalo Selim, tus clases son de otro nivel, Yadi, Struck, Labardini, entre otros.

A la UNAM y al IPN, mis dos alma maters, las llevo siempre en el corazón.

Introducción

En el análisis y seguimiento de la política económica, la cual está conformada por la política fiscal y monetaria, no hay variable que reciba mayor interés como el que recibe el Producto Interno Bruto (PIB). El cual, está definido como el total de los bienes (mercancías o productos tangibles), servicios (actividades intangibles que satisfacen las necesidades) e inversiones que se producen en un país en un intervalo de tiempo determinado (usualmente un año). Es por lo anterior, que las variaciones del PIB reflejan el crecimiento del país en cuestión a lo largo del tiempo. En México este valor es publicado de forma trimestral por el Instituto Nacional de Estadística y Geografía (INEGI) tras una ardua recopilación y procesamiento de información macroeconómica. Aunque este proceso es necesario, puede llegar a considerarse para los actuarios y economistas un problema, dado que la publicación tiene un rezago de más de 5 semanas al momento del cierre del trimestre. Tener una estimación correcta y oportuna del PIB beneficia a las instituciones financieras en su toma de decisiones.

En este documento se presenta al lector un modelo de predicción del PIB de México utilizando la información de dos de los sistemas de pagos electrónicos más importantes que operan en el país: el Sistema de Pagos Electrónicos Interbancarios (SPEI) y el mercado de Tarjetas. El SPEI, el sistema de pagos más importante en el país, ha sido de gran utilidad para que el sistema financiero mexicano funcione correcta y ordenadamente en todos los estratos sociales, es decir, en él se liquidan todas las operaciones realizadas por la población general y además también se liquidan las operaciones realizadas por instituciones gubernamentales y empresas. Lo anterior, nos da más claridad de porqué los sistemas de pagos electrónicos han ayudado a que la economía de México crezca, se refuerce y tenga un impacto mayor en la sociedad; permitiéndole que tenga acceso a todos los beneficios que otorga pertenecer al sistema financiero como lo son: créditos, seguros, pensiones, etcétera.

Se estudia la relación entre las variables de pagos con el PIB para así poder definir el modelo predictivo; se realiza un proceso iterativo pronosticando desde el primer trimestre del 2017 al cuarto trimestre del 2021 y por último se presentan métricas de error para que el lector pueda compararlo contra otros modelos. Cabe señalar que es imposible ignorar la crisis económica del 2020: la pandemia causada por el virus SARS-CoV 2; se mencionan puntos relevantes respecto al tema y como afectó al modelo en cuestión.

Índice general

Agradecimientos	II
Introducción	III
1 Marco Teórico	1
§1.1 Introducción a la Estadística Bayesiana	1
§1.1.1 Teorema de Bayes	1
§1.1.2 Ejemplo	2
§1.1.3 Familias conjugadas	3
§1.2 Regresión lineal	6
§1.2.1 Estimación de parámetros	8
§1.3 Regresión Lineal, enfoque Bayesiano	12
§1.3.1 Distribución final de los parámetros	12
§1.3.2 Distribución no informativa de Jeffreys	20
§1.4 Cointegración	23
§1.4.1 Caso Multivariado. Prueba de Johansen	25
2 Marco Conceptual	28
§2.1 Producto Interno Bruto en México	28
§2.1.1 PIB Trimestral	28
§2.2 Sistemas de pagos en México	29
§2.2.1 SPEI	32
§2.2.2 Tarjetas	33
3 Metodología	35
§3.1 Relación entre PIBT e información de los Sistemas de Pagos	35

§3.1.1	Preparación de la información	36
§3.1.2	Selección de variables, estacionariedad y cointegración	39
§3.2	Modelado	49
§3.2.1	Estimación de distribuciones finales	50
§3.3	Predicción	54
§3.3.1	Predicción fuera de la muestra	56
§3.3.2	Reflexiones periodo confinamiento COVID-19	59
4	Conclusiones	64
§4.0.1	Casos Confirmados COVID 19	66
§4.0.2	ADF	67
§4.0.3	Modelos escogidos	72
§4.0.4	Gráficos de predicciones	73

Índice de figuras

1.1	Ejemplo de estructura bayesiana de un modelo de una variable aleatoria normal con parámetro μ que proviene también de una variable aleatoria normal.	6
2.1	Producto Interno Bruto trimestral (PIBT), del primer trimestre de 1993 al cuarto trimestre de 2021. Millones de pesos a precios de 2013.	30
2.2	Datos relevantes de los sistemas de Pagos a nivel nacional al cierre de 2020. Figura tomada de la introducción del Programa Anual de Inclusión Financiera. [20] Fuente: CNBV.	31
2.3	Evolución del número de operaciones anuales en diversos medios de pago electrónico. Fuente: Panorama Anual IF CNBV y Banxico 2016-2020.	32
2.4	Número de operaciones SPEI TaT liquidadas al año. Fuente: Informe Anual sobre las Infraestructuras de los Mercados Financieros 2020 del Banco de México. [15] .	33
2.5	Proporción del Monto liquidado en operaciones SPEI TaT dividido por el PIB anual. Fuente: Banco de México e INEGI.	34
3.1	Monto transaccionado trimestralmente a través del SPEI en sus 3 grupos por monto. Primer trimestre de 2009 al cuarto trimestre del 2019. Miles de millones de pesos a precios de 2013.	37
3.2	Número de operaciones transaccionadas trimestralmente a través del SPEI en sus 3 grupos por monto. Primer trimestre de 2009 al cuarto trimestre del 2019.	38
3.3	Monto transaccionado trimestralmente con tarjeta de crédito y débito. Primer trimestre de 2009 al cuarto trimestre del 2019. Miles de millones de pesos a precios de 2013.	40
3.4	Número de operaciones transaccionadas trimestralmente con tarjeta de crédito y débito. Primer trimestre de 2009 al cuarto trimestre del 2019.	41
3.5	Correlación entre las variables de pagos y el PIBT.	42

3.6 Modelo seleccionado de 4 covariables que expliquen el PIBT. 49

3.7 Densidad posterior de $\tau \sim Gamma(\alpha_1, \delta_1)$ y de $\sigma^2 \sim GammaInversa(\alpha_1, \delta_1)$ en el modelo de regresión bayesiano seleccionado. 51

3.8 Densidades T-student posteriores de cada uno de los coeficientes del modelo de regresión. De color azul se marca el intervalo de 95 % de credibilidad. De color rojo se marca la media y moda (las cuales coinciden por simetría de la distribución). 53

3.9 Valores ajustados, $PIBT = \hat{Y} = \mathbf{X}\hat{\underline{\beta}}$, y sus intervalos de credibilidad al 95 % para el modelo de regresión desde el enfoque bayesiano. 56

3.10 Predicción del PIBT (línea azul punteada), primer trimestre del 2020 junto con su intervalo de credibilidad al 95 % (línea punteada gris). De color rojo se muestra toda la serie de valores ajustados y de color negro la serie observada de la variable de interés. 57

3.11 Crecimientos trimestrales de las variables del modelo de regresión $PIBT = \mathbf{X}\underline{\beta}$ en el periodo del primer trimestre del 2017 al cuarto trimestre del 2021. 61

4.1 Fuente: Periódico El Economista.[7] 66

4.2 Modelo 1 de 3 el cual se acepta la prueba de Johansen de cointegración de las series. 72

4.3 Modelo 3 de 3 el cual se acepta la prueba de Johansen de cointegración de las series. 72

4.4 Predicciones 2017. 73

4.5 Predicciones 2018. 73

4.6 Predicciones 2019. 74

4.7 Predicciones 2020. 74

4.8 Predicciones 2021. 75

Capítulo 1

Marco Teórico

1.1. Introducción a la Estadística Bayesiana

“I see great value in *both* Frequentist and Bayesian approaches to Statistics; and the last thing I want to do is to dwell to too great an extent on old controversies.” David Williams [21].

1.1.1. Teorema de Bayes

Consideren eventos A y B , subconjuntos de Ω , tales que $\mathbb{P}(A) \neq 0$,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Suponga H_1, H_2, \dots, H_n una partición **exhaustiva y exclusiva** del espacio muestral, Ω , con

$$\mathbb{P}(H_k) > 0, \quad \forall k \in \{1, \dots, n\}$$

Sea Z un evento con $\mathbb{P}(Z) > 0$; se tiene que:

$$\mathbb{P}(Z) = \sum_{j=1}^n \mathbb{P}(H_j \cap Z) = \sum_{j=1}^n \mathbb{P}(H_j) \mathbb{P}(Z|H_j)$$

Otro resultado útil que se desprende del teorema de Bayes, dice que:

$$\mathbb{P}(H_i|Z) = \frac{\mathbb{P}(H_i) \mathbb{P}(Z|H_i)}{\mathbb{P}(Z)}$$

O más convenientemente en el contexto Bayesiano:

$$\mathbb{P}(H_i|Z) \propto \mathbb{P}(H_i) \mathbb{P}(Z|H_i)$$

Podemos pensar en H_1, \dots, H_n como “hipótesis” entre las cuales, una es verdadera; entonces Z es la información que sabemos que es cierta por medio del resultado de un experimento. Cuando nos dicen que Z ocurrió:

1. $\mathbb{P}(H_i)$ Será el nivel de “creencia” **inicial** de la i -ésima hipótesis.
2. $\mathbb{P}(Z|H_i)$ es la probabilidad condicional de observar el resultado Z cuando H_i sucede.
3. $\mathbb{P}(H_i|Z)$ es el nivel **posterior**, es decir, una vez observado el resultado, de “creencia” de la i -ésima hipótesis.

Uno de los principales factores que obtenemos al utilizar la perspectiva bayesiana es que el nivel de creencia inicial está dado por el experto, es decir, la distribución $p(H_i)$ se asume como verdadera antes de observar el experimento. Lo que se pretende, utilizando la información obtenida de observar el experimento, es actualizar el nivel de creencia, de tal forma que recoja ambas partes: el nivel inicial de creencia que algún experto brinda y la información que el experimento nos proporciona. Es importante notar que entre más grande sea la muestra observada, se debe de dar más importancia al experimento que a la información inicial, luego entonces, la estadística bayesiana tiene relevancia en muestras pequeñas o en algoritmos que aprenden con el tiempo, como son las redes neuronales o los modelos de aprendizaje supervisado: la información al tiempo t puede servir como información inicial para el modelo en el futuro y así sucesivamente mejorando cada vez más el mismo.

1.1.2. Ejemplo

Se tiene una muestra aleatoria y_1, y_2, \dots, y_n de una distribución *Bernoulli*(θ). Se quiere dar una estructura bayesiana de la forma:

$$\pi(\theta) \sim \text{Beta}(\theta|\alpha, \beta)$$

con α y β conocidos. A esta distribución del parámetro, se le conoce como **distribución apriori o inicial** del parámetro. Lo que nosotros buscamos, para poder hacer inferencia a cerca del

parámetro θ , es encontrar la distribución **final o posterior**, $p(\theta|y)$, del parámetro θ dados los datos.

Por lo argumentado anteriormente se sabe que

$$\pi(\theta|y) \propto L(y|\theta) \pi(\theta)$$

donde:

1. $\pi(\theta) \sim \text{Beta}(\alpha, \beta)$. La distribución inicial del parámetro.
2. $L(y|\theta) \sim \text{Binomial}(n, \theta)$. La verosimilitud del modelo o experimento.
3. $\pi(\theta|y) \sim ?$. La distribución posterior o final del parámetro.

Por el momento utilizaremos la notación usual de verosimilitud ($L(y|\theta)$). Ya después utilizaremos las notaciones comunes en Bayesiana, $p(\cdot|\cdot)$, de distribuciones condicionales.

Calculemos entonces $\pi(\theta|y)$:

$$\begin{aligned} \pi(\theta|y) &\propto \binom{n}{k} \theta^y (1-\theta)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &\propto \theta^{(y+\alpha-1)} (1-\theta)^{(n-y+\beta-1)} \end{aligned}$$

Lo cual, nótese que es el Kernel de una distribución $\text{Beta}(y+\alpha, n-y+\beta)$. Por lo tanto:

$$\pi(\theta|y) \sim \text{Beta}(y+\alpha, n-y+\beta)$$

Ya que se preservó la familia de la distribución inicial en la distribución final, se dice que es una estructura conjugada.

1.1.3. Familias conjugadas

Una estructura bayesiana se dice que es **conjugada** si para $p(\theta) \in P$ y $p(y|\theta) \in F$ entonces se tiene que $p(\theta|y) \in P$. Donde P y F son familias de distribuciones.

Familia Poisson-Gamma

Sea x_1, x_2, \dots, x_n una muestra aleatoria con distribución $Poisson(\theta)$. Y suponemos como distribución inicial de θ , una $Gamma(\alpha, \beta)$, con α y β conocidos.

Sabemos que

$$p(x|\theta) = \frac{e^{-n\theta} \theta^{n\bar{X}}}{\prod_{i=1}^n x_i!}$$

Luego,

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\theta} \theta^{\alpha-1} 1_{(0,\infty)}(\theta)$$

Por lo tanto nuevamente por el teorema de Bayes, se procede a calcular la distribución final de θ :

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

$$\propto e^{-\theta(\beta+n)} \theta^{\alpha+n\bar{X}-1}$$

Lo cual, notamos que es el kernel de una distribución Gamma

$$\theta|x \sim Gamma(\alpha + n\bar{X}, \beta + n)$$

Por lo que podemos concluir que la distribución Poisson-Gamma es conjugada. Análogamente y como último ejemplo inicial, si

$$p(\theta) \sim Gamma(\alpha, \beta); \quad p(x|\theta) \sim exp(\theta)$$

Se logra fácilmente concluir que es también una estructura conjugada de la forma:

$$p(\theta|x) \sim Gamma(\alpha + n, \beta + n\bar{X})$$

Modelo Normal $N(y|\theta, \sigma^2)$, con σ^2 conocida

Desde el enfoque bayesiano, es común parametrizar a la distribución normal por medio del vector (μ, τ) , con $1/\sigma^2 = \tau$. A τ se le conoce como la “precisión” de la variable aleatoria. Sea y_1, y_2, \dots, y_n una muestra aleatoria de la distribución $N(\theta, \sigma^2)$, con:

$$p(\theta) \sim N(\theta_0, \tau_0),$$

con θ_0 y τ_0 conocidos.

Lo que nos interesa es encontrar la distribución posterior o final del parámetro θ , es decir, $p(\theta|y)$. Y se calcula de la siguiente forma:

$$\begin{aligned} p(\theta|y) &\propto p(\theta|y)p(\theta) \\ &= \frac{\tau^{n/2}}{(2\pi)^{n/2}} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^n (y_i - \theta)^2 \right\} \frac{\tau_0^{1/2}}{\sqrt{2\pi}} \exp \left\{ -\frac{\tau_0}{2} (\theta - \theta_0)^2 \right\} \end{aligned}$$

Los cuales, bajo un conjunto de simplificaciones aritméticas, encontramos nuevamente el kernel de la distribución Normal. logrando así que:

$$p(\theta|y) \sim N(\mu^*, \tau^*),$$

con:

$$\mu^* = \frac{\frac{\theta_0 \tau_0}{2} + \frac{\tau}{2} \sum_{i=1}^n y_i}{\frac{n\tau}{2} + \frac{\tau_0}{2}}, \quad \tau^{2*} = \frac{n\tau}{2} + \frac{\tau_0}{2}$$

Nótese entonces que el modelo Normal, con media normal, es conjugado. Se muestra a continuación en la Figura 1.1 un ejemplo en el cual se supone x_1, x_2, \dots, x_{55} , una muestra aleatoria tal que proviene de una distribución $N(\mu = 9.5, \tau = 1)$ (Semilla = 55). El experto en estadística, el cual quiere estudiar la media de la población, propone una distribución inicial para $\mu \sim N(10, 1)$. Al observar la muestra, obtiene que $\bar{x} = 9.4961$, por lo que la distribución posterior de μ , va a tener los parámetros $\mu^* = 9.5051$ y $\tau^* = 1/28$.

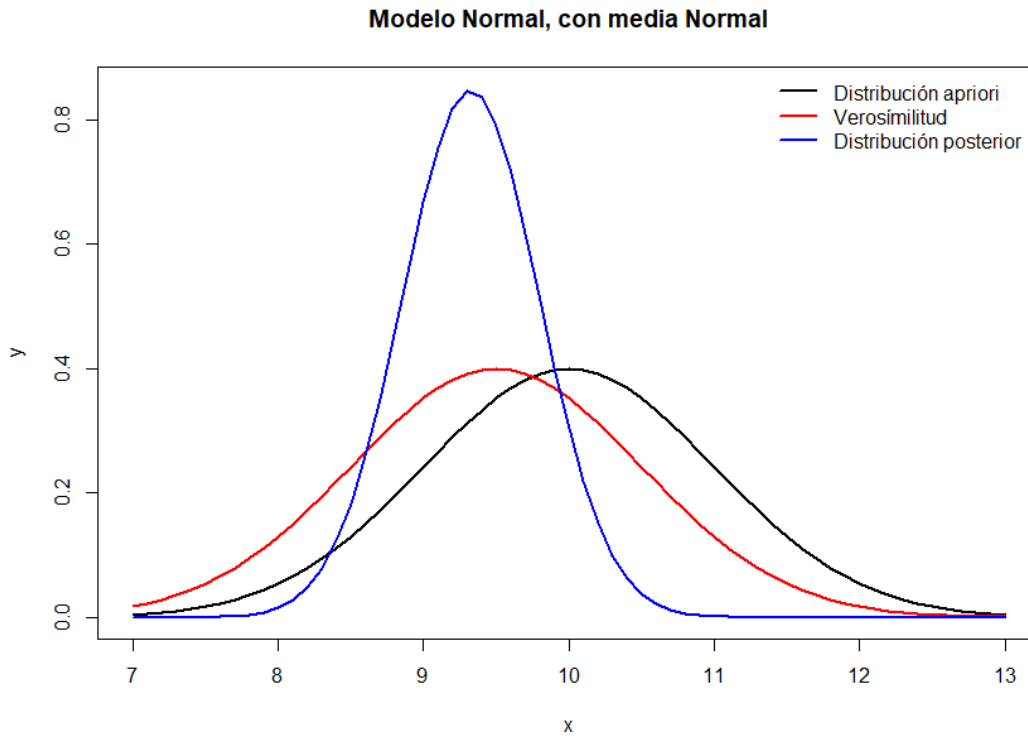


Figura 1.1: Ejemplo de estructura bayesiana de un modelo de una variable aleatoria normal con parámetro μ que proviene también de una variable aleatoria normal.

1.2. Regresión lineal

El término regresión fue acuñado por Francis Galton [11] en el siglo XIX donde observó que las alturas de los descendientes de ancestros altos tienden a regresar hacia abajo, hacia un promedio normal (un fenómeno conocido como regresión a la media). Dada una variable Y de respuesta (también conocida como target), y un conjunto de variables (también conocidas como covariables, variables exógenas, variables independientes, etcétera) \underline{z} , surge de manera natural preguntarnos cuál deberá ser la relación funcional para modelar dicha relación. La más común que se utiliza es:

$$\mathbb{E}(Y|z) = \mu(\underline{z}),$$

donde, en general, μ es una función determinista desconocida. Es decir, modelar la esperanza de Y , como una función determinista de \underline{z} . La forma más simple de aproximar a μ , es

$$\mu(\underline{z}) = \psi(z; \underline{\beta}),$$

donde $\underline{\beta} \in \mathbb{R}^k$, denota un vector de parámetros desconocidos a estimar (o actualmente se dice “calibrar”) de la siguiente forma:

$$\psi(\underline{z}; \underline{\beta}) = \beta_0 + \beta_1 s_1(\underline{z}) + \dots + \beta_k s_k(\underline{z}),$$

donde s_i son funciones conocidas. A la función ψ , se le trata como si fuese la verdadera función de regresión μ ; por lo que todo se reduce a estimar al vector de parámetros $\underline{\beta}$. Agregado a lo anterior, usualmente se acostumbra un supuesto distribucional de la forma:

$$Y|\underline{z} \sim N(\mu(\underline{z}), \sigma^2),$$

con $\sigma^2 > 0$ otro parámetro a estimar. También se tiene un supuesto estructural en las observaciones de la forma, $Cov(Y_i, Y_j) = 0, \forall i \neq j$. Otra forma usual de escribir el modelo es:

$$Y_i = \beta_0 + \beta_1 s_1(\underline{z}_i) + \dots + \beta_k s_k(\underline{z}_i) + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

Se define $x_{ij} := s_j(\underline{z}_i)$ $i \in \{1, \dots, n\}$; $j \in \{1, \dots, k\}$, logrando reescribir el modelo como:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

Dada la relación lineal que hemos impuesto resulta conveniente utilizar la siguiente notación matricial y escribir el modelo lineal como (definimos $p = k + 1$):

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}; \quad \underline{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

donde:

$$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ 1 & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}_{n \times p} \quad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1}$$

El vector de parámetros:

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{p \times 1}$$

1.2.1. Estimación de parámetros

Dado que ahora conocemos la forma de \underline{Y} , podemos encontrar la función de verosimilitud

$$f(\underline{Y}; \underline{\beta}; \sigma^2) = \frac{1}{(2\pi)^{n/2} |\sigma^2 \mathbf{I}_n|^{1/2}} \exp \left\{ -\frac{1}{2} (\underline{Y} - \mathbf{X}\underline{\beta})^T (\sigma^2 \mathbf{I}_n)^{-1} (\underline{Y} - \mathbf{X}\underline{\beta}) \right\}$$

Como $|\sigma^2 \mathbf{I}_n| = \sigma^{2n}$ y $(\sigma^2 \mathbf{I}_n)^{-1} = \frac{1}{\sigma^2} \mathbf{I}_n$ entonces la verosimilitud esta dada por:

$$L(\underline{\beta}; \sigma^2; \underline{Y}) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\underline{Y} - \mathbf{X}\underline{\beta})^T (\underline{Y} - \mathbf{X}\underline{\beta}) \right\}$$

como también su logverosimilitud está dada por:

$$\log L(\underline{\beta}; \sigma^2; \underline{Y}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\underline{Y} - \mathbf{X}\underline{\beta})^T (\underline{Y} - \mathbf{X}\underline{\beta})$$

Maximizamos con respecto a $(\underline{\beta}, \sigma^2)$ para ello derivamos e igualamos a cero:

$$\frac{\partial}{\partial \underline{\beta}} \log L(\underline{\beta}; \sigma^2; \underline{Y}) = -\frac{1}{2\sigma^2} (-2\mathbf{X}^T \underline{Y} + 2\mathbf{X}^T \mathbf{X}\underline{\beta})$$

$$\frac{\partial}{\partial \sigma^2} \log L(\underline{\beta}; \sigma^2; \underline{Y}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\underline{Y} - \mathbf{X}\underline{\beta})^T (\underline{Y} - \mathbf{X}\underline{\beta})$$

De la primera derivada parcial obtenemos las ecuaciones normales, es decir:

$$\mathbf{X}^T \mathbf{X}\underline{\beta} = \mathbf{X}^T \underline{Y}$$

Esta ecuación nos muestra la primera restricción a nuestro modelo, y es que $\mathbf{X}^T \mathbf{X}$ debe ser invertible (\mathbf{X} debe tener rango completo) en cuyo caso el estimador máximo verosímil de $\underline{\beta}$ está dado por:

$$\hat{\underline{\beta}}_{MV} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underline{Y}$$

Mientras que para σ^2 , obtenemos:

$$-n + \frac{1}{\sigma^2} (\underline{\mathbf{Y}} - \mathbf{X}\underline{\beta})^T (\underline{\mathbf{Y}} - \mathbf{X}\underline{\beta}) = 0$$

$$\hat{\sigma}^2 = \frac{(\underline{\mathbf{Y}} - \mathbf{X}\underline{\beta})^T (\underline{\mathbf{Y}} - \mathbf{X}\underline{\beta})}{n}$$

por lo tanto, al sustituir a $\hat{\underline{\beta}}_{MV}$ y definiendo $\hat{\mathbf{Y}} = \mathbf{X}\hat{\underline{\beta}}_{MV}$ se obtiene

$$\hat{\sigma}_{MV}^2 = \frac{(\underline{\mathbf{Y}} - \hat{\mathbf{Y}})^T (\underline{\mathbf{Y}} - \hat{\mathbf{Y}})}{n} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Algunas propiedades que tienen nuestros estimadores máximo verosímiles son:

1. $\mathbb{E}(\hat{\underline{\beta}}_{MV}) = \underline{\beta}$ (es insesgado)
2. $Var(\hat{\underline{\beta}}_{MV}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$
3. $\hat{\underline{\beta}}_{MV} \sim N_p(\underline{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$
4. $\mathbb{E}(\hat{\sigma}_{MV}^2) = \sigma^2 \frac{n-p}{n}$ (es sesgado)
5. $Var(\hat{\sigma}_{MV}^2) = 2 \frac{n-p}{n} \sigma^4$
6. $\hat{\sigma}_{MV}^2 \sim Gamma(\frac{n-p}{2}, \frac{n}{2\sigma^2})$
7. $\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ (insesgado)
8. Ya que $\hat{\underline{\beta}}_{MV} \sim N_p(\underline{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$, si definimos $C = (\mathbf{X}^T \mathbf{X})^{-1}$ se tiene:

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 C_{(i+1)(i+1)})$$

Logrando que:

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}} \sim t_{(n-p)}$$

Inferencia sobre $\underline{\beta}$

De la expresión 8, se desprende la prueba de hipótesis e intervalos de confianza correspondientes para el parámetro β_i , $i \in \{0, \dots, k\}$ de la forma

$$H_0 : \beta_i = b_i \text{ vs } H_1 : \beta_i \neq b_i; \quad i \in \{0, \dots, k\}$$

La regla es rechazar H_0 cuando $|t| \geq \tau_{n-p}^{(1-\alpha/2)}$, donde $t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}$

Esta prueba toma relevancia cuando tomamos $b_i = 0$, ya que nos ayuda a determinar si la variable asociada a ese coeficiente es estadísticamente diferente de cero, lo que se traduce en verificar si hay un efecto de la variable X_i en Y .

La estadística $t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}$ definida previamente, es una cantidad pivotal para β_i , por lo que podemos definir un intervalo de confianza de la forma:

$$\mathbb{P} \left(-\tau_{n-p}^{(1-\alpha/2)} \leq \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}} \leq \tau_{n-p}^{(1-\alpha/2)} \right) = 1 - \alpha$$

$$\mathbb{P} \left(\hat{\beta}_i - \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)} \tau_{n-p}^{(1-\alpha/2)}} \leq \beta_i \leq \hat{\beta}_i + \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)} \tau_{n-p}^{(1-\alpha/2)}} \right) = 1 - \alpha$$

Por lo tanto un intervalo de confianza al $100 * (1 - \alpha) \%$ para β_i está dado por:

$$\left(\hat{\beta}_i - \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)} \tau_{n-p}^{(1-\alpha/2)}}, \hat{\beta}_i + \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)} \tau_{n-p}^{(1-\alpha/2)}} \right)$$

donde $\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}$ se le conoce como error estándar del estimador.

La tabla ANOVA nos ayuda a contrastar la hipótesis conjunta

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \text{ vs } H_1 : \beta_i \neq 0, \text{ p.a } i \in \{1, \dots, k\},$$

a continuación en la Tabla 1.1 se muestra la estructura general de la misma:

Var.	S.C.	G.Lib	S.C.M.	F
(SCR)	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	p-1	$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{p-1}$	$\frac{(n-p) \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{(p-1) \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
(SCE)	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	n-p	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p}$	
(SCT)	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	n-1	$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n-1}$	

Tabla 1.1: Estructura general de la tabla ANOVA para regresión lineal.

Y se rechaza H_0 al nivel de significancia α si el valor del estadístico F cuando toma valores mayores al cuantíl $F_{(p-1, n-p)}^{1-\alpha}$ de una distribución F de Fischer de parámetros $p-1$ y $n-p$.

Inferencia sobre σ^2

Por otro lado la inferencia sobre σ^2 se resume a la cantidad pivotal:

$$\frac{(\underline{Y} - \hat{\underline{Y}})^T (\underline{Y} - \hat{\underline{Y}})}{\sigma^2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi_{n-p}^2$$

por lo que análogamente, podemos dar un intervalo de confianza para σ^2 , a saber:

$$\left(\frac{1}{\chi_{n-p}^{2(1-\alpha/2)}} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \frac{1}{\chi_{n-p}^{2(\alpha/2)}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)$$

Inferencia sobre y

Es claro que una vez calibrado el modelo, nos gustaría poder utilizarlo para dada un conjunto nuevo de covariables \underline{x}^* , hacer la inferencia utilizando la relación $\mathbb{E}(Y|\underline{x}^*)$, logrando estimar a y :

$$\hat{y}^* = \underline{x}^{*T} \hat{\underline{\beta}}_{MV} = \underline{x}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underline{Y}$$

Como habíamos visto \underline{Y} es un vector normal multivariado, por lo que se sigue:

$$\hat{y}^* \sim N(\mathbb{E}(y|\underline{x}^*), \sigma^2 \underline{x}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \underline{x}^*)$$

logrando encontrar la cantidad pivotal:

$$\frac{\hat{y}^* - \mathbb{E}(y|\underline{x}^*)}{\sqrt{\sigma^2 \underline{x}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \underline{x}^*}} \sim t_{n-p}$$

De la cual se desprenden las pruebas de hipótesis e intervalos de confianza correspondientes al valor $\mathbb{E}(y|\underline{\mathbf{x}}^*)$ a saber:

$$\left(\hat{y}^* - t_{(n-p)}^{(1-\alpha/2)} \sqrt{\sigma^2 \underline{\mathbf{x}}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \underline{\mathbf{x}}^*}, \hat{y}^* + t_{(n-p)}^{(1-\alpha/2)} \sqrt{\sigma^2 \underline{\mathbf{x}}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \underline{\mathbf{x}}^*} \right)$$

Sin embargo, este intervalo solo hace inferencias sobre $\mathbb{E}(y|\underline{\mathbf{x}}^*)$, si se desea hacer inferencia sobre y^* se utiliza la definición del modelo de regresión, es decir:

$$y^* = \underline{\mathbf{x}}^{*T} \underline{\beta} + \varepsilon^*, \quad \varepsilon^* \sim N(0, \sigma^2)$$

construyendo así, una cantidad pivotal para y^* de la forma:

$$\frac{\hat{y}^* - y^*}{\sqrt{\hat{\sigma}^2 (1 + \underline{\mathbf{x}}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \underline{\mathbf{x}}^*)}} \sim t_{n-p}$$

para poder dar un intervalo que se le conoce intervalo de **predicción**:

$$\left(\hat{y}^* \pm t_{(n-p)}^{(1-\alpha/2)} \sqrt{\hat{\sigma}^2 (1 + \underline{\mathbf{x}}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \underline{\mathbf{x}}^*)} \right)$$

Debe de observarse que este intervalo incrementa su longitud respecto al intervalo anterior, esto se debe a que este reconoce dos tipos de incertidumbres, la primera referente a la estimación de los parámetros utilizada en $\hat{y}^* = \underline{\mathbf{x}}^{*T} \hat{\underline{\beta}}_{MV}$ y la segunda se refiere a la variabilidad presente en la definición del modelo al suponer $\varepsilon \sim N(0, \sigma^2)$.

1.3. Regresión Lineal, enfoque Bayesiano

Utilizaremos la notación de Soriano [9], la cual es clara para el lector y se sigue de la sección introductoria al enfoque bayesiano.

1.3.1. Distribución final de los parámetros

Sea $\underline{\beta} \in \mathbb{R}^p$ un vector de parámetros, y $\mathbf{X} \in \mathbb{R}^{n \times p}$ una matriz de diseño conocida. Definamos el modelo lineal:

$$\underline{\mathbf{y}} = \mathbf{X} \underline{\beta} + \underline{\varepsilon}$$

donde $\underline{\varepsilon} \sim N_n(\underline{\varepsilon}|0, \tau \mathbf{I}_n)$ con $\tau = \sigma^{-2} > 0$. Suponemos una matriz de precisión con correlación 0 entre las variables, lo que implica independencia entre las ε_i , logrando:

$$\underline{y} \sim N_n(\underline{y}|\mathbf{X}\underline{\beta}, \tau \mathbf{I}_n)$$

El objetivo inicial, es encontrar las distribuciones finales de los parámetros $\underline{\beta}$ y τ utilizando la propiedad de Bayes, a saber:

$$p(\underline{\beta}, \tau|\underline{y}) \propto p(\underline{y}|\underline{\beta}, \tau) p(\underline{\beta}, \tau)$$

$$p(\underline{\beta}|\underline{y}) = \int_0^\infty p(\underline{\beta}, \tau|\underline{y}) d\tau; \quad p(\tau|\underline{y}) = \int_{\mathbb{R}^p} p(\underline{\beta}, \tau|\underline{y}) d\underline{\beta}$$

Obsérvese que la verosimilitud de nuestro modelo está dado por:

$$\begin{aligned} p(\underline{y}|\underline{\beta}, \tau) &= N(\underline{y}|\mathbf{X}\underline{\beta}, \tau \mathbf{I}_n) \\ &\propto \tau^{n/2} \exp\left\{-\frac{\tau}{2}(\underline{y} - \mathbf{X}\underline{\beta})^T(\underline{y} - \mathbf{X}\underline{\beta})\right\} \end{aligned}$$

Se puede probar que:

$$(\underline{y} - \mathbf{X}\underline{\beta})^T(\underline{y} - \mathbf{X}\underline{\beta}) = (\underline{y} - \mathbf{X}\hat{\underline{\beta}})^T(\underline{y} - \mathbf{X}\hat{\underline{\beta}}) + (\underline{\beta} - \hat{\underline{\beta}})^T \mathbf{X}^T \mathbf{X} (\underline{\beta} - \hat{\underline{\beta}})$$

donde $\hat{\underline{\beta}} = \hat{\underline{\beta}}_{MV} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underline{y}$. Por lo que la verosimilitud puede escribirse como:

$$\begin{aligned} p(\underline{y}|\underline{\beta}, \tau) &\propto \tau^{n/2} \exp\left\{-\frac{\tau}{2} \left((\underline{y} - \mathbf{X}\hat{\underline{\beta}})^T(\underline{y} - \mathbf{X}\hat{\underline{\beta}}) + (\underline{\beta} - \hat{\underline{\beta}})^T \mathbf{X}^T \mathbf{X} (\underline{\beta} - \hat{\underline{\beta}}) \right)\right\} \\ &\propto \tau^{n/2} \exp\left\{-\frac{\tau}{2} \left(\tilde{\beta}_0 + (\underline{\beta} - \hat{\underline{\beta}})^T \mathbf{X}^T \mathbf{X} (\underline{\beta} - \hat{\underline{\beta}}) \right)\right\} \end{aligned}$$

Lo cual es claro observar que es el kernel de una distribución Normal-multivariada-Gamma. Lo que nos hace proponer a dicha distribución como la apriori, es decir:

$$p(\underline{\beta}, \tau|\mathbf{P}_0, \alpha_0, \delta_0) \propto N_p(\underline{\beta}|\mu_0, \tau \mathbf{P}_0) \text{Gamma}(\tau|\alpha_0, \delta_0)$$

donde $\mathbf{P}_0 \in \mathbb{R}^{p \times p}$, $\mu_0 \in \mathbb{R}^p$, $\alpha_0, \delta_0 \in \mathbb{R}$ son hiperparámetros que se deben de ajustar para modelar nuestro conocimiento inicial de los parámetros $\underline{\beta}$ y τ . Algunas propiedades que ya sabemos de las marginales iniciales son

1. $\underline{\beta} \sim T_p \left(\underline{\beta} | 2\alpha_0, \mu_0, \frac{\alpha_0}{\delta_0} \mathbf{P}_0 \right)$
2. $\mathbb{E}(\underline{\beta}) = \mu_0$
3. $Var(\underline{\beta}) = \frac{\delta_0}{\alpha_0 - 1} \mathbf{P}_0^{-1}$
4. $\tau \sim Gamma(\tau | \alpha_0, \delta_0)$
5. $\mathbb{E}(\tau) = \frac{\alpha_0}{\delta_0}$
6. $Var(\tau) = \frac{\alpha_0}{\delta_0^2}$

Retomando el cálculo de la distribución final de los parámetros $\underline{\beta}$ y τ

$$\begin{aligned}
 p(\underline{\beta}, \tau | \mathbf{y}) &\propto p(\mathbf{y} | \underline{\beta}, \tau) p(\underline{\beta}, \tau) \\
 &\propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} \left((\underline{\beta} - \hat{\underline{\beta}})^T \mathbf{X}^T \mathbf{X} (\underline{\beta} - \hat{\underline{\beta}}) + (\mathbf{y} - \mathbf{X} \hat{\underline{\beta}})^T (\mathbf{y} - \mathbf{X} \hat{\underline{\beta}}) \right) \right\} \\
 &\quad \tau^{p/2} \exp \left\{ -\frac{\tau}{2} (\underline{\beta} - \mu_0)^T \mathbf{P}_0 (\underline{\beta} - \mu_0) \right\} \tau^{\alpha_0 - 1} e^{-\delta_0 \tau} \\
 &\propto \tau^{n/2 + \alpha_0 - 1} \exp \left\{ -\tau \left(\delta_0 + \frac{1}{2} (\mathbf{y} - \mathbf{X} \hat{\underline{\beta}})^T (\mathbf{y} - \mathbf{X} \hat{\underline{\beta}}) \right) \right\} \\
 &\quad \tau^{p/2} \exp \left\{ -\frac{\tau}{2} \left((\underline{\beta} - \hat{\underline{\beta}})^T \mathbf{X}^T \mathbf{X} (\underline{\beta} - \hat{\underline{\beta}}) + (\underline{\beta} - \mu_0)^T \mathbf{P}_0 (\underline{\beta} - \mu_0) \right) \right\}
 \end{aligned}$$

Se demuestra que:

$$\begin{aligned}
 (\underline{\beta} - \hat{\underline{\beta}})^T \mathbf{X}^T \mathbf{X} (\underline{\beta} - \hat{\underline{\beta}}) + (\underline{\beta} - \mu_0)^T \mathbf{P}_0 (\underline{\beta} - \mu_0) = \\
 (\underline{\beta} - \mu_1)^T \mathbf{P}_1 (\underline{\beta} - \mu_1) + \hat{\underline{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\underline{\beta}} + \mu_0^T \mathbf{P}_0 \mu_0 - \mu_1^T \mathbf{P}_1 \mu_1
 \end{aligned}$$

donde:

$$\mathbf{P}_1 = \mathbf{X}^T \mathbf{X} + \mathbf{P}_0; \quad \mu_1 = (\mathbf{X}^T \mathbf{X} + \mathbf{P}_0)^{-1} (\mathbf{X}^T \mathbf{y} + \mathbf{P}_0 \mu_0)$$

Sustituyendo y reorganizando $p(\underline{\beta}, \tau | \mathbf{y})$ es proporcional a:

$$\tau^{n/2 + \alpha_0 - 1} \exp \left\{ -\tau \left(\delta_0 + \frac{1}{2} \left((\mathbf{y} - \mathbf{X} \hat{\underline{\beta}})^T (\mathbf{y} - \mathbf{X} \hat{\underline{\beta}}) + \hat{\underline{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\underline{\beta}} + \mu_0^T \mathbf{P}_0 \mu_0 - \mu_1^T \mathbf{P}_1 \mu_1 \right) \right) \right\}$$

$$\tau^{p/2} \exp \left\{ -\frac{\tau}{2} (\underline{\beta} - \mu_1)^T \mathbf{P}_1 (\underline{\beta} - \mu_1) \right\}$$

Se puede probar que:

$$(\mathbf{y} - \mathbf{X} \hat{\underline{\beta}})^T (\mathbf{y} - \mathbf{X} \hat{\underline{\beta}}) + \hat{\underline{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\underline{\beta}} + \mu_0^T \mathbf{P}_0 \mu_0 - \mu_1^T \mathbf{P}_1 \mu_1 =$$

$$(\mathbf{y} - \mathbf{X} \mu_1)^T (\mathbf{y} - \mathbf{X} \mu_1) + (\mu_1 - \mu_0)^T \mathbf{P}_0 (\mu_1 - \mu_0)$$

Por lo tanto $p(\underline{\beta}, \tau | \mathbf{y})$ es proporcional:

$$\propto \tau^{n/2 + \alpha_0 - 1} \exp \left\{ -\tau \left(\delta_0 + \frac{1}{2} \left((\mathbf{y} - \mathbf{X} \mu_1)^T (\mathbf{y} - \mathbf{X} \mu_1) + (\mu_1 - \mu_0)^T \mathbf{P}_0 (\mu_1 - \mu_0) \right) \right) \right\}$$

$$\tau^{p/2} \exp \left\{ -\frac{\tau}{2} (\underline{\beta} - \mu_1)^T \mathbf{P}_1 (\underline{\beta} - \mu_1) \right\}$$

con:

1. $\mathbf{P}_1 = \mathbf{X}^T \mathbf{X} + \mathbf{P}_0$
2. $\mu_1 = (\mathbf{X}^T \mathbf{X} + \mathbf{P}_0)^{-1} (\mathbf{X}^T \mathbf{y} + \mathbf{P}_0 \mu_0)$
3. $\alpha_1 = \frac{n}{2} + \alpha_0$
4. $\delta_1 = \delta_0 + \frac{1}{2} \left((\mathbf{y} - \mathbf{X} \mu_1)^T (\mathbf{y} - \mathbf{X} \mu_1) + (\mu_1 - \mu_0)^T \mathbf{P}_0 (\mu_1 - \mu_0) \right)$

se obtiene que:

$$p(\underline{\beta}, \tau | \underline{y}) = N_p(\underline{\beta} | \mu_1, \tau \mathbf{P}_1) \text{Gamma}(\tau | \alpha_1, \delta_1)$$

lo cual, por las propiedades de la distribución se tiene que:

$$p(\beta | \underline{y}) = T_p \left(\underline{\beta} | 2\alpha_1, \mu_1, \frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)$$

$$p(\tau | \underline{y}) = \text{Gamma}(\alpha_1, \delta_1) \implies p(\sigma^2 | \underline{y}) = \text{GammaInversa}(\alpha_1, \delta_1)$$

Inferencia sobre τ

1. $\mathbb{E}(\hat{\tau}) = \frac{\alpha_1}{\delta_1}$
2. $\text{Moda}(\hat{\tau}) = \frac{\alpha_1 - 1}{\delta_1}$
3. $\mathbb{E}(\hat{\sigma}^2) = \frac{\delta_1}{\alpha_1 - 1}$
4. $\text{Moda}(\hat{\sigma}^2) = \frac{\delta_1}{\alpha_1 + 1}$

Inferencia sobre $\underline{\beta}$

Haciendo inferencia sobre $\underline{\beta}$, sabemos que utilizando la función de utilidad de pérdida cuadrática, es decir, utilizando la media de la distribución posterior, podemos dar un estimador puntual de β , es decir:

$$\hat{\underline{\beta}} = \mu_1 = (\mathbf{X}^T \mathbf{X} + \mathbf{P}_0)^{-1} (\mathbf{X}^T \underline{y} + \mathbf{P}_0 \mu_0)$$

En cuanto a dar un intervalo de credibilidad a β y sabiendo que la densidad T-multivariada pertenece a las distribuciones elípticas y por tanto sabemos que:

$$R^2 = (\underline{\beta} - \mu_1)^T \frac{\alpha_1}{\delta_1} \mathbf{P}_1 (\underline{\beta} - \mu_1) \sim pF_{(p, 2\alpha_1)}$$

Por lo tanto:

$$\mathbb{P} \left(\frac{1}{p} (\underline{\beta} - \mu_1)^T \frac{\alpha_1}{\delta_1} \mathbf{P}_1 (\underline{\beta} - \mu_1) \leq \mathbf{F}_{(p, 2\alpha_1)}^{(1-\alpha)} \right) = 1 - \alpha$$

Suponiendo que solo estamos interesados en β_i , sabemos que

$$p(\beta_i|\underline{Y}) = T_1 \left(\beta_i | 2\alpha_1, \mu_{1(i+1)}, \frac{1}{c_{(i+1)(i+1)}} \right), \quad i \in \{0, \dots, k\}$$

donde $c_{jj} = \left(\left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \right)_{jj}$

De la ecuación anterior, se tiene que el estimador puntual para β_i está dado por:

$$\hat{\beta}_i = \mu_{1(i+1)} = ((\mathbf{X}^T \mathbf{X} + \mathbf{P}_0)^{-1} (\mathbf{X}^T \underline{Y} + \mathbf{P}_0 \mu_0))_{(i+1)}$$

Hablando de un intervalo de credibilidad para β_i , se sabe:

$$R^2 = (\beta_i - \mu_{1(i+1)}) \left(\frac{1}{c_{(i+1)(i+1)}} \right) (\beta_i - \mu_{1(i+1)}) \sim F_{(1, 2\alpha_1)}$$

Por lo tanto la región de tamaño $1 - \alpha$ está dada por:

$$(\beta_i - \mu_{1(i+1)}) \left(\frac{1}{c_{(i+1)(i+1)}} \right) (\beta_i - \mu_{1(i+1)}) \leq F_{(1, 2\alpha_1)}^{(1-\alpha)}$$

$$|\beta_i - \mu_{1(i+1)}| \left(\frac{1}{\sqrt{c_{(i+1)(i+1)}}} \right) \leq t_{2\alpha_1}^{1-\alpha/2}$$

Prueba Lindley

Supóngase que se está interesado en probar:

$$H_0 : \mathbf{A}\underline{\beta} = \mathbf{b} \quad \text{vs} \quad H_1 : \mathbf{A}\underline{\beta} \neq \mathbf{b}$$

con $\mathbf{A} \in \mathbb{R}^{m \times p}$ de rango completo y $\mathbf{b} \in \mathbb{R}^m$.

Una forma de probar esta hipótesis (Lindley) es construir la región de credibilidad para $\mathbf{A}\underline{\beta}$, es decir encontrar la densidad final del vector aleatorio $\mathbf{A}\underline{\beta}$, y luego construir D^* la región de credibilidad de tamaño $1 - \alpha$, para en ella verificar si \mathbf{b} se encuentra en dicha región y decidir si rechazar H_0 o no. Utilizando el hecho de que:

$$\beta|\underline{Y} \sim T_p \left(\underline{\beta} \mid 2\alpha_1, \mu_1, \frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)$$

se sigue con las propiedades de las distribuciones elípticas:

$$\mathbf{A}\underline{\beta}|\underline{Y} \sim T_p \left(\underline{\beta} \mid 2\alpha_1, \mathbf{A}\mu_1, \left(\mathbf{A} \left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \mathbf{A}^T \right)^{-1} \right)$$

Le sigue entonces:

$$R^2 = (\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1)^T \left(\mathbf{A} \left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \mathbf{A}^T \right)^{-1} (\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1) \sim mF_{(m, 2\alpha_1)}$$

La región es:

$$\frac{1}{m} (\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1)^T \left(\mathbf{A} \left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \mathbf{A}^T \right)^{-1} (\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1) \leq F_{(m, 2\alpha_1)}^{1-\alpha}$$

por lo que rechazo H_0 si:

$$\frac{1}{m} (\mathbf{b} - \mathbf{A}\mu_1)^T \left(\mathbf{A} \left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \mathbf{A}^T \right)^{-1} (\mathbf{b} - \mathbf{A}\mu_1) > F_{(m, 2\alpha_1)}^{1-\alpha}$$

Con esta prueba construida, podemos constatar la hipótesis análoga a la forma frecuentista

$$H_0 : \beta_1 = \dots = \beta_k = 0 \quad vs \quad H_1 : \beta_i \neq 0 \quad p.a. \quad i \in \{1, \dots, k\}$$

Definimos a nuestra matriz \mathbf{A} de la siguiente forma:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{(p-1) \times p}$$

a lo que es equivalente a:

$$H_0 : \mathbf{A}\underline{\beta} = \underline{0} \in \mathbb{R}^{p-1} \quad vs \quad H_1 : \mathbf{A}\underline{\beta} \neq \underline{0} \in \mathbb{R}^{p-1}$$

por lo que como definimos anteriormente nuestra región de rechazo, la hipótesis se rechaza si:

$$\frac{1}{p-1} (\mathbf{A}\mu_1)^T \left(\mathbf{A} \left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \mathbf{A}^T \right)^{-1} (\mathbf{A}\mu_1) > F_{(p-1, 2\alpha_1)}^{1-\alpha}$$

Inferencia sobre nuevas y

Imaginemos ahora que deseamos hacer inferencia sobre nuevas observaciones del modelo:

$$\mathbf{w} = \mathbf{Z}\underline{\beta} + \mathbf{e}$$

donde:

1. $\mathbf{Z} \in \mathbb{R}^{k \times p}$ es una nueva matriz de k observaciones.
2. $\mathbf{e} \sim N_k(\mathbf{e}|\underline{0}, \tau\mathbf{I})$
3. Por lo que: $\mathbf{w} \sim N_k(\mathbf{w}|\mathbf{Z}\underline{\beta}, \tau\mathbf{I})$

Por el enfoque bayesiano que se está utilizando, lo que queremos encontrar es $p(\mathbf{w}|y)$:

$$\begin{aligned} p(\mathbf{w}|y) &= \int_0^\infty \int_{\mathbb{R}^p} p(\mathbf{w}, \underline{\beta}, \tau|\underline{y}) d\underline{\beta}d\tau = \int_0^\infty \int_{\mathbb{R}^p} \frac{p(\mathbf{w}, \underline{\beta}, \tau, \underline{y})}{p(\underline{y})} d\underline{\beta}d\tau \\ &= \int_0^\infty \int_{\mathbb{R}^p} \frac{p(\mathbf{w}|\underline{\beta}, \tau, \underline{y})p(\underline{\beta}, \tau, \underline{y})}{p(\underline{y})} d\underline{\beta}d\tau \\ &= \int_0^\infty \int_{\mathbb{R}^p} p(\mathbf{w}|\underline{\beta}, \tau, \underline{y})p(\underline{\beta}, \tau|\underline{y}) d\underline{\beta}d\tau \end{aligned}$$

Sin embargo, por independencia condicional:

$$\begin{aligned} p(\mathbf{w}|\underline{\beta}, \tau, \underline{y}) &= \frac{p(\mathbf{w}, \underline{\beta}, \tau, \underline{y})}{p(\underline{\beta}, \tau, \underline{y})} = \frac{p(\mathbf{w}, \underline{y}|\underline{\beta}, \tau)p(\underline{\beta}, \tau)}{p(\underline{\beta}, \tau, \underline{y})} \\ &= \frac{p(\mathbf{w}|\underline{\beta}, \tau)p(\underline{y}|\underline{\beta}, \tau)p(\underline{\beta}, \tau)}{p(\underline{\beta}, \tau, \underline{y})} = \frac{p(\mathbf{w}|\underline{\beta}, \tau)p(\underline{y}|\underline{\beta}, \tau)p(\underline{\beta}, \tau)}{p(\underline{y}|\underline{\beta}, \tau)p(\underline{\beta}, \tau)} \\ &= p(\mathbf{w}|\underline{\beta}, \tau) \end{aligned}$$

Por lo tanto:

$$p(\mathbf{w}|y) = \int_0^\infty \int_{\mathbb{R}^p} p(\mathbf{w}|\underline{\beta}, \tau, \underline{y})p(\underline{\beta}, \tau|\underline{y}) d\underline{\beta}d\tau$$

$$= \int_0^\infty \int_{\mathbb{R}^p} p(\mathbf{w}|\underline{\beta}, \tau) p(\underline{\beta}, \tau|\mathbf{y}) d\underline{\beta} d\tau$$

donde, como ya sabemos:

$$p(\mathbf{w}|\underline{\beta}, \tau) = N_k(\mathbf{w}|\mathbf{Z}\underline{\beta}, \tau\mathbf{I}); \quad p(\underline{\beta}, \tau|\mathbf{y}) = \text{NormalGamma}(\underline{\beta}, \tau|\mu_1, \mathbf{P}_1, \alpha_1, \delta_1)$$

es decir:

$$\begin{aligned} p(\mathbf{w}|\mathbf{y}) &\propto \int_0^\infty \int_{\mathbb{R}^p} \tau^{k/2} \exp\left\{-\frac{\tau}{2}(\mathbf{w} - \mathbf{Z}\underline{\beta})^T(\mathbf{w} - \mathbf{Z}\underline{\beta})\right\} \cdots \\ &\cdots \tau^{p/2} \exp\left\{-\frac{\tau}{2}(\underline{\beta} - \mu_1)^T \mathbf{P}_1(\underline{\beta} - \mu_1)\right\} \tau^{\alpha_1-1} e^{-\tau\delta_1} d\underline{\beta} d\tau \end{aligned}$$

En el integrando se puede completar un kernel de una Normal-Multivariada en $\underline{\beta}$, por lo que nos quedamos con los términos restantes:

$$\begin{aligned} p(\mathbf{w}|\mathbf{y}) &\propto \int_0^\infty \underbrace{\tau^{k/2} \exp\left\{-\frac{\tau}{2}(\mathbf{w} - \mathbf{Z}\mu_1)^T (\mathbf{I} - \mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z})^{-1}(\mathbf{w} - \mathbf{Z}\mu_1)\right\}}_{N_k\left(\mathbf{w} \middle| \mathbf{Z}\mu_1, (\mathbf{I} + \mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z}^T)^{-1}\right)} \underbrace{\tau^{\alpha_1-1} e^{-\tau\delta_1}}_{\text{Gamma}(\tau|\alpha_1, \delta_1)} d\tau \\ &\propto \int_0^\infty \text{NormalGamma}\left(\mathbf{w}, \tau \middle| \mathbf{Z}\mu_1, (\mathbf{I} + \mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z}^T)^{-1}, \alpha_1, \delta_1\right) d\tau \\ &\propto T_K\left(\mathbf{w} \middle| 2\alpha_1, \mathbf{Z}\mu_1, \frac{\alpha_1}{\delta_1} (\mathbf{I} + \mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z}^T)^{-1}\right) \end{aligned}$$

1.3.2. Distribución no informativa de Jeffreys

Si queremos utilizar éste modelo y no contamos con información inicial para los hiperparámetros, podemos calcular la distribución inicial no informativa de Jeffreys:

Obtenemos la matriz de información de Fisher $I(\underline{\beta}, \tau)$

$$p(\mathbf{y}|\underline{\beta}, \tau) \propto \tau^{n/2} \exp\left\{-\frac{\tau}{2}(\mathbf{y} - \mathbf{X}\underline{\beta})^T(\mathbf{y} - \mathbf{X}\underline{\beta})\right\}$$

por lo que:

$$\log(p(\underline{y}|\underline{\beta}, \tau)) \propto \frac{n}{2} \log(\tau) - \frac{\tau}{2} (\underline{y} - \mathbf{X}\underline{\beta})^T (\underline{y} - \mathbf{X}\underline{\beta})$$

Obtenemos la primera derivada:

$$\begin{pmatrix} \frac{\partial \log(p(\underline{y}|\underline{\beta}, \tau))}{\partial \underline{\beta}} \\ \frac{\partial \log(p(\underline{y}|\underline{\beta}, \tau))}{\partial \tau} \end{pmatrix}_{(p+1) \times 1} = \begin{pmatrix} \tau \mathbf{X}^T (\underline{y} - \mathbf{X}\underline{\beta}) \\ \frac{n}{2\tau} - \frac{1}{2} (\underline{y} - \mathbf{X}\underline{\beta})^T (\underline{y} - \mathbf{X}\underline{\beta}) \end{pmatrix}_{(p+1) \times 1}$$

Después obtenemos la segunda derivada:

$$\begin{pmatrix} \frac{\partial^2 \log(p(\underline{y}|\underline{\beta}, \tau))}{\partial \underline{\beta}^2} & \frac{\partial^2 \log(p(\underline{y}|\underline{\beta}, \tau))}{\partial \tau \partial \underline{\beta}} \\ \frac{\partial^2 \log(p(\underline{y}|\underline{\beta}, \tau))}{\partial \underline{\beta} \partial \tau} & \frac{\partial^2 \log(p(\underline{y}|\underline{\beta}, \tau))}{\partial \tau^2} \end{pmatrix} = \begin{pmatrix} -\tau \mathbf{X}^T \mathbf{X} & \mathbf{X}^T (\underline{y} - \mathbf{X}\underline{\beta}) \\ (\underline{y} - \mathbf{X}\underline{\beta})^T \mathbf{X} & -\frac{n}{2\tau^2} \end{pmatrix}$$

Por lo tanto:

$$I(\underline{\beta}, \tau) = -\mathbb{E} \left(\begin{pmatrix} -\tau \mathbf{X}^T \mathbf{X} & \mathbf{X}^T (\underline{y} - \mathbf{X}\underline{\beta}) \\ (\underline{y} - \mathbf{X}\underline{\beta})^T \mathbf{X} & -\frac{n}{2\tau^2} \end{pmatrix} \right) = \begin{pmatrix} -\tau \mathbf{X}^T \mathbf{X} & 0 \\ 0 & -\frac{n}{2\tau^2} \end{pmatrix}$$

Por lo que, la distribución inicial es:

$$\begin{aligned} p(\underline{\beta}, \tau) &\propto \sqrt{\det(I(\underline{\beta}, \tau))} \propto \sqrt{\det(\tau \mathbf{X}^T \mathbf{X}) \det\left(\frac{n}{2\tau^2}\right)} \propto \sqrt{\tau^{p-2} \det(\mathbf{X}^T \mathbf{X})} \frac{n}{2} \\ &\propto \tau^{\frac{p-2}{2}} \sqrt{\det(\mathbf{X}^T \mathbf{X})} \frac{n}{2} \propto \tau^{\frac{p-2}{2}} \end{aligned}$$

Nótese que es una distribución impropia, lo cual podría generar problemas para nuestro modelo posterior, si no se logra que sea propia. Aplicando el teorema de Bayes con la inicial de Jeffreys se tiene:

$$\begin{aligned} p(\underline{\beta}, \tau | \underline{y}) &\propto (\underline{y} | \underline{\beta}, \tau) p(\underline{\beta}, \tau) \\ &\propto \tau^{n/2} \exp \left\{ -\frac{\tau}{2} \left((\underline{y} - \mathbf{X}\hat{\underline{\beta}})^T (\underline{y} - \mathbf{X}\hat{\underline{\beta}}) + (\hat{\underline{\beta}} - \underline{\beta})^T \mathbf{X}^T \mathbf{X} (\hat{\underline{\beta}} - \underline{\beta}) \right) \right\} \tau^{\frac{p-2}{2}} \end{aligned}$$

$$\propto \underbrace{\tau^{p/2} \exp\left\{(\underline{\beta} - \hat{\underline{\beta}})^T \mathbf{X}^T \mathbf{X} (\underline{\beta} - \hat{\underline{\beta}})\right\}}_{N_p(\underline{\beta} | \hat{\underline{\beta}}, \tau \mathbf{X}^T \mathbf{X})} \underbrace{\tau^{n/2-1} \exp\left\{-\frac{\tau}{2} (\underline{y} - \mathbf{X} \hat{\underline{\beta}})^T (\underline{y} - \mathbf{X} \hat{\underline{\beta}})\right\}}_{Gamma\left(\tau | \frac{n}{2}, \frac{1}{2} (\underline{y} - \mathbf{X} \hat{\underline{\beta}})^T (\underline{y} - \mathbf{X} \hat{\underline{\beta}})\right)}$$

Con la inicial de Jeffreys obtenemos que la distribución final es una distribución Normal-Gamma.

$$p(\underline{\beta}, \tau | \underline{y}) = NormalGamma\left(\underline{\beta}, \tau \mid \hat{\underline{\beta}}, \mathbf{X}^T \mathbf{X}, \frac{n}{2}, \frac{1}{2} (\underline{y} - \mathbf{X} \hat{\underline{\beta}})^T (\underline{y} - \mathbf{X} \hat{\underline{\beta}})\right)$$

En este caso de la inicial de Jeffreys, los parámetros de la distribución final son:

1. $\mu_1 = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \underline{y}) = \hat{\underline{\beta}}_{MV}$
2. $\mathbf{P}_1 = \mathbf{X}^T \mathbf{X}$
3. $\alpha_1 = \frac{n}{2}$
4. $\delta_1 = \frac{1}{2} (\underline{y} - \mathbf{X} \mu_1)^T (\underline{y} - \mathbf{X} \mu_1)$

es decir, se logra haciendo en la inicial $\mathbf{P}_0 = \mathbf{0}$, $\alpha_0 = 0$ y $\delta_0 = 0$

Inferencia sobre τ

De la sección anterior, sabemos que la distribución marginal de τ y de $\sigma^2 = \tau^{-1}$ tomando la distribución inicial de Jeffreys está dada por:

$$p(\tau | \underline{y}) = Gamma\left(\tau \mid \frac{n}{2}, \frac{n-p}{2} \hat{\sigma}^2\right) \implies p(\sigma^2 | \underline{y}) = GammaInversa\left(\sigma^2 \mid \frac{n}{2}, \frac{n-p}{2} \hat{\sigma}^2\right)$$

Estimación puntual sobre τ

1. $\mathbb{E}(\hat{\tau}) = \frac{\alpha_1}{\delta_1} = \frac{n}{(n-p)\hat{\sigma}^2} = \hat{\sigma}^{-2} MV$
2. $Moda(\hat{\tau}) = \frac{\alpha_1 - 1}{\delta_1} = \frac{n-1}{(n-p)\hat{\sigma}^2}$
3. $\mathbb{E}(\hat{\sigma}^2) = \frac{\delta_1}{\alpha_1 - 1} = \frac{n-p}{n-2} \hat{\sigma}^2$
4. $Moda(\hat{\sigma}^2) = \frac{\delta_1}{\alpha_1 + 1} = \frac{n-p}{n+2} \hat{\sigma}^2$

Intervalo de credibilidad para τ : No tiene forma cerrada.

Inferencia sobre $\underline{\beta}$

De la sección anterior, sabemos que la distribución marginal de $\underline{\beta}$ tomando la distribución inicial de Jeffreys está dada por:

$$p(\underline{\beta}|\underline{y}) = T_p \left(\underline{\beta} \middle| n, \hat{\underline{\beta}}_{MV}, \frac{n}{n-p} \hat{\sigma}^{-2} \mathbf{X}^T \mathbf{X} \right)$$

Estimación puntual sobre $\underline{\beta}$

Utilizando la función de pérdida cuadrática nuevamente, es decir, utilizando la media de la distribución posterior, se tiene que el estimador puntual del parámetro $\underline{\beta}$ está dado por:

$$\hat{\underline{\beta}} = \mu_1 = \hat{\underline{\beta}}_{MV}$$

Intervalo de credibilidad para $\underline{\beta}$

Ya que la densidad T-multivariada pertenece a las distribuciones elípticas sabemos que:

$$R^2 = \left(\underline{\beta} - \hat{\underline{\beta}}_{MV} \right)^T \frac{n}{n-p} \hat{\sigma}^{-2} \mathbf{X}^T \mathbf{X} \left(\underline{\beta} - \hat{\underline{\beta}}_{MV} \right) \sim pF_{(p,n)}$$

Por lo tanto:

$$\mathbb{P} \left(\frac{1}{p} \left(\underline{\beta} - \hat{\underline{\beta}}_{MV} \right)^T \frac{n}{n-p} \hat{\sigma}^{-2} \mathbf{X}^T \mathbf{X} \left(\underline{\beta} - \hat{\underline{\beta}}_{MV} \right) \leq F_{(p,n)}^{1-\alpha} \right) = 1 - \alpha$$

1.4. Cointegración

Tomando la notación de Campbell y Perron (1991) [22], se dice que 2 series de tiempo, x_t, y_t , están cointegradas si son integradas de orden 1 (de forma abreviada $I(1)$) y existe $\alpha \in \mathbb{R}$ tal que

$$u_t = y_t - \alpha x_t$$

es un proceso estacionario. Decir que x_t y y_t sean $I(1)$, nos indica que individualmente deben de ir a la deriva o con una cierta tendencia, pero esa tendencia o comportamiento de cada una, debe de estar relacionado con el de la otra. Es importante esta característica cuando hablamos de cointegración. Consideremos los siguientes casos:

1. Consideremos la siguiente ecuación de regresión:

$$y_t = \alpha_0 + \alpha_1 x_t + u_t,$$

donde x_t y y_t son caminatas aleatorias independientes una de la otra. Es claro que α_1 es igual a 0, y que la distribución del estimador $\hat{\alpha}_1$ converge a un movimiento Browniano. Esta regresión espuria fue descubierta por Monte Carlo y estudiada por Granger y Newbold (1974) [12]. Este ejemplo traería inconsistencia en las estimaciones, dado que ellos demostraron que el estadístico t diverge a una tasa \sqrt{T} , por lo que el problema se detecta al momento de estimar por mínimos cuadrados el vector α y se obtiene un valor de R^2 alto y un valor de Durbin-Watson bajo. Normalmente aquí se necesita revisar el problema de raíces unitarias, dado que la relación es sospechosa.

2. Considere el ejemplo anterior, suponiendo ahora que $\alpha_1 \neq 0$ y que u_t tiene raíz unitaria. En este caso igualmente se tendrían relaciones inconsistentes y se tendría que estimar la relación:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + e_t.$$

A esto también se le conoce como relación espuria.

3. El último ejemplo consiste en el planteamiento inicial, suponiendo que u_t es estacionario. En este caso $\hat{\alpha}_1$ converge al valor real a tasa T . Se dice que el estimador por mínimos cuadrados es superconsistente. Se puede demostrar que $T * (\hat{\alpha}_1 - \alpha_1)$ distribuye asintóticamente como $\int_0^1 B_2 dB_1 / \int_0^1 B_2^2 dt$, donde B_1 y B_2 son dos movimientos Brownianos independientes; esta distribución tiene media cero, pero más aún la prueba t distribuye asintóticamente normal.

Se dice que un proceso estocástico es integrado de orden p , $I(p)$, si se debe diferenciar el proceso p veces para garantizar estacionariedad. En general, x_t y y_t se dice que son cointegrados de orden $CI(d, p)$, si x_t y y_t son integrados de orden d y existe α tal que $y_t - \alpha x_t$ es integrada de orden $d - p$. Nosotros hablaremos de “cointegración” a las relaciones $CI(1, 1)$.

Definición. Un vector de series de tiempo $I(1)$, y_t , se dice que está cointegrado si existe un vector β_i , tal que $\beta_i^T y_t$ es estacionario. Si existen r vectores linealmente independientes $\beta_i, i = 1, \dots, r$, entonces se dice que y_t está cointegrado con rango de cointegración r . La matriz $B = (\beta_1, \dots, \beta_r)$ se le conoce como la matriz de cointegración.

1.4.1. Caso Multivariado. Prueba de Johansen

Para generalizar las ideas anteriores al caso multivariado, utilizaremos la notación de Gerald P. Dwyer (2015) [6]. El vector autorregresivo VAR en niveles y sin constante puede ser escrito de la forma:

$$x_t = \sum_{i=1}^k A_i x_{t-i} + u_t$$

Para $k > 1$, este VAR siempre puede ser escrito de la forma:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Pi_i \Delta x_{t-i} + u_t$$

Si Π fuese cero, es trivial notar que no hay cointegración. Este es el modelo implícito en la prueba Box-Jenkins. Si Π tuviese rango completo, y_t debe ser estacionario dado que los dos lados de la ecuación lo son (limitándonos a series $I(1)$ y $I(0)$).

El caso más interesante y que aborda Johansen (1995) [14] es cuando Π no tiene rango completo pero no es cero. A este caso le llamamos cointegración. Aquí, Π se puede escribir como $\Pi = \alpha\beta^T$ (con la β introducida en la sección anterior), y α, β son matrices $n \times r$. Nótese que α y β no tienen una representación única. Más aún $\Pi = \alpha\beta^T = \alpha F^{-1}(\beta F^T)^T$ para alguna matriz F no singular. Este problema de normalización de β (y por consiguiente α) es la parte más complicada de realizar.

Como se mencionó en la sección anterior, si las variables son cointegradas, el rango de $\Pi \neq 0$, de hecho, el rango de Π coincide con el número de vectores de cointegración, además, este número es menor o igual al número de series de tiempo y estrictamente menor al número de series de tiempo si las variables tienen raíces unitarias. Los valores propios son útiles para resolver este problema porque el determinante de una matriz cuadrada es igual al producto de los valores

propios. Si el rango de la matriz es menor que el número de filas y columnas de la matriz, entonces uno o más valores propios son cero y el determinante es cero.

Las pruebas de Johansen se basan en los valores y vectores propios de transformaciones específicas de los datos y utilizan combinaciones lineales que tienen la máxima correlación (correlaciones canónicas). Utilizando estas transformaciones se garantiza que los valores propios son no negativos y reales. Entrar en este detalle, sale del alcance de este documento. Supóngase que se calcula de forma explícita los n valores propios y se ordenan de forma descendente, es decir, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ y recordemos que $\lambda_i > 0$ para todo i . Si $\lambda_1 = 0$, el rango de Π es cero y no existen vectores de cointegración. Si $\lambda_1 > 0$, el rango de Π es por lo menos 1, por lo que existe por lo menos 1 vector de cointegración. Iterativamente, si $\lambda_{n-1} > 0$, se prueba si $\lambda_n = 0$. Si $\lambda_n = 0$, entonces existen $n - 1$ vectores de cointegración.

Las pruebas de Johansen, consideradas como una generalización multivariada de la prueba aumentada de Dickey-Fuller, se denominan prueba del valor propio máximo y prueba de la traza. Sea r el rango de Π . Sabemos por la discusión anterior, que r coincide con el número de vectores de cointegración. Se realizan 2 pruebas por medio de cocientes de verosimilitudes. La primera se denomina prueba del valor propio máximo y la segunda de la traza. Para ambas pruebas, la prueba inicial consiste en contrastar la hipótesis nula de no cointegración, con la hipótesis alternativa de cointegración.

Prueba de máximo valor propio

Esta prueba examina si el valor propio más grande es cero en relación con la hipótesis alternativa de que el siguiente valor propio más grande sea cero. La primera prueba consiste en contrastar la hipótesis que el rango de $\Pi = 0$ en relación con la alternativa de que el rango de $\Pi = 1$. Para las siguientes pruebas, la hipótesis nula será que $\text{rango}(\Pi) = 1, 2, \dots$ y la hipótesis alternativa será que $\text{rango}(\Pi) = 2, 3, \dots$. Profundizando la primera prueba $\text{rango}(\Pi) = 0$, si el rango de la matriz Π es cero, el valor propio más grande, λ_1 , es cero, y no existe cointegración en las series. Si λ_1 no es cero, el rango de la matriz es por lo menos 1 y podría haber más vectores de cointegración. Así sucesivamente hasta que la hipótesis nula de que el n -ésimo valor propio es igual a cero no pueda ser rechazada. La prueba que estamos mencionando es una prueba basada en el cociente de verosimilitud y su estadístico está dado por

$$LR(r_0, r_0 + 1) = -T \ln(1 - \lambda_{r_0+1})$$

Donde $LR(r_0, r_0 + 1)$ es el estadístico resultante del cociente de verosimilitud donde $\text{rango}(\Pi) = r_0$ contra la hipótesis alternativa donde $\text{rango}(\Pi) = r_0 + 1$. Lamentablemente y similar a la prueba Dickey-Fuller, este estadístico no tiene una distribución asintótica usual de χ^2 .

Prueba de la Traza.

Esta prueba, “trace test” en inglés, se basa en que el rango de la matriz Π es r_0 . La hipótesis nula es $\text{rango}(\Pi) = r_0$ mientras que la hipótesis alternativa es que $r_0 < \text{rango}(\Pi) \leq n$. La siguiente hipótesis nula es $\text{rango}(\Pi) = r_0 + 1$ mientras que la hipótesis alternativa es que $r_0 + 1 < \text{rango}(\Pi) \leq n$ y así sucesivamente. El estadístico de prueba basado en el cociente de verosimilitud, el cual es similar a la prueba de máximo valor propio, está dado por

$$LR(r_0, n) = -T \sum_{i=r_0+1}^n \ln(1 - \lambda_i)$$

donde aquí $LR(r_0, n)$ es el cociente de verosimilitud referente a la hipótesis $\text{rango}(\Pi) = r$ en contra la hipótesis alternativa de $\text{rango}(\Pi) \leq n$.

Cabe destacar que esta prueba no tiene relación con la traza de Π , sino se llama de esta forma dado que la distribución asintótica del estadístico de prueba es la traza de una matriz basada en funciones de movimientos brownianos (estándar de Wiener) el cual se aborda en Econometrica 1983 [19] [18].

Capítulo 2

Marco Conceptual

2.1. Producto Interno Bruto en México

En economía política, existen diversos indicadores para medir el desempeño de la economía del país. Estos índices nos describen el comportamiento de la coyuntura económica y se miden con información nacional, por ejemplo: la evolución mensual de la producción industrial, los resultados del comercio exterior, el desempeño de las variables fiscales, financieras y monetarias, así como índices de precios al productor o al consumidor. El PIB, según el Banco Mundial, se define como: “la suma del valor añadido bruto de todos los productores residentes en la economía, más los impuestos sobre los productos, menos las subvenciones no incluidas en el valor de estos. Se calcula sin hacer deducciones por la depreciación de los activos fabricados ni por el agotamiento y la degradación de los recursos naturales.”

2.1.1. PIB Trimestral

Para dar un seguimiento más oportuno, el INEGI a partir de 1987, publicó el primer cálculo trimestral del PIB (PIBT), bajo el marco conceptual del Sistema de Cuentas Nacionales de México [13], el cual consiste en elaborar índices mensuales o trimestrales de volumen físico de la producción para cada una de las clases seleccionadas, con base fija en el año 2013, con estos índices se realiza la extrapolación de la producción del año base. EL PIBT, lo que nos reporta trimestre a trimestre, es un cálculo oportuno del PIB anual, logrando que el promedio de los 4 crecimientos anuales del PIBT al año T coincide muy cercanamente con el crecimiento anual del PIB anual al año T.

Los valores de la serie PIBT están a base de precios del 2013 con el fin de que sean comparables unas con otras, se quita el efecto del aumento generalizado de precios (inflación) continuando con las metodologías del Sistema de Cuentas Nacionales de México. Además, añade los lineamientos internacionales sobre contabilidad nacional, que se han establecido en el Manual del Sistema de Cuentas Nacionales 2008 (SCN2008) y en el Manual de Cuentas Nacionales Trimestrales, Conceptos, Fuentes de Datos y Compilación 2001 (CNT 2001) editado por el Fondo Monetario Internacional [1].

Es importante destacar que la publicación del PIBT tiene un rezago de varias semanas (5 semanas en promedio), reto para los actuarios y economistas, ya que, tener un buen pronóstico del PIB puede darnos información valiosa y **oportuna** de la actividad económica del país y así una anticipada toma de decisiones. En la Figura 2.1, se observa la serie de tiempo del PIBT del primer trimestre de 1993 al cuarto trimestre de 2021.

2.2. Sistemas de pagos en México

Un sistema de pagos, según Padoa-Schioppa (2004) [16] es el conjunto de instrumentos, reglas, procedimientos y tecnologías utilizadas para liquidar las transferencias de dinero entre los agentes económicos. Podemos distinguir entre pagos al por mayor (wholesale) y pagos al por menor (retail). En los primeros, el sistema bancario suele gestionar pagos de gran cuantía (transacciones interbancarias), normalmente relacionados con los flujos de los mercados financieros y las operaciones de refinanciación con los bancos centrales; los segundos se refieren a las transacciones dentro del circuito de particulares y empresas y están estrechamente relacionadas con la actividad económica.

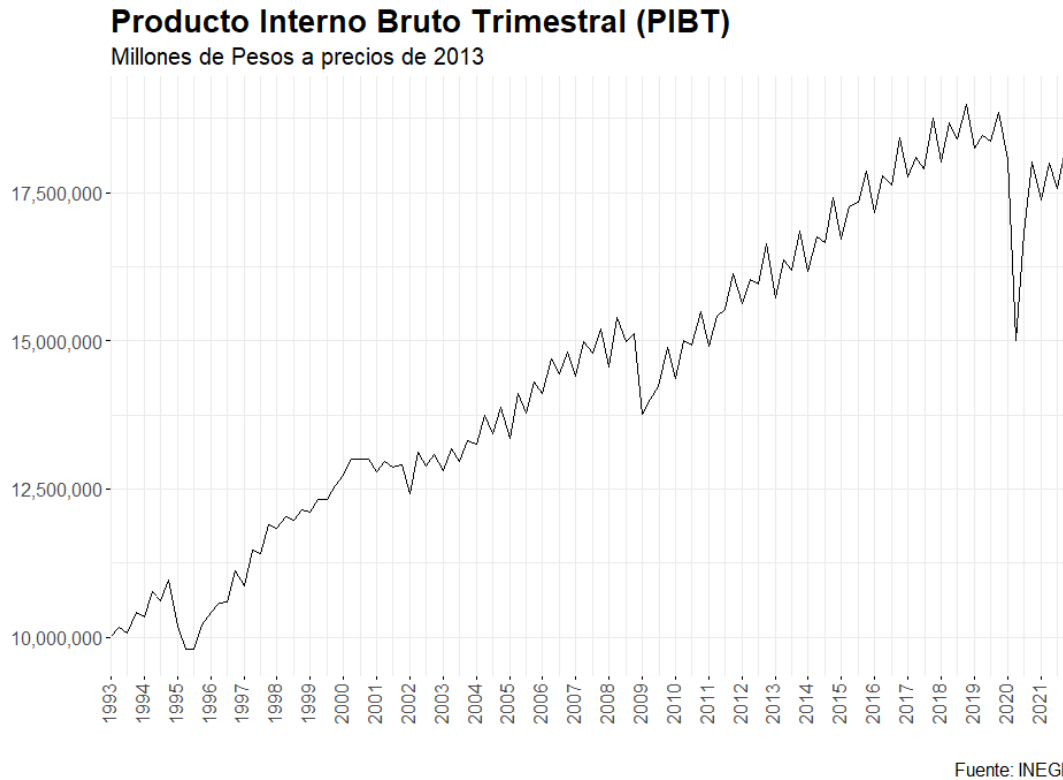


Figura 2.1: Producto Interno Bruto trimestral (PIBT), del primer trimestre de 1993 al cuarto trimestre de 2021. Millones de pesos a precios de 2013.

Según el Panorama anual de Inclusión Financiera (IF) [20] de la Comisión Nacional Bancaria y de Valores (CNBV) se cuenta con los siguientes datos al cierre de 2020:

1. De cada 10,000 mexicanos adultos existían 13,932 cuentas de la banca, entidades de ahorro y crédito popular e instituciones de fondos de pago electrónico (Con una meta de 16,680 para 2024).
2. 43.9 transacciones y transferencias electrónicas per cápita anual (Con una meta de 60.4 para 2024).
3. Entre otros que se muestran en la Figura 2.2

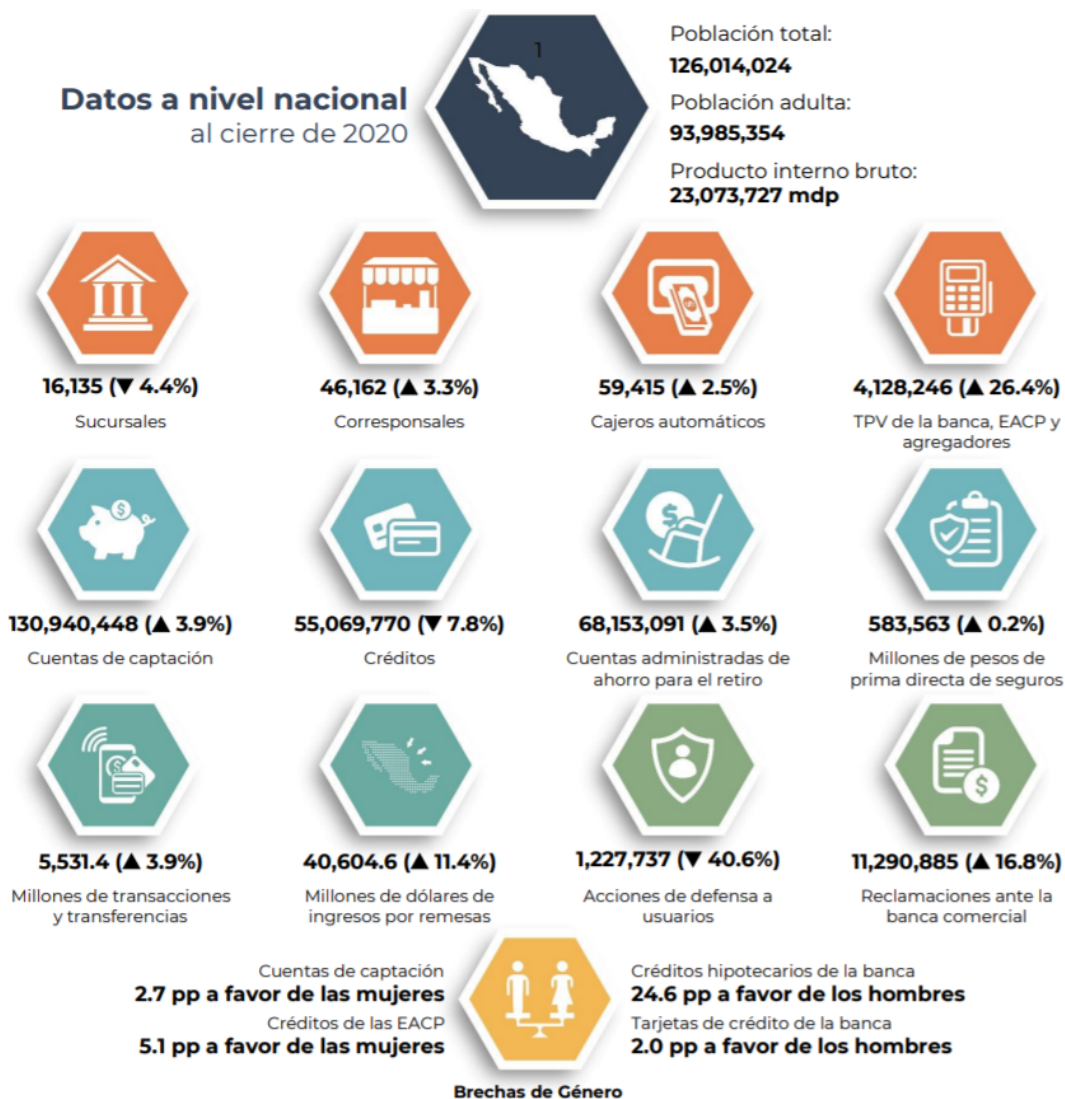


Figura 2.2: Datos relevantes de los sistemas de Pagos a nivel nacional al cierre de 2020. Figura tomada de la introducción del Programa Anual de Inclusión Financiera. [20] Fuente: CNBV.

En la Figura 2.3, tomada del mismo Panorama Anual IF [20], podemos observar un aumento generalizado en el número de operaciones liquidadas a través de los sistemas de pagos electrónicos en México en los últimos 5 años.

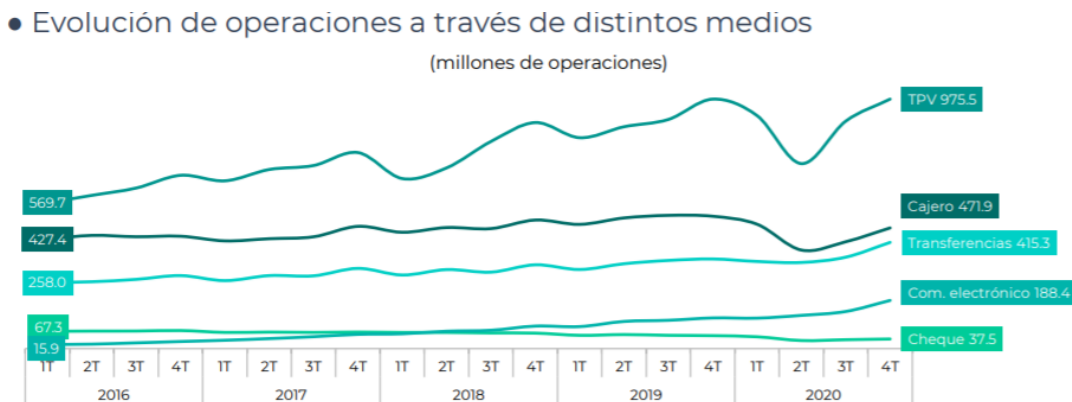


Figura 2.3: Evolución del número de operaciones anuales en diversos medios de pago electrónico. Fuente: Panorama Anual IF CNBV y Banxico 2016-2020.

2.2.1. SPEI

El 13 de agosto de 2004 comenzó a operar el Sistema de Pagos Electrónicos Interbancarios (SPEI), es el sistema o infraestructura de pagos más grande e importante en México en cuanto a número de operaciones y en monto. En él se procesan todas las transacciones interbancarias (entre bancos, casas de bolsa, sofipos y otras entidades financieras reguladas) en el país, desde los montos más bajos realizados por el público general (comúnmente conocido como transacciones Tercero a Tercero o TaT), como las transacciones de montos altos realizadas por las empresas y las instituciones financieras [15]. El SPEI operaba de 6:00 am a 5:30 pm hasta el día 28 de diciembre de 2017: “A partir de esta fecha todos los bancos tendrán la obligación de procesar los pagos SPEI de menos de \$8,000 pesos en tiempo real, todos los días, las 24 horas”, informó Lorenza Martínez, directora general de Sistema de Pagos y Servicios Corporativos del Banco de México en ese entonces.

Para dimensionar el tamaño del SPEI, a continuación, se presentan 2 Figuras ejemplificando el número y monto liquidado en el Sistema de Pagos.

1. En la Figura 2.4 se observa el número de operaciones TaT liquidadas anualmente del 2011 al 2020; se observa un crecimiento acelerado en las operaciones en los últimos años, y nos habla de que el proceso de adopción del sistema de pagos es cada vez mayor en la población general (el 82 % de las operaciones del 2020 fueron menores a \$8,000).
2. En la Figura 2.5 se observa la proporción del monto liquidado en transacciones SPEI TaT y

el PIB anual, destacando 4.7 para 2021.

El Banco de México reporta de forma semanal el agregado del número y monto transaccionado en el SPEI TaT en una estructura diaria; se divide en 3 cortes: los pagos menores a \$8,000 pesos (los cuales la regulación obliga a todos los participantes del SPEI a que se puedan enviar y recibir pagos de este monto las 24 horas de los 365 días del año), los pagos entre \$8,000 y \$300,000 pesos y los pagos mayores de \$300,000 pesos.

2.2.2. Tarjetas

Retomando la Figura 2.3, y exceptuando el periodo de completo confinamiento a raíz de la pandemia de COVID-19 (segundo trimestre del 2020), el mercado de tarjetas crece de manera exponencial (tasa promedio de crecimiento anual de 12 % según el Banco de México) día con día.

Evolución del número de transferencias en el SPEI, 2011-2020 entre usuarios finales del sistema
Millones de operaciones

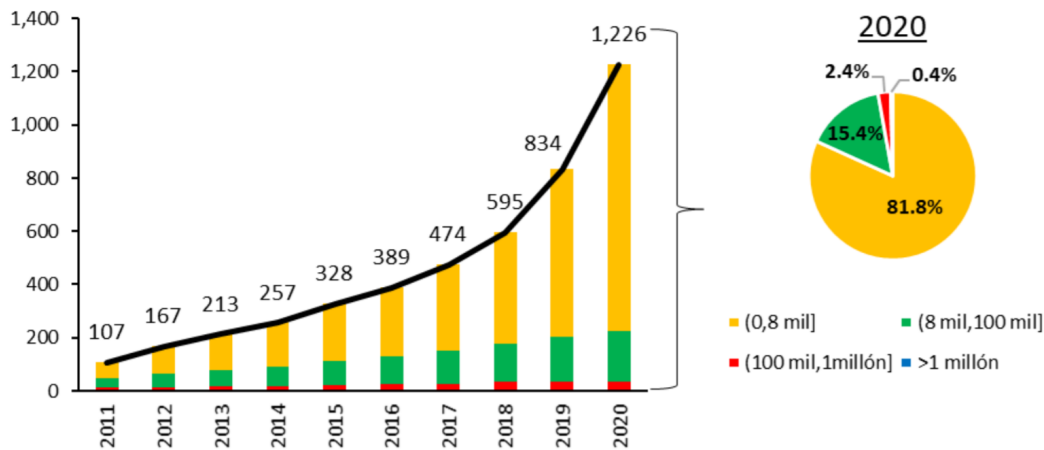


Figura 2.4: Número de operaciones SPEI TaT liquidadas al año. Fuente: Informe Anual sobre las Infraestructuras de los Mercados Financieros 2020 del Banco de México. [15]

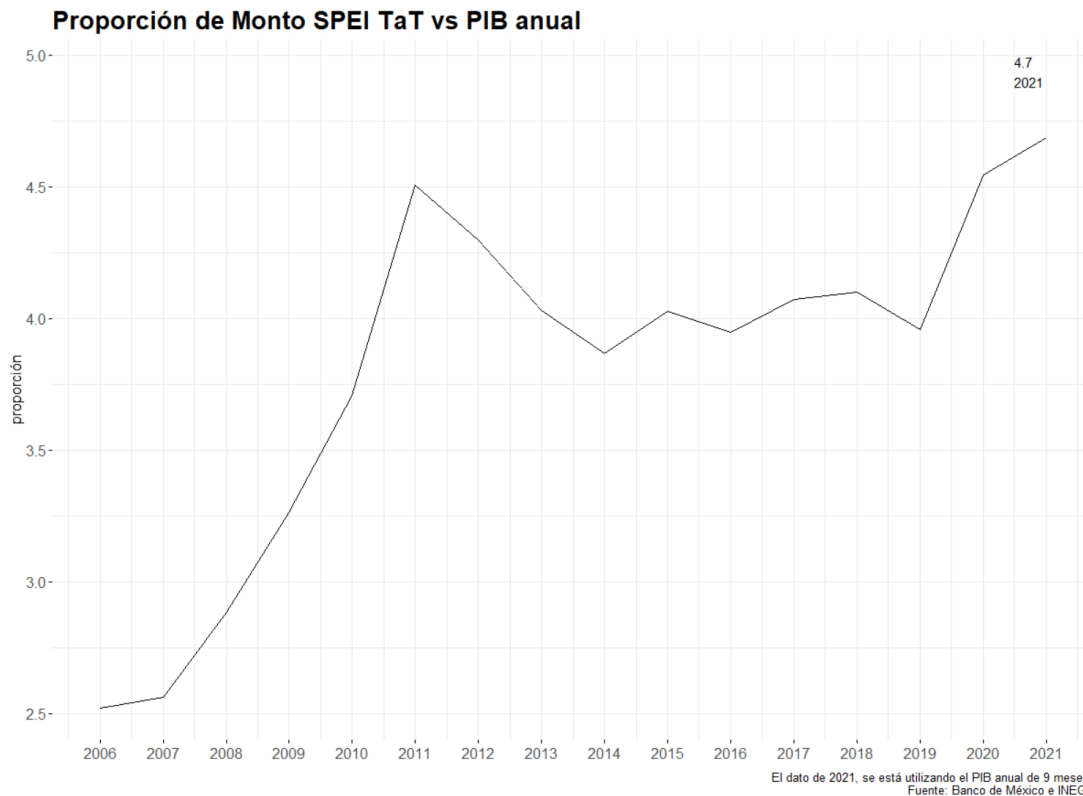


Figura 2.5: Proporción del Monto liquidado en operaciones SPEI TaT dividido por el PIB anual. Fuente: Banco de México e INEGI.

“En 1968 se emitió la primera tarjeta de crédito, llamada Bancomático, en México lanzada por el Banco Nacional de México (Banamex). En seguida en 1969, Bancomer emitió su tarjeta afiliada al sistema BankAmericard. En ese mismo año, un grupo de bancos emitió la tercera tarjeta bancaria en el mercado mexicano: Carnet (actualmente MasterCard). Gracias a sus beneficios y a la característica de poder comprar cualquier tipo de objeto o servicio sin tener que pagar en efectivo en el momento, mucha gente optó por ellas. En 1967, el First American National Bank of Nashville puso a disposición una nueva tarjeta de crédito denominada MasterCharge, que en 1979 cambió su nombre por el que ahora todos conocemos como MasterCard” menciona el gobierno de México [5].

Como se puede observar, las tarjetas de crédito y débito llegaron a resolver una necesidad de la población mexicana: disminuir el uso de efectivo en compras de monto considerable, acceso a servicios financieros como: préstamos, créditos hipotecarios, inversiones, compra de activos financieros, entre otros; lo que comúnmente se conoce como Inclusión Financiera.

Capítulo 3

Metodología

Desde la gran recesión de 2008, la necesidad de nuevas herramientas de previsión macroeconómica, principalmente basadas en información monetaria y financiera ha aumentado considerablemente, existen estudios en diversas partes del mundo entre ellos los que destacan son: Canadá (Galbraith y Tkacz 2007, 2009, 2018) [10], Portugal (Esteves 2009; Duarte, Rodrigues y Rua 2017) [8], Dinamarca (Carlsen y Storgaard 2010) [4], Estados Unidos (Barnett et al. 2016) [3], Italia (Aprigliano, et al. 2019) [2] y por supuesto India (Sudhakar Raju et al. 2019) [17]. Dichos artículos encuentran que las operaciones electrónicas pueden ayudar a estimaciones del Producto Interno Bruto y del consumo privado a corto plazo. En el caso de este documento, la intención es de realizar análisis similares con información de México. La oportunidad de tener la información de SPEI y Tarjetas de manera diaria y oportuna, nos hace pensar que es información valiosa y útil para dar pronósticos a diversas variables reales macroeconómicas, por ejemplo, el Indicador Mensual de la Actividad Industrial, el Índice Global de la Actividad Económica (IGAE) y por supuesto, el Producto Interno Bruto.

3.1. Relación entre PIBT e información de los Sistemas de Pagos

El objetivo principal de este documento es argumentar que la información de pagos es útil para dar estimaciones confiables del PIBT. Se cuenta con información de SPEI TaT y Tarjetas desde el 1 de enero del 2009, por lo que se tomará la información desde el primer trimestre del

2009 al cuarto trimestre del 2019 para el siguiente desarrollo.

3.1.1. Preparación de la información

SPEI TaT

Como se mencionó en el capítulo anterior, se cuenta con el número y monto transaccionado en el SPEI TaT en 3 grupos:

1. Pagos menores a \$8,000. Monto y número: (nosa.real.SPME8K y nosa.vol.SPME8K)
2. Pagos entre \$8,000 y \$300,000. Monto y número:
(nosa.real.SP8KY300K y vol.SP8KY300K)
3. Pagos mayores a \$300,000. Monto y número: (nosa.real.SPMA300K y vol.SPMA300K)

Para que la información fuese comparable con la del PIBT, primero se procedió a obtener los montos y números acumulados por trimestre. Algo importante es que los montos reportados por Banxico son montos a pesos corrientes o del año correspondiente, por lo que para que sean a pesos constantes (también conocido como reales) de 2013 como lo está el PIBT, se utiliza el Índice Nacional de Precios al Consumidor (INPC) el cual según el INEGI: “mide la variación de los precios de una canasta de bienes y servicios representativa del consumo de los hogares mexicanos”. A continuación, se muestra en las Figuras 3.1 y 3.2 las 6 series de tiempo (3 de monto y 3 de número) que se construyeron. También cada una de ellas, muestra las correlaciones muestrales con el PIBT, valores que son 0.88, 0.98 y 0.78 para monto y 0.8, 0.94 y 0.96 para número. El análisis pertinente de la relación con el PIBT se hablará en la siguiente subsección.

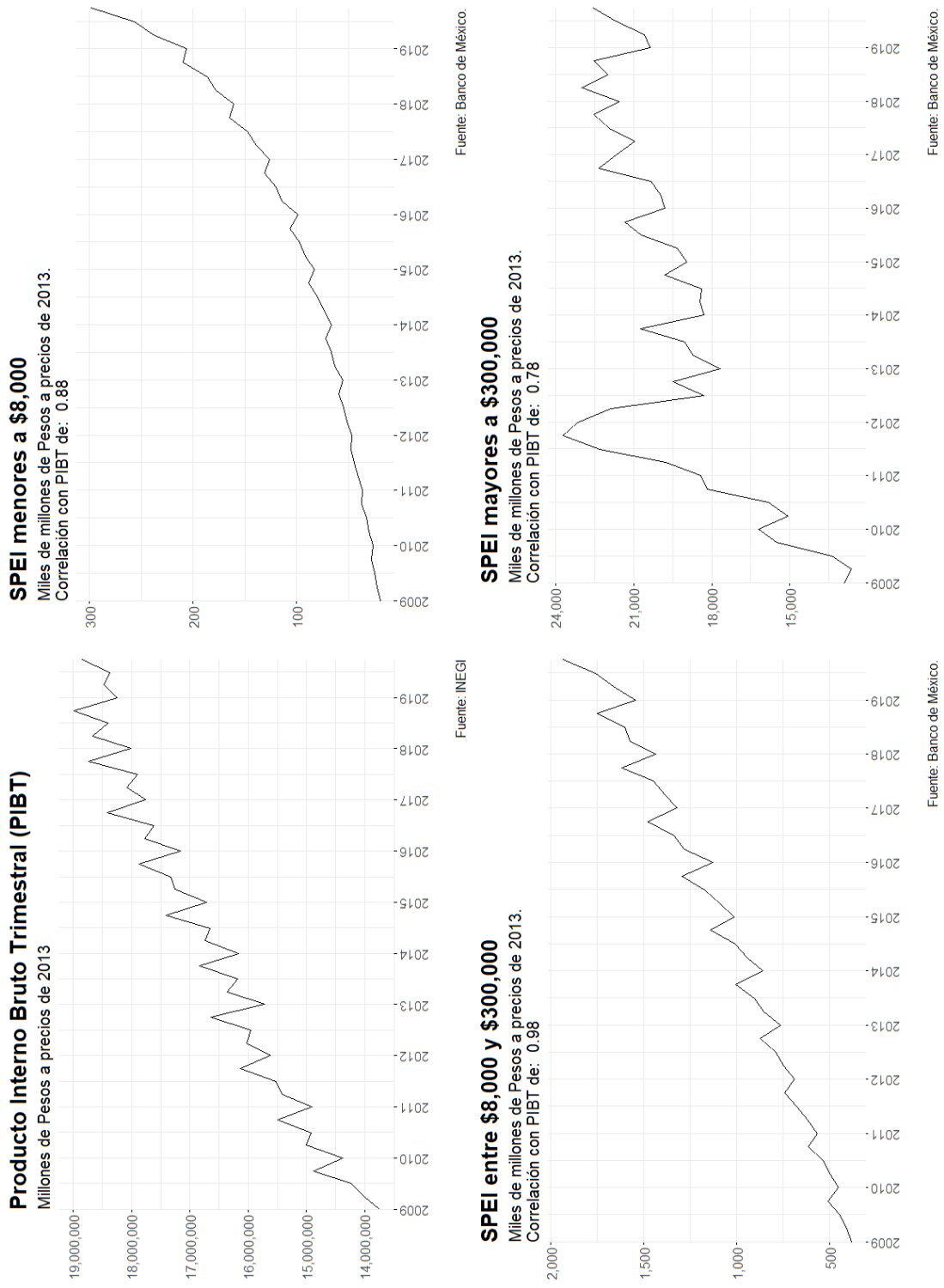


Figura 3.1: Monto transaccionado trimestralmente a través del SPEI en sus 3 grupos por monto. Primer trimestre de 2009 al cuarto trimestre del 2019. Miles de millones de pesos a precios de 2013.

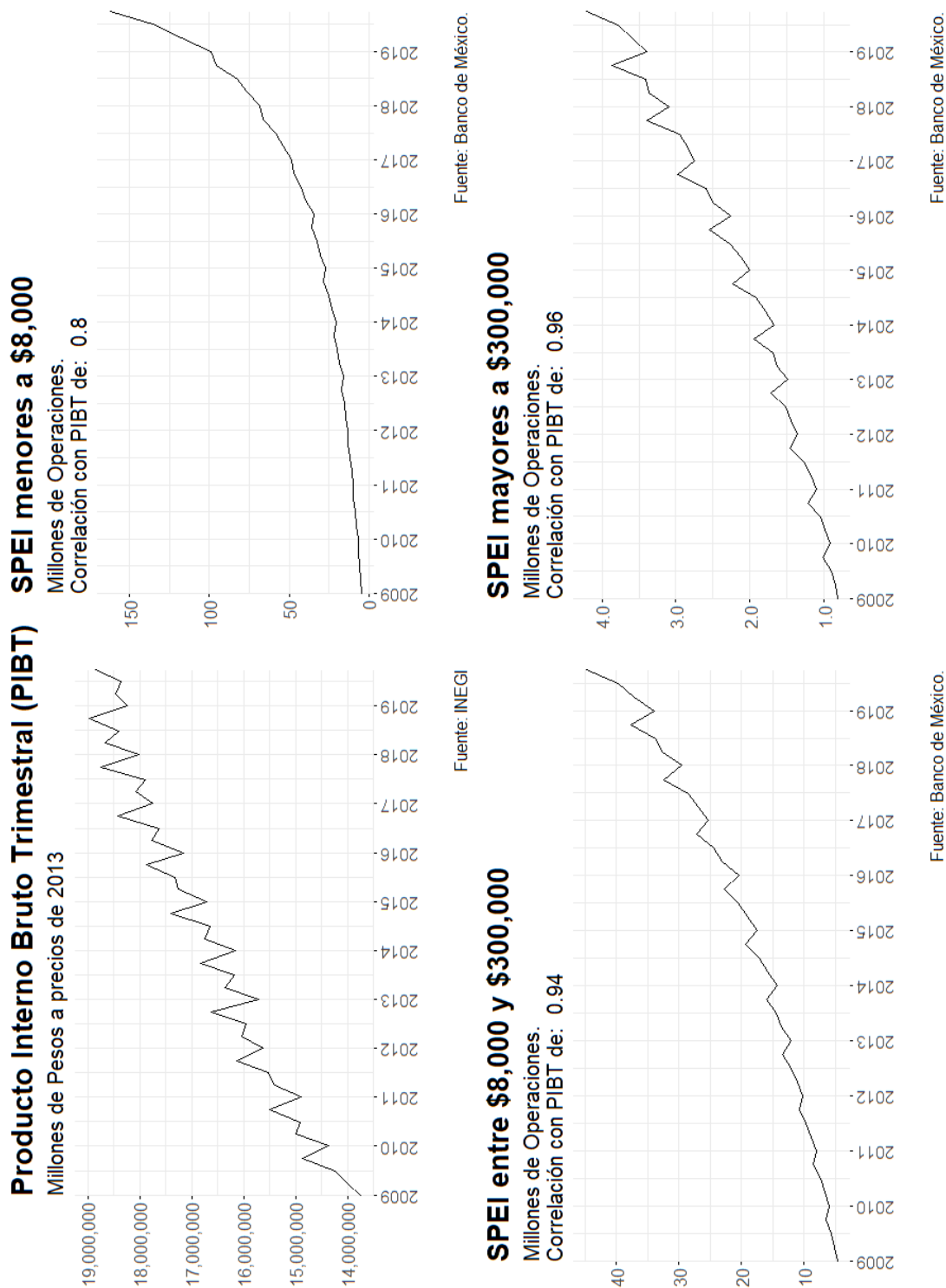


Figura 3.2: Número de operaciones transaccionadas trimestralmente a través del SPEI en sus 3 grupos por monto. Primer trimestre de 2009 al cuarto trimestre del 2019.

Tarjetas

Para la información de tarjetas, se utilizará los totales de monto y número transaccionado con tarjeta de débito y crédito y de igual forma que la información del SPEI, los montos son a pesos constantes de 2013 utilizando el INPC.

A continuación, se muestra en las Figuras 3.3 y 3.4, las 4 series de tiempo (2 de monto y 2 de número) que se construyeron. También cada una de ellas, muestra las correlaciones muestrales con el PIBT, valores que son 0.98 y 0.95 para monto y 0.97 y 0.95 para número.

3.1.2. Selección de variables, estacionariedad y cointegración

Como se mencionó en la subsección anterior, cuando estamos realizando un análisis en el cual un conjunto de covariables o variables exógenas puedan explicar a una variable objetivo dependiente, claramente se debe de analizar si existe correlación entre ellas; la cual, en las Figuras 3.1 a 3.4 se observa que existe y es considerablemente grande, en algunas variables la correlación es cercana a 1. La Figura 3.5 muestra de forma gráfica y numérica, la correlación muestral entre las variables preparadas en la anterior sección y el PIBT. Destacan las variables Monto Tarjetas de Crédito y Monto SPEI entre \$8,000 y \$300,000 como variables con mayor correlación con el PIBT. (Los nombres de las variables se encuentran en la Tabla 3.1).

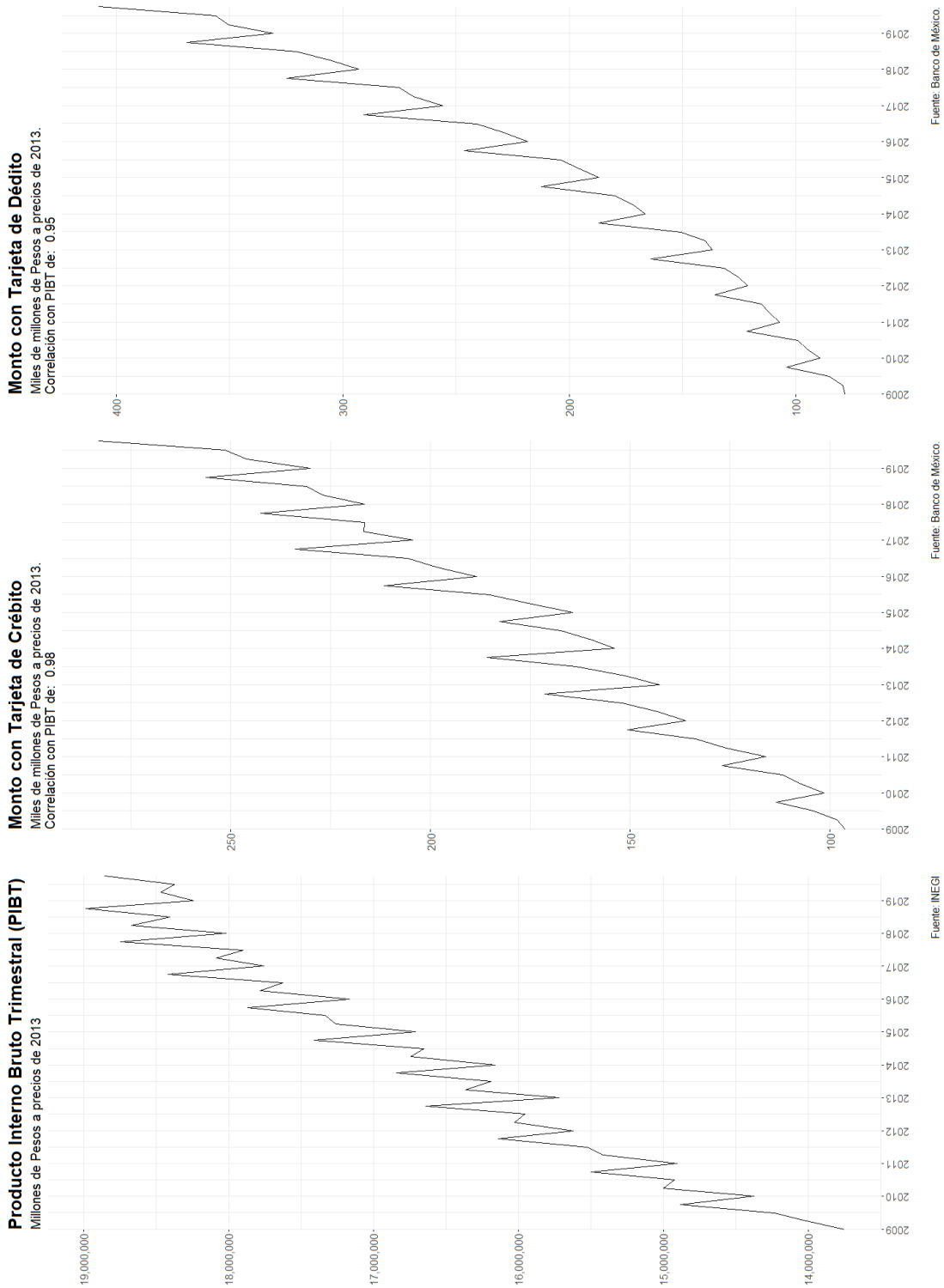


Figura 3.3: Monto transaccionado trimestralmente con tarjeta de crédito y débito. Primer trimestre de 2009 al cuarto trimestre del 2019. Miles de millones de pesos a precios de 2013.

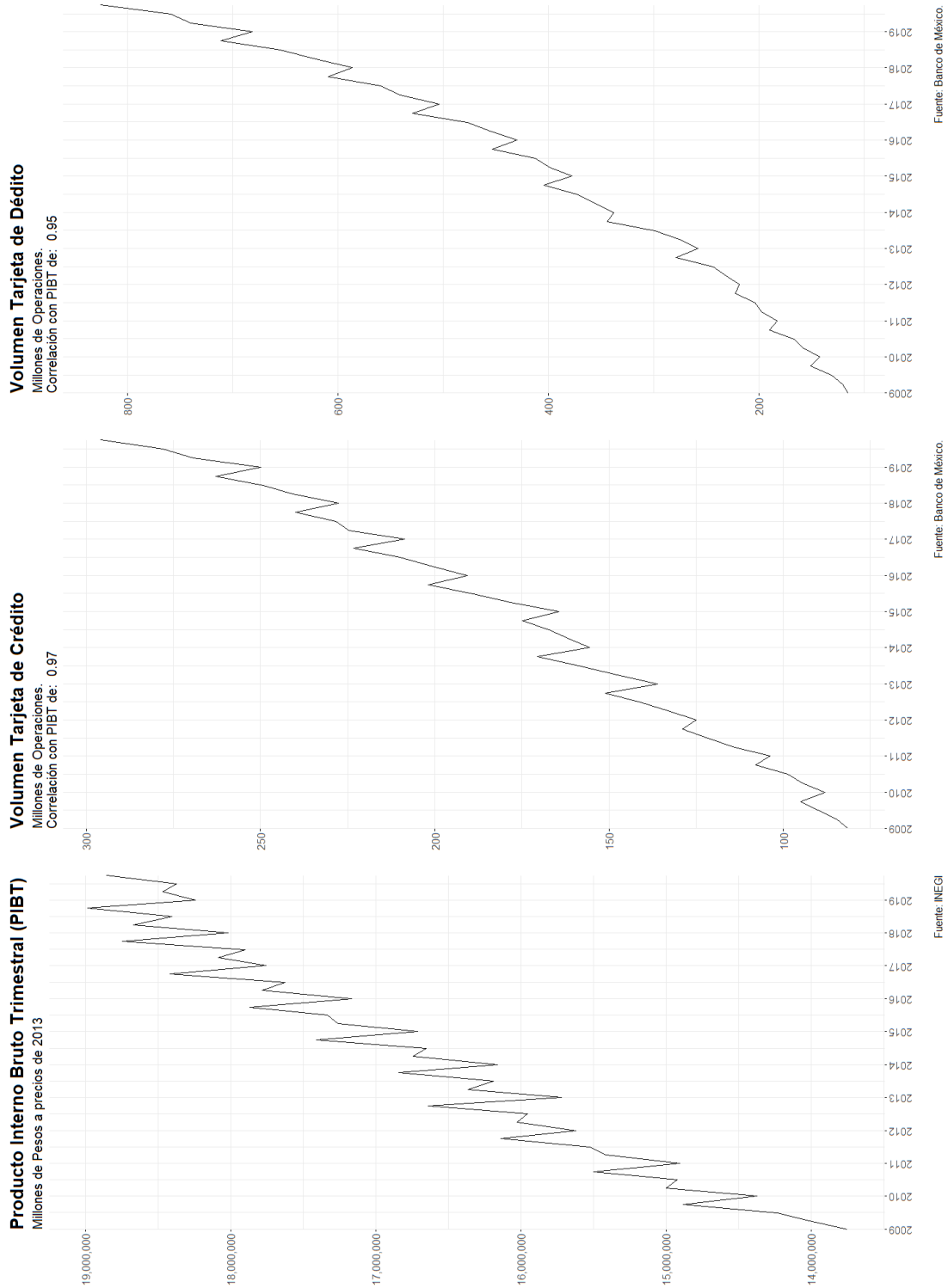


Figura 3.4: Número de operaciones transaccionadas trimestralmente con tarjeta de crédito y débito. Primer trimestre de 2009 al cuarto trimestre del 2019.

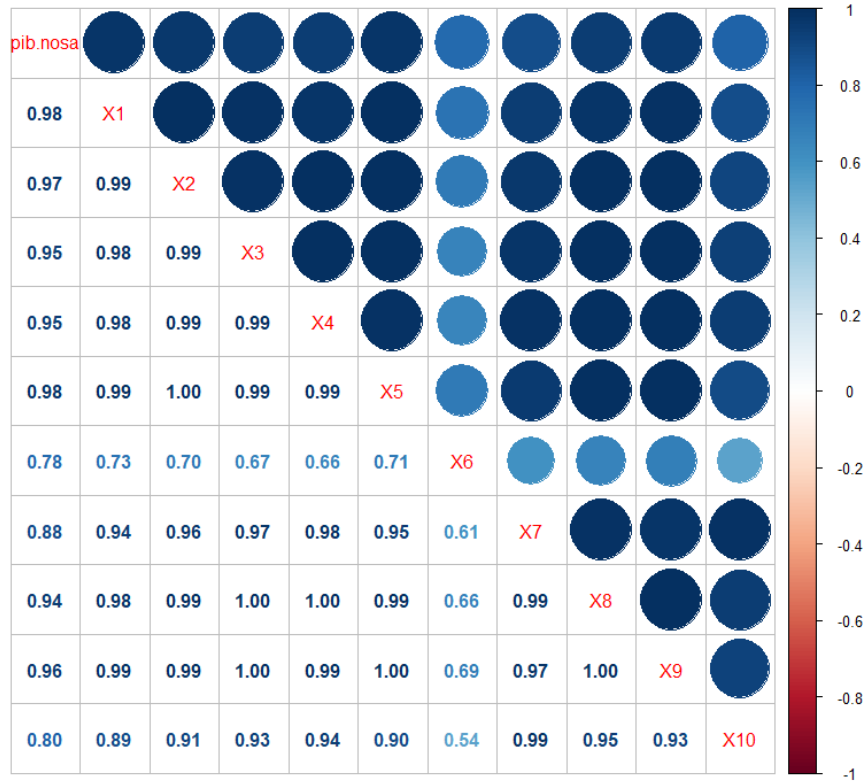


Figura 3.5: Correlación entre las variables de pagos y el PIBT.

donde:

Variable	Nombre completo	Descripción
pib.nosa	pib.nosa	PIBT
X_1	nosa.cards.Credito.M	Monto Tarjetas Crédito
X_2	nosa.cards.Credito.O	Número Tarjetas Crédito
X_3	nosa.cards.Debito.M	Monto Tarjetas Débito
X_4	nosa.cards.Debito.O	Número Tarjetas Débito
X_5	nosa.real.SP8KY300K	Monto SPEI entre 8K y 300K
X_6	nosa.real.SPMA300K	Monto SPEI mayor a 300K
X_7	nosa.real.SPME8K	Monto SPEI menor a 8K
X_8	nosa.vol.SP8KY300KV	Volumen SPEI entre 8K y 300K
X_9	nosa.vol.SPMA300KV	Volumen SPEI mayor a 300K
X_{10}	nosa.vol.SPME8KV	Volumen SPEI menor a 8K

Tabla 3.1: Conjunto de variables utilizadas como potenciales predictores del PIBT. Se está utilizando el prefijo “nosa”, haciendo referencia a que los datos no están ajustados por estacionalidad (abreviatura del inglés).

Selección de variables

Al momento de realizar un modelo de predicción de una variable en función de otras, ya que se realizó un análisis descriptivo y exploratorio, se debe definir métricas y criterios para seleccionar las covariables o variables independientes que se van a utilizar en el mismo. Para esto, los detalles que se van a considerar son:

1. El objetivo primordial de nuestro modelo claramente es que dada la información observada de las covariables hasta el tiempo T y de nuestra variable dependiente al $(T - 1)$, poder empatar a la misma prediciéndola. Por lo anterior, se realizarán ejercicios iterativos, desde t_0 hasta T en los cuales suponiendo que se tenía información hasta al tiempo t_k , $k \in \{0, \dots, T\}$, ajustaremos por mínimos cuadrados un modelo de regresión múltiple donde las variables de pagos expliquen al PIBT; después se realizará un proceso a pasos basándonos en el criterio de Akaike (AIC) para eliminar variables que no aportan información significativa al ajuste de la variable target. Se recopilarán las variables que se preservan la mayor cantidad de veces en los escenarios planteados. Se propone definir a t_0 como la observación que recopila

el 80 % de la muestra y así sucesivamente hasta utilizar el 100 % de la misma.

2. Problemas de multicolinealidad. Como se menciona en el primer punto, se utilizará el método a pasos (Step), el cual se basa en el Criterio de Información de Akaike (AIC), para seleccionar el mejor modelo, penalizando por el número de variables, sin perder error de ajuste.

Como se mencionó, se decidió tomar a partir del primer 80 % de la muestra, planteando en cada uno de los casos la suposición de que se tenía información hasta ese momento. El primer caso es teniendo información hasta el 3T del 2017, el segundo caso teniendo información hasta el 4T del 2017 y así sucesivamente hasta el 4T de 2019. En cada caso, se ajustó un modelo de regresión lineal múltiple en el cual todas las variables explican al PIBT, de ahí se realiza un proceso a pasos (Step) basado en el AIC, y dejamos que el algoritmo agregue o quite las variables que no aportan información adicional. A continuación, se muestra en la Tabla 3.2 el resumen de los resultados del proceso anterior, se coloca una \checkmark cuando la variable fue seleccionada para el modelo y un X cuando no se consideró.

Periodo	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	R^2_{adj}
2017-3	X	X	\checkmark	\checkmark	\checkmark	\checkmark	X	\checkmark	\checkmark	\checkmark	0.989
2017-4	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.987
2018-1	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.987
2018-2	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.988
2018-3	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.989
2018-4	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.990
2019-1	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.989
2019-2	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.989
2019-3	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.989
2019-4	X	X	\checkmark	X	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	0.990

Tabla 3.2: Resumen de mejor modelo seleccionado tomando información hasta el trimestre T (primera columna).

Como se puede observar en la Tabla 3.2, las variables X_1 , X_2 y X_4 que se refieren al Monto y Número de Tarjetas de Crédito y al Número de Tarjetas de Débito fueron las variables que

casi todas las veces fueron excluidas de los modelos. Se asume que dichas variables no aportan información adicional por lo que se excluirán del modelo propuesto.

Estacionariedad y cointegración

La Figura 3.5 muestra que las variables escogidas y el PIBT tienen una correlación bastante alta; que realmente, era de esperarse puesto que todas las variables están directa o indirectamente reflejando el crecimiento económico del país. Por ello, no podemos aún asumir que X puede predecir a Y , puesto que las relaciones pueden ser espurias. Lo que se revisará a continuación es si las variables entre sí, son relación $CI(1, 1)$, es decir, están cointegradas. Siguiendo la metodología de Johansen que se abordó en el primer capítulo, se debe revisar primero que todas las variables sean $I(1)$, es decir, que tengan una raíz unitaria. Debido a la forma de cada una de las series (clara tendencia positiva), es claro notar que las series deben de ser diferenciadas para obtener estacionariedad, sin embargo, se realizaron pruebas Dickey-Fuller aumentada (ADF test) a cada serie, para contrastar la hipótesis si dicha serie tiene una raíz unitaria o no. En las Tablas 4.1 a 4.8 del Anexo se pueden observar los detalles de este proceso. Se concluyó que al analizar cada una de las series no se rechaza la hipótesis nula, es decir, presencia de raíz unitaria, implicando $I(1)$.

Continuando con la metodología de Johansen, se procede a realizar las pruebas de máximo valor propio, la hipótesis nula: $rango(\Pi) \neq 0$ y la prueba de la traza, la cual iterativamente prueba para $r \in \{1, \dots, (p - 1)\}$, con p el número de series de tiempo a contrastar, la hipótesis $rango(\Pi) \leq r$, donde Π , es la matriz de tamaño $T \times p$ del modelo $VAR(k)$ que se abordó en la sección de Marco Teórico.

A continuación, se muestra en la Tabla 3.3, el resumen de cada una de las pruebas que realiza el proceso de Johansen, recordando que en este caso $p = 8$, dado que se cuenta con 7 covariables y 1 variable dependiente.

H_0	test	10 %	5 %	1 %
$r \leq 7$	7.19	7.52	9.24	12.97
$r \leq 6$	14.51	13.75	15.67	20.2
$r \leq 5$	18.93	19.77	22	26.81
$r \leq 4$	38.58	25.56	28.14	33.24
$r \leq 3$	53.21	31.66	34.4	39.79
$r \leq 2$	79.79	37.45	40.3	46.82
$r \leq 1$	101.36	43.25	46.45	51.91
$r=0$	127.6	48.91	52	57.95

Tabla 3.3: Resumen de las pruebas de hipótesis del proceso de Johansen. En la cual, la hipótesis nula $H_0 : rango(\Pi) \leq r$. La hipótesis se rechaza si el valor **test** es mayor que el valor de significancia dado.

Como se puede observar en la Tabla 3.3, para $0 \leq r \leq 4$ (leyendo de abajo hacia arriba), la prueba de Johansen considera estadísticamente significativo rechazar la hipótesis $rango(\Pi) \leq r$, es decir, podríamos pensar que r fuese entre 4 y 5. Algo importante que recordar de la prueba de Johansen, es que la prueba nos dice cuál es el rango de la matriz Π , más no nos dice que variables están cointegradas; por lo que se esperaría que existiese por lo menos una combinación de 4 covariables y el PIBT, que logren aceptarse la hipótesis de que dicha submatriz es de rango completo, es decir, que estén cointegradas.

Búsqueda de variables cointegradas con el PIBT

Ya que sabemos con la prueba de Johansen que el rango de Π es a lo más 5, decidimos de nuestras 7 covariables, realizar todas las combinaciones ($7C4 = 35$) de 4 covariables y el PIBT, buscando por lo menos una combinación que logre significativamente rechazar la hipótesis de Johansen ($rango(\Pi) \leq 4$). Añadido a esta prueba, a cada combinación se le ajusta un modelo de regresión múltiple para así a sus residuales contrastar las pruebas estructurales y distribucionales, es decir, que los errores tienen como distribución un ruido blanco gaussiano de media nula.

Cabe destacar las pruebas que se están tomando, es decir:

1. Breusch-Pagan de varianza constante
2. Anderson-Darling de normalidad
3. Durbin-Watson de autocorrelación.

De los 35 modelos, nos quedamos solamente con los modelos (3) en los que se rechazara la hipótesis $\text{rango}(\Pi) \leq 4$ como también los que no se rechazara el contraste a los residuales de la hipótesis de Breush-Pagan, la cual, nos traería más problemas al no ser estacionario. En la Tabla 3.4 se recupera el resumen de los 3 modelos que cumplen las anteriores condiciones; se decidió tomar el “Modelo 2”, ya que, aunque los 3 rechazan la hipótesis de Johansen a un nivel de significancia del 5% (su respectivo “jo_test” es mayor que su “jo test 5%”), el modelo acepta los supuestos estructurales y distribucionales anteriormente mencionados.

De los 3 modelos, se muestra en la Figura 3.6 el “Modelo 2”, el cual, está conformado por las siguientes 4 variables: Monto y Volumen SPEI entre \$8,000 y \$300,000 y Monto y Volumen SPEI menores a \$8,000. Dicha Figura también muestra, el análisis de residuales lo cumple de forma adecuada a un nivel de significancia del 5%, a saber: las pruebas Breusch- Pagan, Anderson-Darling y Durbin-Watson no se rechazan. El coeficiente R^2 es alto, y $\hat{\sigma}^2$ es el más pequeño de los 3. Se muestra en las Figuras 4.2 y 4.3 del Anexo los gráficos análogos de los modelos que faltaron y se descartaron porque no se aceptaba la prueba Durbin-Watson.

	Modelo 1	Modelo 2	Modelo 3
jo_test	10.0826	9.83	12.19
jo test 10%	7.5200	7.52	7.52
jo test 5%	9.2400	9.24	9.24
jo test 1%	12.9700	12.97	12.97
ncv_p	0.1175	0.24	0.07
ad.test_p	0.3942	0.81	0.11
dw_test	0.0106	0.06	0.00
$\hat{\sigma}^2$	0.0189	0.02	0.03
R^2	0.9833	0.98	0.96
V_1	nosa.real.SP8KY300K	nosa.real.SP8KY300K	nosa.cards.Debito.M
V_2	nosa.vol.SP8KY300KV	nosa.real.SPME8K	nosa.real.SPME8K
V_3	nosa.vol.SPMA300KV	nosa.vol.SP8KY300KV	nosa.vol.SP8KY300KV
V_4	nosa.vol.SPME8KV	nosa.vol.SPME8KV	nosa.vol.SPMA300KV

Tabla 3.4: Resumen de los 3 modelos de 4 covariables y el PIBT que cumplen con la condición de rechazar a un nivel de significancia del 5% la hipótesis de Johansen $\text{rango}(\Pi) \leq 4$.

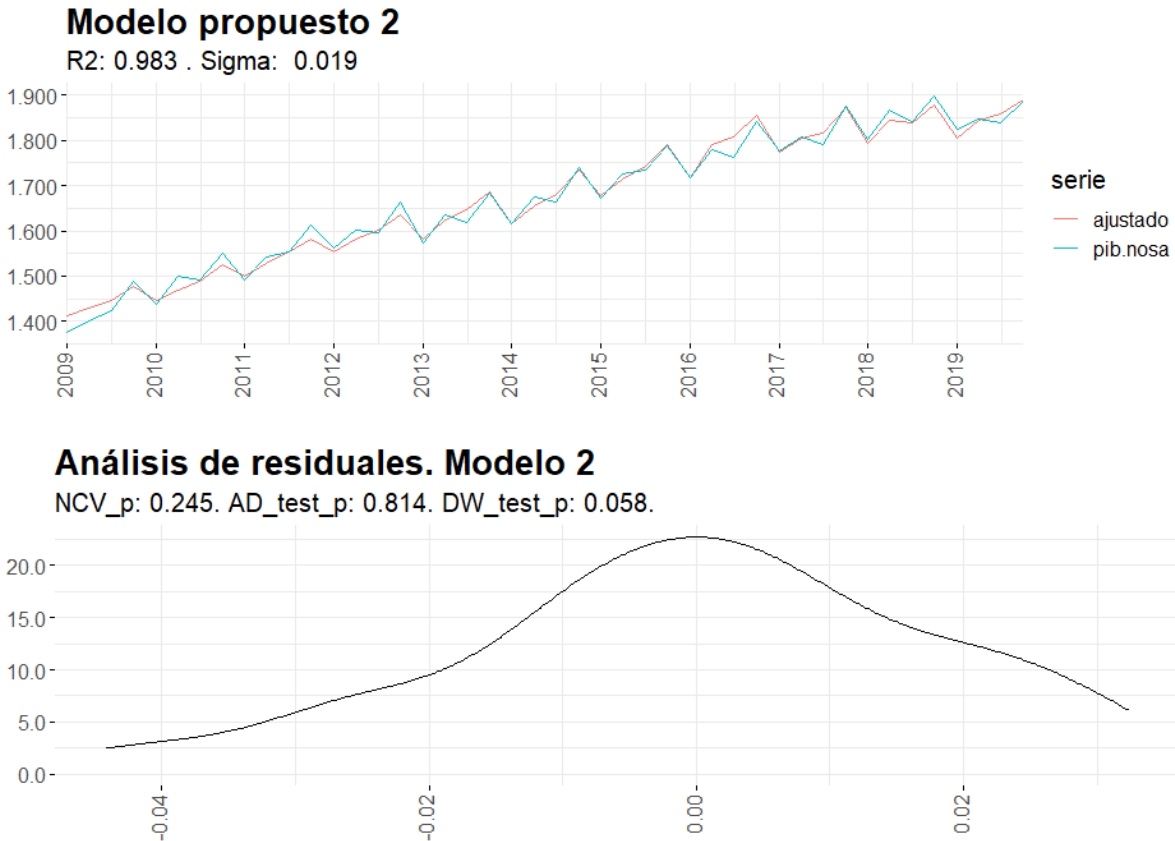


Figura 3.6: Modelo seleccionado de 4 covariables que expliquen el PIBT.

3.2. Modelado

En el primer capítulo se describió la forma en la cual se ajusta un modelo de regresión desde el enfoque bayesiano. Este modelo nos brinda diversos beneficios, ya que, al darle una densidad a los parámetros, podemos ser más flexibles y tener la posibilidad de si es necesario, dar ideas iniciales sobre los mismos. Además, no necesariamente damos un valor puntual para el parámetro, si no, sabiendo que no podemos muestrear infinitamente, nos da la posibilidad de dar densidades posteriores, intervalos de credibilidad, etc. Con las variables escogidas en la sección anterior (número y monto de las operaciones SPEI TaT menores a \$8,000 y número y monto de las operaciones SPEI TaT entre \$8,000 y \$300,000), se ajustará un modelo de regresión bayesiano para describir la relación lineal que existe con el PIBT. Sabemos previamente que dichas variables están

cointegradas con el PIBT, es decir, existe una combinación lineal de las variables la cual resulta ser estacionaria (en este caso, los residuales); esto nos da pie a que la relación que obtendremos no es espuria y tiene una fuerza predictiva mayor a que si los residuales no lo fuesen. Aunado a lo anterior, sabemos que, desde una perspectiva frecuentista, los residuales también pasan los supuestos estructurales y distribucionales, es decir, no autocorrelación entre los errores, varianza homocedástica y distribución normal.

El modelo que se planteó fue de la forma:

$$Y = \mathbf{X}\underline{\beta} + \underline{\varepsilon},$$

donde $\underline{\varepsilon} \sim N_n(\underline{\varepsilon}|\underline{0}, \tau\mathbf{I}_n)$. Al suponer que la estructura de correlación sea de la forma $\tau\mathbf{I}_n$, implica que la correlación es 0 entre las variables, logrando que las ε_i sean independientes y logrando que $Y \sim N_n(Y|\mathbf{X}\underline{\beta}, \tau\mathbf{I}_n)$. Suponemos como distribuciones iniciales para los parámetros $\underline{\beta} \sim N_p(\underline{\beta}|\mu_0, \tau\mathbf{P}_0)$ y para $\tau \sim Gamma(\tau|\alpha_0, \delta_0)$, en el primer capítulo también se abordó que dicha estructura inicial es conjugada, logrando que las distribuciones posteriores sean de la misma familia, es decir, Normal-Gamma. Dado que no se cuenta con información inicial para los hiperparámetros, se decidió utilizar la distribución inicial de Jeffreys, la cual, nos indica que los parámetros de las distribuciones finales de $\underline{\beta}$ y τ son $\mu_1 = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^TY) = \hat{\underline{\beta}}_{MV}$ y $\mathbf{P}_1 = \mathbf{X}^T\mathbf{X}$ para $\underline{\beta}$ y $\alpha_1 = n/2$ y $\delta_1 = 1/2(Y - \mathbf{X}\mu_1)^T(Y - \mathbf{X}\mu_1)$ para τ .

3.2.1. Estimación de distribuciones finales

Se construyó la matriz de diseño $\mathbf{X} = (\underline{1}_n, X_1, \dots, X_k)_{n \times p}$, la cual consiste de las 4 covariables y el vector $\underline{1}_n$ (para obtener β_0 o intercepto) y donde $p = k + 1$ y k en este caso es 4. Después se calcularon los parámetros de las distribuciones finales dada la muestra. A continuación, en la Figura 3.7, se observan las densidades posteriores de τ y σ^2 ; se marca de color rojo la media posterior y de color azul la moda posterior. Algo que es interesante recalcar, es que las distribuciones Gamma y GammaInversa son distribuciones sesgadas por lo que la media es mayor a la moda en ambos casos.

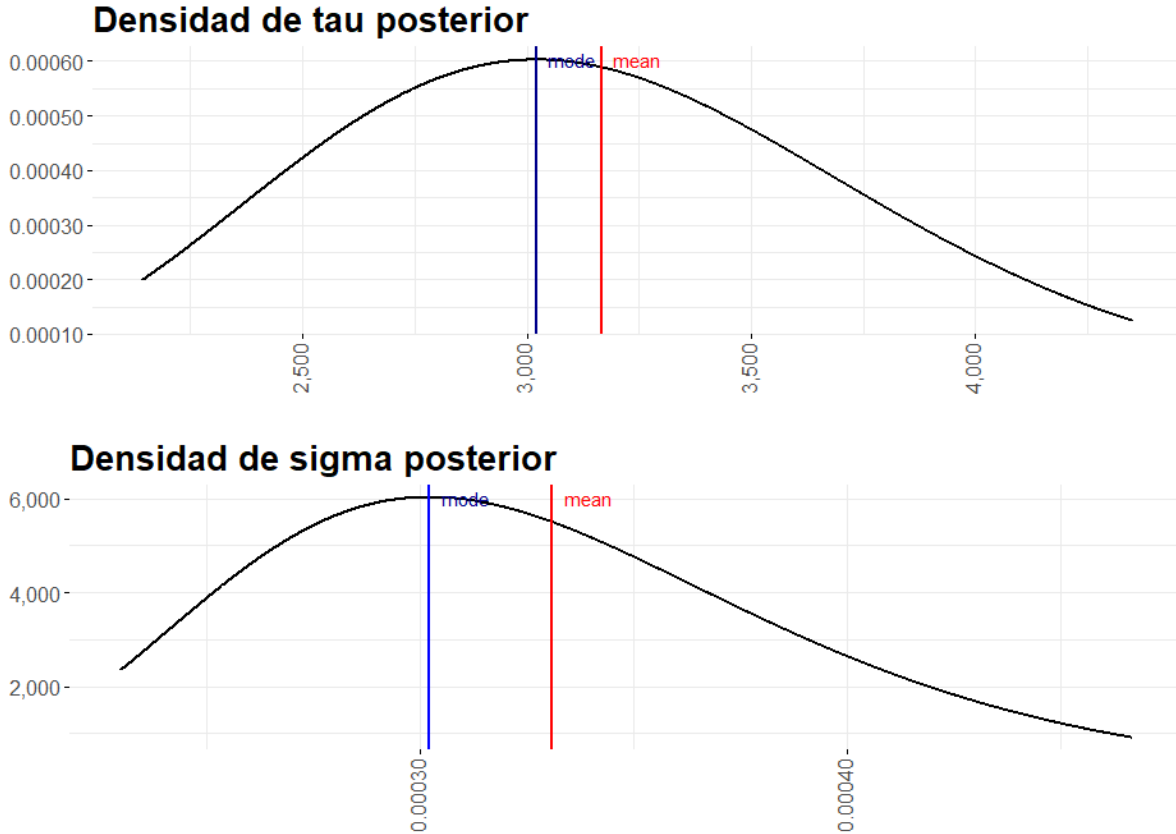


Figura 3.7: Densidad posterior de $\tau \sim \text{Gamma}(\alpha_1, \delta_1)$ y de $\sigma^2 \sim \text{GammaInversa}(\alpha_1, \delta_1)$ en el modelo de regresión bayesiano seleccionado.

Respecto al vector $\underline{\beta}$, ya que se utilizó la distribución inicial de Jeffreys y la función de pérdida cuadrática, el estimador bayesiano del modelo coincide con el estimador máximo verosímil de la idea frecuentista, sin embargo, al obtener toda la distribución T multivariada, podemos encontrar regiones de credibilidad fácilmente. En la siguiente Tabla 3.5, se observan los coeficientes estimados para cada una de las covariables. Recordando que el i -ésimo coeficiente representa la variación de la variable Y si la i -ésima covariable varía en una unidad.

Variable	$L_{0.975}$	$L_{0.95}$	$L_{0.90}$	$\hat{\beta}_i y$	$U_{0.90}$	$U_{0.95}$	$U_{0.975}$
Intercept	1.16565	1.17209	1.17917	1.21462	1.25008	1.25715	1.26359
X_1	0.00041	0.00044	0.00048	0.00065	0.00083	0.00086	0.00090
X_2	-0.00117	-0.00048	0.00027	0.00406	0.00785	0.00861	0.00930
X_3	-0.04239	-0.03992	-0.03720	-0.02358	-0.00996	-0.00724	-0.00476
X_4	-0.01021	-0.00947	-0.00867	-0.00462	-0.00057	0.00023	0.00097

Tabla 3.5: Coeficientes $\hat{\beta}_i$ estimados y sus intervalos al 90%, 95% y 97.5% de credibilidad para el modelo de regresión bayesiana $PIBT = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$.

También se muestra a continuación en la Figura 3.8 las distribuciones T de Student para cada β_i . De color azul se muestra el intervalo de 95% credibilidad para coeficiente. Algunos puntos que podemos recalcar de los mismos son:

1. El coeficiente de mayor aportación con un valor central de 1.215 es el intercepto, lo que nos habla de que, si nuestras 4 covariables tuviesen valores cercanos a cero, el valor ajustado del PIBT sería un valor muy cercano a él.
2. Dos variables que pueden ser de preocupación, son el monto y volumen de las operaciones SPEI TaT menores a \$8,000 (*nosa.real.SPME8K* y *nosa.vol.SPME8KV*) ya que el valor 0 vive dentro del intervalo de credibilidad al 95%, sin embargo, al 90% de credibilidad, ya no pertenece el 0 a ninguna variable, esto se puede ver con detalle en la Tabla 3.5.

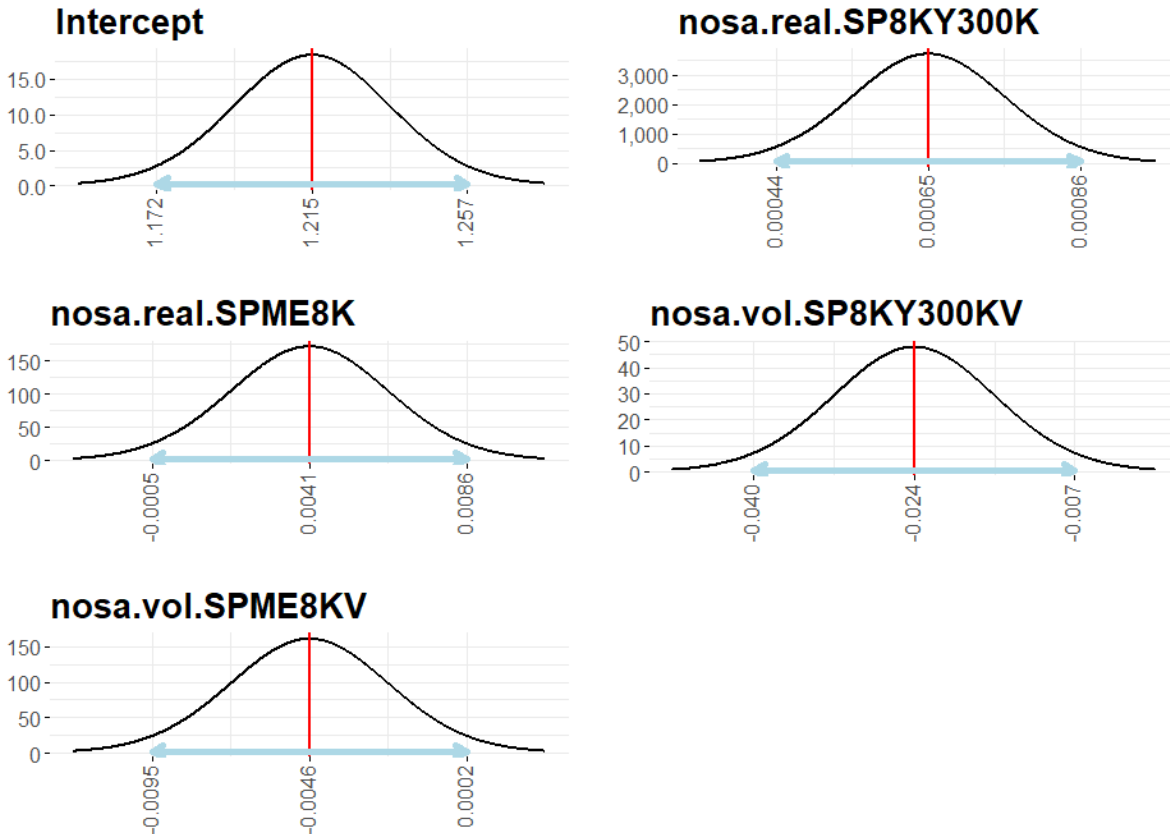


Figura 3.8: Densidades T-student posteriores de cada uno de los coeficientes del modelo de regresión. De color azul se marca el intervalo de 95 % de credibilidad. De color rojo se marca la media y moda (las cuales coinciden por simetría de la distribución).

Prueba Lindley para $\underline{\beta}$

Una prueba de hipótesis importante en las regresiones frecuentistas es la prueba F de Fischer la cual, la hipótesis nula refiere a contrastar si el vector de coeficientes, $\underline{\beta}$, sea igual a $\underline{0}$. Para ello se basa en utilizar el cociente de verosimilitud generalizado, y se logra definir una región de rechazo la cual distribuye, como su nombre lo indica, F de Fischer. En nuestro contexto bayesiano, no es tan directo poder definir dichas pruebas de hipótesis, por lo que, lo que se hace, es construir la región de credibilidad para $(\beta_1, \dots, \beta_4)^T$ utilizando la matriz \mathbf{A} definida como:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

para que $\mathbf{A}\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ y preguntarnos si el vector $\underline{0} \in \mathbb{R}^4$ pertenece a la región.

\mathbf{A} está definida de esta forma ya que nos dará un intervalo de confianza para todos los parámetros exceptuando el intercepto: nos interesa saber si existe por lo menos una covariable, $x_k, k \in \{1, \dots, 4\}$, la cual tenga un valor significativamente diferente de cero, con esto podemos argumentar estadísticamente que la relación lineal planteada, $PIBT = \mathbf{X}\underline{\beta}$, existe y tiene sentido. Utilizando la información de nuestro modelo, y siguiendo el desarrollo del primer capítulo, la hipótesis nula se rechaza si:

$$\frac{1}{4}(\mathbf{A}\mu_1)^T \left(\mathbf{A} \left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \mathbf{A}^T \right)^{-1} (\mathbf{A}\mu_1) > F_{(4, 2\alpha_1)}^{1-\alpha},$$

con todos los parámetros definidos en las secciones anteriores. En la Tabla 3.6, se muestra la comparación del estadístico F en comparación con los cuantiles $F_{(4, 2\alpha_1)}$. Se observa que el estadístico es considerablemente mayor a los cuantiles a niveles de significancia de 10 %, 5 % y 2.5 %, por lo que podemos concluir que hay evidencia que el cero no pertenece a la región de credibilidad.

F	0.1	0.05	0.025
715.138919	2.07719416	2.58366743	3.09328796

Tabla 3.6: Tabla resumen de la prueba Lindley para la hipótesis nula $H_0 : (\beta_1, \dots, \beta_4)^T = \underline{0}$. La hipótesis se rechaza a los niveles de significancia 10 %, 5 % y 2.5 % si el estadístico F es mayor que los valores críticos de la distribución F de Fischer.

3.3. Predicción

Valores ajustados (Predicción dentro de la muestra)

Con el vector $\hat{\underline{\beta}}$, se define a $\hat{Y} := \mathbf{X}\hat{\underline{\beta}}$, el vector de valores ajustados con el modelo de regresión, dicho vector nos sirve para estimar el valor de $\hat{\sigma}^2$ y por consiguiente de $\hat{\tau}$, ya que, refleja los valores

estimados de nuestra variable Y dados los datos, \mathbf{X} y el modelo. Es importante recalcar que este valor ajustado, es el valor esperado de una distribución completa, (distribución predictiva) lo cual nos da pie a poder dar regiones o intervalos de credibilidad fácilmente. Recordemos que el modelo planteado es de la forma:

$$Y = \mathbf{X}\underline{\beta} + \varepsilon,$$

Por lo que supóngase que Y no es conocido, digamos W sea el vector de valores a estimar en el modelo dados los datos, es decir, lo que nos interesa saber, es la distribución final de $W|Y$, la cual, en el primer capítulo se hizo el desarrollo llegando a la conclusión de que dicha distribución tiene forma de T-Student multivariada de la forma:

$$p(W|Y) \propto T_n \left(W \middle| 2\alpha_1, \mathbf{X}\mu_1, \frac{\alpha_1}{\delta_1} (\mathbf{I} + \mathbf{X}\mathbf{P}_1\mathbf{X}^T)^{-1} \right)$$

por lo que, individualmente, el intervalo de credibilidad al $(1 - \alpha) \%$ para w_i , $i \in \{1, \dots, n\}$, está dado por:

$$\left((\mathbf{Z}\mu_1)_i \pm t_{2\alpha_1}^{1-\alpha/2} \right) \sqrt{d_{(i,i)}},$$

donde $\mathbf{D} = \frac{\delta_1}{\alpha_1} (\mathbf{I} + \mathbf{X}\mathbf{P}_1\mathbf{X}^T)$. En la Figura 3.9 se observa de color negro el vector Y observado, de color rojo los valores ajustados, \hat{Y} , con el modelo y de color gris punteado, los intervalos de credibilidad al 95%. Cabe mencionar que a estos valores se les dice ajustados, ya que fueron predicciones de la variable dependiente dentro de la muestra, es decir, nosotros sabemos inicialmente el valor real de la variable Y y por consiguiente es común que si el modelo fue ajustado de forma satisfactoria, (un error considerablemente pequeño y que cumpla los supuestos estructurales y distribucionales) todos los valores reales pertenezcan al intervalo de credibilidad; sin embargo, al ver que 43 de 44 valores reales viven dentro de los intervalos de credibilidad, nos indica que el modelo es bueno y tiene una fuerza predictiva esperanzadora.

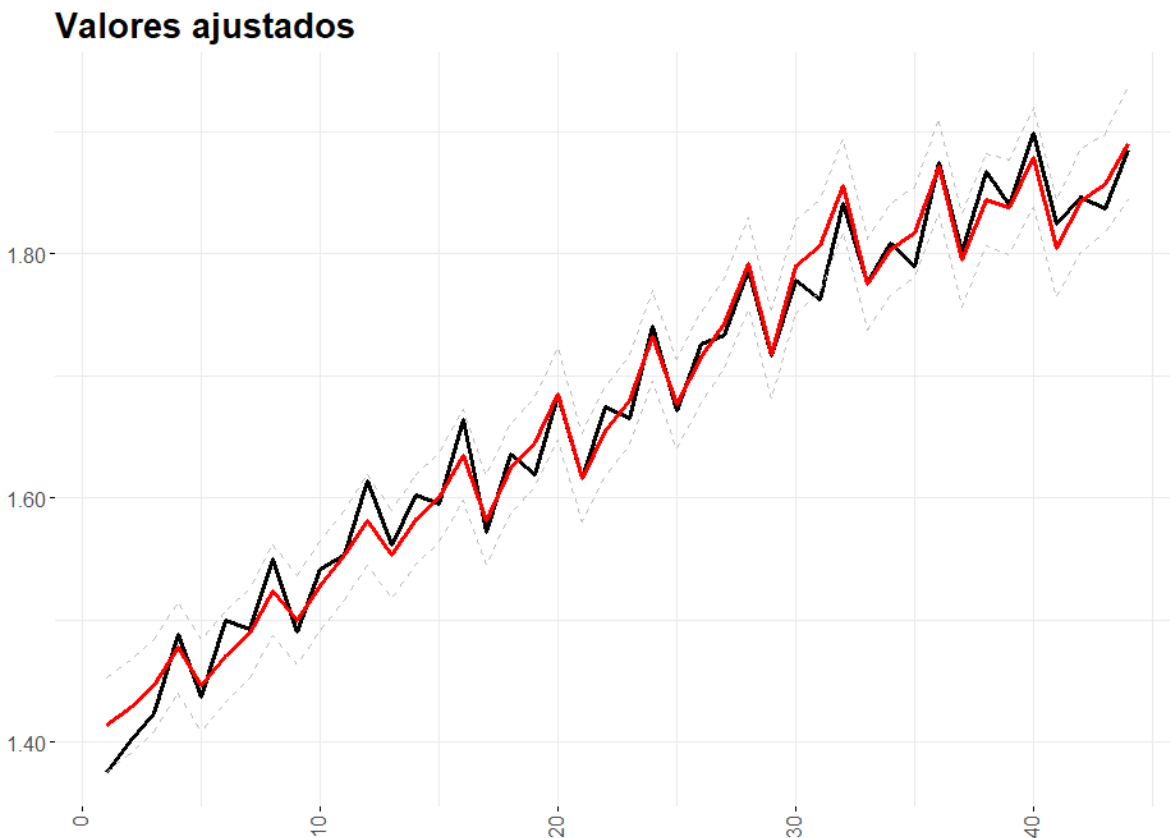


Figura 3.9: Valores ajustados, $PIBT = \hat{Y} = \mathbf{X}\hat{\beta}$, y sus intervalos de credibilidad al 95 % para el modelo de regresión desde el enfoque bayesiano.

3.3.1. Predicción fuera de la muestra

Tomando la misma idea anterior, y teniendo una observación adicional (1er Trimestre de 2020), \mathbf{Z} , de nuestras variables independientes, se procede a obtener a $\hat{W} = \mathbf{Z}\hat{\beta}$ junto con su intervalo de credibilidad al 95 %. Cabe destacar que este valor ya es una predicción (también conocido como pronóstico) fuera de la muestra con información de la variable independiente no utilizada para calibrar al modelo. Este valor es ya el que nosotros como econométristas estamos buscando: el valor del PIBT oportuno antes de que el INEGI lo publique. En la Figura 3.10 podemos observar los valores ajustados de nuestro modelo, agregando de color azul y de forma punteada, el valor que está prediciendo el modelo para el primer trimestre del 2020. Añadido a lo anterior, en la misma Figura se observa de color negro el valor real del PIBT el cual, pertenece al intervalo de

credibilidad de nuestra predicción.

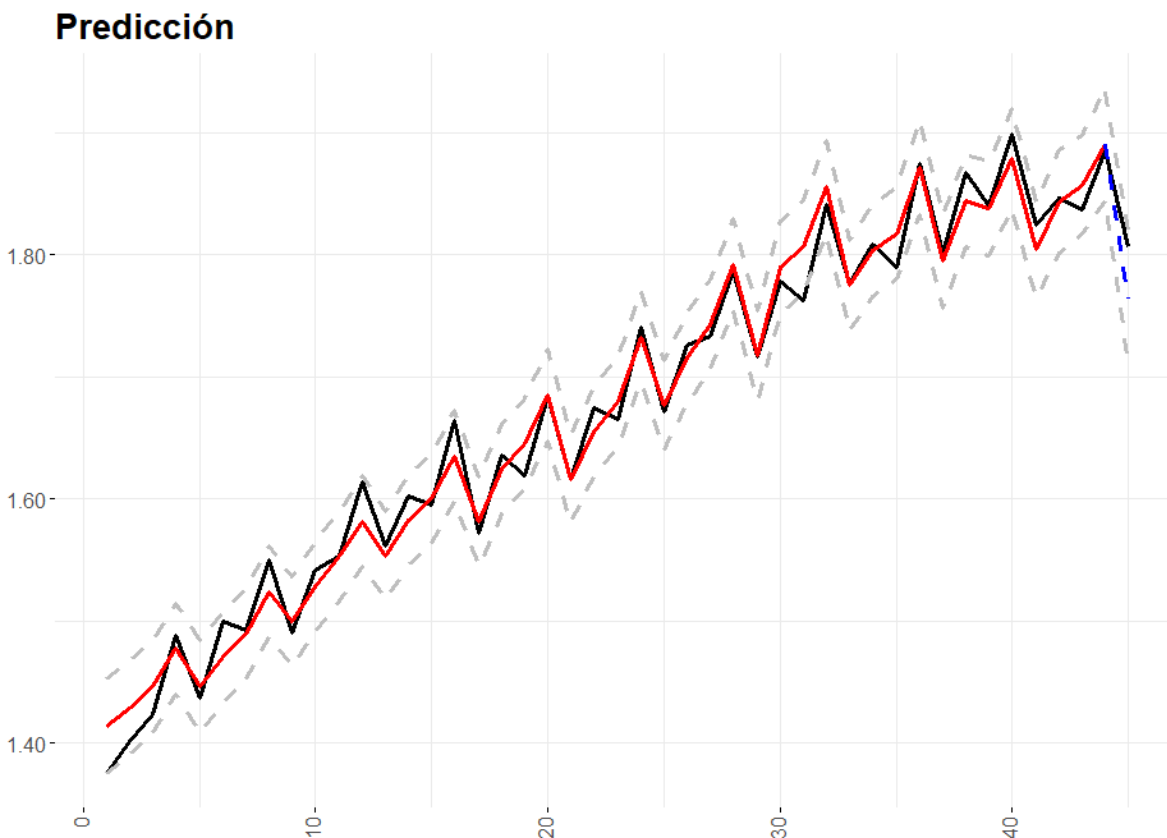


Figura 3.10: Predicción del PIBT (línea azul punteada), primer trimestre del 2020 junto con su intervalo de credibilidad al 95 % (línea punteada gris). De color rojo se muestra toda la serie de valores ajustados y de color negro la serie observada de la variable de interés.

Resultados

Con base en el proceso anterior, se decidió hacer un proceso iterativo desde el primer trimestre de 2017, hasta el último trimestre del 2021. Es decir, para cada iteración al tiempo t , se toma la muestra observada de las covariables y del PIBT hasta el tiempo $t - 1$, se calibra el modelo con esa información y se prosigue a pronosticar el PIBT al tiempo t con la información al tiempo t de las covariables. En la Tabla 3.8 se observa trimestre por trimestre el resultado del ejercicio anterior. La última columna nos marca con un \checkmark si el observado pertenece al intervalo de credibilidad de 95 % y con un X si no pertenece. Las Figuras que muestran cada una de las predicciones de forma

visual se muestran en las Figuras 4.4 a 4.8 del Anexo.

Además, al poder tener la información observada del PIBT, podemos calcular métricas de error, como por ejemplo el error cuadrático medio (MSE por si siglas en inglés), el error absoluto medio (MAE), entre otros. En la Tabla 3.7 podemos observar estos valores; cabe resaltar que el valor que se puede interpretar mejor ya que es independiente de la escala es el error porcentual absoluto medio (MAPE), el cual, nuestro ejercicio tuvo un valor de 3.47 %.

Tipo	Expresión	Error
MAE	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	0.0597
MSE	$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	0.0074
RMSE	$\text{sqrt}(MSE)$	0.0860
MAPE	$\frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} * 100 \right $	3.4742 %
sMAPE	$\frac{100}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{(y_i + \hat{y}_i)/2}$	3.5025 %

Tabla 3.7: Resumen de métricas del error de predicción para el modelo de regresión $PIBT = \mathbf{X}\beta$.

Año	T	Obs	L	Pred	U	I
2017	1	1.7758	1.6738	1.7196	1.7654	X
2017	2	1.8085	1.7326	1.7800	1.8275	✓
2017	3	1.7901	1.7590	1.8030	1.8471	✓
2017	4	1.8746	1.7972	1.8418	1.8863	✓
2018	1	1.8018	1.7416	1.7940	1.8464	✓
2018	2	1.8669	1.8103	1.8564	1.9026	✓
2018	3	1.8406	1.8157	1.8597	1.9037	✓
2018	4	1.8987	1.8560	1.9016	1.9472	✓
2019	1	1.8244	1.8143	1.8759	1.9375	✓
2019	2	1.8469	1.8483	1.9074	1.9664	X
2019	3	1.8374	1.8474	1.9035	1.9596	X
2019	4	1.8855	1.8392	1.9001	1.9611	✓
2020	1	1.8062	1.7108	1.7638	1.8167	✓
2020	2	1.5013	1.6440	1.7083	1.7727	X
2020	3	1.6805	1.3422	1.4564	1.5706	X
2020	4	1.8023	1.6090	1.6908	1.7726	X
2021	1	1.7378	1.5271	1.6123	1.6976	X
2021	2	1.7997	1.6081	1.7161	1.8241	✓
2021	3	1.7565	1.6928	1.7901	1.8874	✓
2021	4	1.8221	1.7308	1.8253	1.9197	✓

Tabla 3.8: Resumen de predicciones con el modelo de regresión bayesiano $PIBT = \mathbf{X}\beta$ desde el 1er trimestre del 2017 al 4to trimestre del 2021.

3.3.2. Reflexiones periodo confinamiento COVID-19

Retomando la Tabla 3.8, es claro que durante del periodo del segundo trimestre del 2020 al primer trimestre del 2021 se observan los errores de predicción de mayor magnitud, los cuales, se refieren al periodo de confinamiento por la pandemia a causa del virus SARS-CoV 2. Es por eso por lo que se decidió hacer un análisis más detallado a este periodo. En primera instancia calculamos los crecimientos trimestrales de las variables involucradas en el modelo. Dichos crecimientos se

definen como

$$\Delta X_T = \left(\frac{X_T - X_{T-1}}{X_{T-1}} \right) * 100 \% = \left(\frac{X_T}{X_{T-1}} - 1 \right) * 100 \%$$

y se pueden observar en la Figura 3.11. Es alentador observar que, durante el periodo del primer trimestre del 2017 al primer trimestre del 2020, el comportamiento de las 5 variables es armonioso (parece que forman letras “M”), podemos observar la cointegración que existe en ellas, cuando una variable sube, las demás suben y lo mismo pasa cuando una variable baja. El comportamiento casi caótico comienza a partir del segundo trimestre del 2020. Es pertinente realizar un análisis detallado de este periodo.

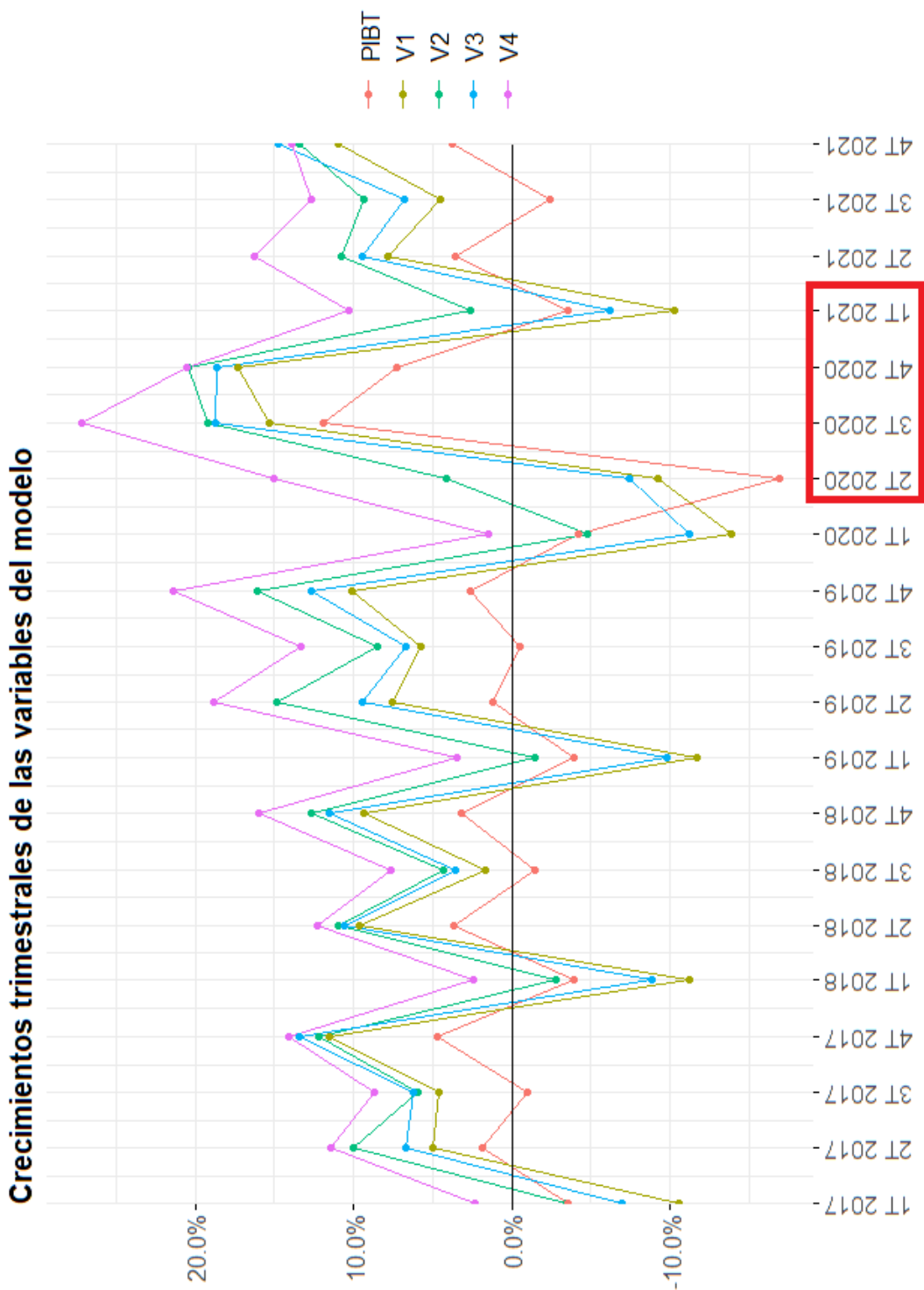


Figura 3.11: Crecimientos trimestrales de las variables del modelo de regresión $PIBT = \mathbf{X}\beta$ en el periodo del primer trimestre del 2017 al cuarto trimestre del 2021.

2T 2020. El segundo trimestre (abril a junio) del 2020, fue el periodo más duro para la población global. En México la mayoría de los negocios cerraron de forma presencial, muchos negocios cayeron en banca rota y los que pudieron continuar sus operaciones de forma no presencial o por medio de entregas a domicilio, tuvieron que buscar nuevas formas de pago. Debido a lo anterior, as 4 variables de pagos del modelo tuvieron alzas, sin embargo, el PIB tuvo una caída preocupante. Es el primer periodo en el cual el PIB no sigue el comportamiento de las otras variables: hubo una descointegración de las mismas implicando que la predicción fuera mayor de lo que fue lo observado. Un dato interesante es, que en este periodo de prueba, es el primer año en el cual el segundo trimestre tiene un crecimiento negativo.

3T 2020. En este trimestre (julio a septiembre) del 2020, muchos de los negocios y centros comerciales reabrieron sus puertas con las medidas pertinentes de higiene; esto reactivó la economía de forma sustancial y dado que las variables de pagos habían tenido un crecimiento considerable el trimestre anterior, el cambio tan abrupto del PIB en comparación con las variables de pagos no pudo ser captado por el modelo implicando que la predicción fuera mucho menor de lo que fue realmente. Un dato interesante es que, en este periodo de prueba del modelo, es el primer año el cual el tercer trimestre tiene un crecimiento positivo.

4T 2020. En este trimestre las variables de pagos estuvieron relativamente estables, sin embargo, se sigue observando esa falta de cointegración con la variable del PIB, a saber, 3 de las 4 variables de pagos tuvieron un alza, sin embargo, el PIB volvió a tener una caída a raíz del segundo confinamiento debido a la segunda ola de contagios. Esto se puede observar en la Figura 4.1.

1T 2021. En este último trimestre de este periodo de confinamiento, se observa que las variables de pagos y el PIB tienen un comportamiento similar, sin embargo, la variación tan abrupta de las variables de pagos, debido a que muchas transacciones en el país ya volvieron a ser en efectivo, se esperaba que el PIB fuera menor de lo observado.

A partir del segundo trimestre del 2021, las 5 variables involucradas en el modelo recuperaron casi en su totalidad la estructura de cointegración observada antes del periodo de confinamiento. Los pronósticos del segundo trimestre al cuarto trimestre del 2021 fueron mejores y el observado perteneció al intervalo de credibilidad de 95 %.

Para concluir y omitiendo el evento desafortunado que vivió la humanidad en 2020, existen pruebas estadísticas de que las variables de pagos pueden ser de utilidad para predicciones de variables macroeconómicas, en este caso el PIB. Algo interesante que rescatar de este ejercicio, es que aunado a que nuestra muestra de prueba era muy pequeña, se pudieron realizar ejercicios de predicción con un error menor al 5%. Teniendo un buen pronóstico del PIB, las instituciones financieras pueden tener una toma de decisiones informada y anticipada.

Capítulo 4

Conclusiones

En este documento se abordó el uso de la información de los sistemas de pagos para realizar pronósticos del Producto Interno Bruto en México. El Banco de México ha realizado esfuerzos continuos para que el público en general tenga la información de los sistemas de pagos de forma puntual, por lo que, esta información nos es de utilidad para generar modelos de predicción de variables macroeconómicas, en este caso, el PIBT. Se demostró que existe un subconjunto de variables de pagos que están cointegradas con nuestra variable de interés, por lo que los pronósticos son fidedignos y altamente oportunos (más de 5 semanas). Cabe destacar que se sabe de antemano que este ejercicio deja de lado muchos factores de análisis de series de tiempo, sin embargo, lo que se quiere mostrar en este documento es ejemplificarle al lector que la información de pagos electrónicos en México tiene un comportamiento similar a la actividad económica nacional. Sudhakar Raju et. al. [17] en 2019 realizaron en India un ejercicio análogo y concluyeron similarmente. Esto nos hace pensar que la información de pagos electrónicos es útil para poder explicar efectos de coyuntura económica.

Hablando en profundidad del modelo propuesto en este documento, respecto a los coeficientes que acompañan a las variables exógenas, podemos observar que los pesos relacionados con el monto operado a través del SPEI TaT son positivas, lo que nos hace afirmar que, si el monto operado aumenta, está hablando que a la economía le está yendo en promedio de forma favorable. Sin embargo, no podemos ignorar el suceso que marcó la economía mundial en 2020: la pandemia por el virus SARS-CoV-2. Este evento obligó que una parte considerable de la actividad económica nacional se detuviese y el resto de los negocios tuvieran que optar por nuevos procesos logísticos y claramente buscar nuevos medios de pago. Esto logró que durante este periodo los pagos electrónicos

tuvieran gran relevancia y crecimiento contradiciendo el comportamiento de la economía. Hubo una “desintegración” de las series: el modelo pronosticó que la caída de la economía no sería tan grave y los intervalos de credibilidad para ese periodo no fueron lo suficientemente amplios para que el observado perteneciese a ellos. Algo que se propone y se invita al lector, es diseñar estrategias para prevenirnos de estos eventos atípicos. Ya que la información de pagos es oportuna y diaria, se pueden hacer análisis profundos intertrimestrales y ayudar al modelo a captar estos focos rojos. Para lo anterior, nos es de utilidad que el modelo se definió desde una perspectiva bayesiana dando espacio a dar ideas iniciales de análisis econométricos previos. Cabe resaltar que se busca que el lector note la utilidad de la perspectiva bayesiana y lo agregue a sus herramientas cotidianas, para así, lograr análisis de mayor robustez para utilidad y beneficio de la sociedad.

Anexo

4.0.1. Casos Confirmados COVID 19



Figura 4.1: Fuente: Periódico El Economista.[7]

4.0.2. ADF

Se muestra cada una de las tablas relacionadas a la prueba aumentada de Dickey-Fuller aplicada a cada una de las variables involucradas en el análisis. Se realiza de 0 a 3 órdenes y en sus tres variantes principales: sin media y sin tendencia, con media y sin tendencia y con media y tendencia.

variable	tipo	lag	ADF	p_value
pib.nosa	no_drift_no_trend	0	1.30276	0.94841
pib.nosa	no_drift_no_trend	1	3.66888	0.99
pib.nosa	no_drift_no_trend	2	2.90376	0.99
pib.nosa	no_drift_no_trend	3	6.65186	0.99
pib.nosa	yes_drift_no_trend	0	-1.6735	0.45159
pib.nosa	yes_drift_no_trend	1	-1.86809	0.37863
pib.nosa	yes_drift_no_trend	2	-1.70067	0.44141
pib.nosa	yes_drift_no_trend	3	-3.73396	0.01
pib.nosa	yes_drift_yes_trend	0	-7.96804	0.01
pib.nosa	yes_drift_yes_trend	1	-3.25507	0.0912
pib.nosa	yes_drift_yes_trend	2	-3.76306	0.03129
pib.nosa	yes_drift_yes_trend	3	0.19789	0.99

Tabla 4.1: Prueba aumentada de Dickey-Fuller aplicada a la variable pib.nosa.

variable	tipo	lag	ADF	p_value
nosa.cards.Debito.M	no_drift_no_trend	0	2.07446	0.98912
nosa.cards.Debito.M	no_drift_no_trend	1	3.41852	0.99
nosa.cards.Debito.M	no_drift_no_trend	2	3.98688	0.99
nosa.cards.Debito.M	no_drift_no_trend	3	24.61422	0.99
nosa.cards.Debito.M	yes_drift_no_trend	0	0.25414	0.97172
nosa.cards.Debito.M	yes_drift_no_trend	1	0.95838	0.99
nosa.cards.Debito.M	yes_drift_no_trend	2	1.35489	0.99
nosa.cards.Debito.M	yes_drift_no_trend	3	11.2424	0.99
nosa.cards.Debito.M	yes_drift_yes_trend	0	-3.62293	0.04244
nosa.cards.Debito.M	yes_drift_yes_trend	1	-2.09314	0.52504
nosa.cards.Debito.M	yes_drift_yes_trend	2	-1.53496	0.75325
nosa.cards.Debito.M	yes_drift_yes_trend	3	-0.28701	0.98684

Tabla 4.2: Prueba aumentada de Dickey-Fuller aplicada a la variable nosa.cards.Debito.M.

variable	tipo	lag	ADF	p_value
nosa.real.SP8KY300K	no_drift_no_trend	0	2.14012	0.99
nosa.real.SP8KY300K	no_drift_no_trend	1	3.49512	0.99
nosa.real.SP8KY300K	no_drift_no_trend	2	3.74891	0.99
nosa.real.SP8KY300K	no_drift_no_trend	3	12.70497	0.99
nosa.real.SP8KY300K	yes_drift_no_trend	0	-0.09132	0.94155
nosa.real.SP8KY300K	yes_drift_no_trend	1	0.4526	0.9812
nosa.real.SP8KY300K	yes_drift_no_trend	2	0.62933	0.98822
nosa.real.SP8KY300K	yes_drift_no_trend	3	3.17619	0.99
nosa.real.SP8KY300K	yes_drift_yes_trend	0	-6.31468	0.01
nosa.real.SP8KY300K	yes_drift_yes_trend	1	-4.54606	0.01
nosa.real.SP8KY300K	yes_drift_yes_trend	2	-4.56014	0.01
nosa.real.SP8KY300K	yes_drift_yes_trend	3	-1.87816	0.61293

Tabla 4.3: Prueba aumentada de Dickey-Fuller aplicada a la variable nosa.real.SP8KY300K.

variable	tipo	lag	ADF	p_value
nosa.real.SPMA300K	no_drift_no_trend	0	0.75673	0.85532
nosa.real.SPMA300K	no_drift_no_trend	1	0.81144	0.87098
nosa.real.SPMA300K	no_drift_no_trend	2	0.70432	0.84031
nosa.real.SPMA300K	no_drift_no_trend	3	0.76747	0.85839
nosa.real.SPMA300K	yes_drift_no_trend	0	-2.47244	0.15203
nosa.real.SPMA300K	yes_drift_no_trend	1	-2.79476	0.07265
nosa.real.SPMA300K	yes_drift_no_trend	2	-2.88669	0.05896
nosa.real.SPMA300K	yes_drift_no_trend	3	-2.30598	0.21444
nosa.real.SPMA300K	yes_drift_yes_trend	0	-2.68972	0.29458
nosa.real.SPMA300K	yes_drift_yes_trend	1	-2.9491	0.19505
nosa.real.SPMA300K	yes_drift_yes_trend	2	-3.21107	0.09785
nosa.real.SPMA300K	yes_drift_yes_trend	3	-2.47306	0.37772

Tabla 4.4: Prueba aumentada de Dickey-Fuller aplicada a la variable nosa.real.SPMA300K.

variable	tipo	lag	ADF	p_value
nosa.real.SPME8K	no_drift_no_trend	0	7.25593	0.99
nosa.real.SPME8K	no_drift_no_trend	1	7.99952	0.99
nosa.real.SPME8K	no_drift_no_trend	2	3.30285	0.99
nosa.real.SPME8K	no_drift_no_trend	3	5.58161	0.99
nosa.real.SPME8K	yes_drift_no_trend	0	4.89295	0.99
nosa.real.SPME8K	yes_drift_no_trend	1	6.80011	0.99
nosa.real.SPME8K	yes_drift_no_trend	2	3.66183	0.99
nosa.real.SPME8K	yes_drift_no_trend	3	6.36641	0.99
nosa.real.SPME8K	yes_drift_yes_trend	0	2.77305	0.99
nosa.real.SPME8K	yes_drift_yes_trend	1	5.56726	0.99
nosa.real.SPME8K	yes_drift_yes_trend	2	3.97326	0.99
nosa.real.SPME8K	yes_drift_yes_trend	3	11.01089	0.99

Tabla 4.5: Prueba aumentada de Dickey-Fuller aplicada a la variable nosa.real.SPME8K.

variable	tipo	lag	ADF	p_value
nosa.vol.SP8KY300KV	no_drift_no_trend	0	3.63927	0.99
nosa.vol.SP8KY300KV	no_drift_no_trend	1	5.36946	0.99
nosa.vol.SP8KY300KV	no_drift_no_trend	2	4.87003	0.99
nosa.vol.SP8KY300KV	no_drift_no_trend	3	19.32885	0.99
nosa.vol.SP8KY300KV	yes_drift_no_trend	0	1.36577	0.99
nosa.vol.SP8KY300KV	yes_drift_no_trend	1	2.44783	0.99
nosa.vol.SP8KY300KV	yes_drift_no_trend	2	2.70207	0.99
nosa.vol.SP8KY300KV	yes_drift_no_trend	3	16.45718	0.99
nosa.vol.SP8KY300KV	yes_drift_yes_trend	0	-1.53463	0.75338
nosa.vol.SP8KY300KV	yes_drift_yes_trend	1	-0.188	0.99
nosa.vol.SP8KY300KV	yes_drift_yes_trend	2	0.08506	0.99
nosa.vol.SP8KY300KV	yes_drift_yes_trend	3	3.30069	0.99

Tabla 4.6: Prueba Dickey-Fuller aplicada a la variable nosa.vol.SP8KY300KV.

variable	tipo	lag	ADF	p_value
nosa.vol.SPMA300KV	no_drift_no_trend	0	2.3951	0.99
nosa.vol.SPMA300KV	no_drift_no_trend	1	3.97705	0.99
nosa.vol.SPMA300KV	no_drift_no_trend	2	4.14287	0.99
nosa.vol.SPMA300KV	no_drift_no_trend	3	14.12186	0.99
nosa.vol.SPMA300KV	yes_drift_no_trend	0	0.36176	0.97758
nosa.vol.SPMA300KV	yes_drift_no_trend	1	1.13382	0.99
nosa.vol.SPMA300KV	yes_drift_no_trend	2	1.38306	0.99
nosa.vol.SPMA300KV	yes_drift_no_trend	3	5.94609	0.99
nosa.vol.SPMA300KV	yes_drift_yes_trend	0	-3.6222	0.0425
nosa.vol.SPMA300KV	yes_drift_yes_trend	1	-1.99393	0.56561
nosa.vol.SPMA300KV	yes_drift_yes_trend	2	-1.63462	0.7125
nosa.vol.SPMA300KV	yes_drift_yes_trend	3	-1.039	0.92122

Tabla 4.7: Prueba Dickey-Fuller aplicada a la variable nosa.vol.SPMA300KV.

variable	tipo	lag	ADF	p_value
nosa.vol.SPME8KV	no_drift_no_trend	0	11.85668	0.99
nosa.vol.SPME8KV	no_drift_no_trend	1	6.3657	0.99
nosa.vol.SPME8KV	no_drift_no_trend	2	0.69957	0.83896
nosa.vol.SPME8KV	no_drift_no_trend	3	0.99091	0.9097
nosa.vol.SPME8KV	yes_drift_no_trend	0	10.10806	0.99
nosa.vol.SPME8KV	yes_drift_no_trend	1	7.39201	0.99
nosa.vol.SPME8KV	yes_drift_no_trend	2	1.73714	0.99
nosa.vol.SPME8KV	yes_drift_no_trend	3	2.21074	0.99
nosa.vol.SPME8KV	yes_drift_yes_trend	0	7.56553	0.99
nosa.vol.SPME8KV	yes_drift_yes_trend	1	9.36214	0.99
nosa.vol.SPME8KV	yes_drift_yes_trend	2	3.58234	0.99
nosa.vol.SPME8KV	yes_drift_yes_trend	3	5.99165	0.99

Tabla 4.8: Prueba aumentada de Dickey-Fuller aplicada a la variable nosa.vol.SPME8KV.

4.0.3. Modelos escogidos

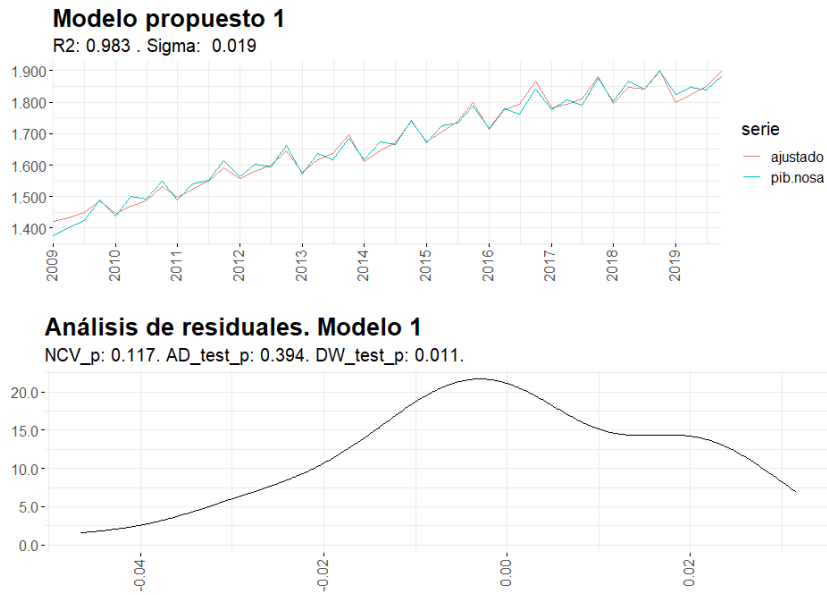


Figura 4.2: Modelo 1 de 3 el cual se acepta la prueba de Johansen de cointegración de las series.

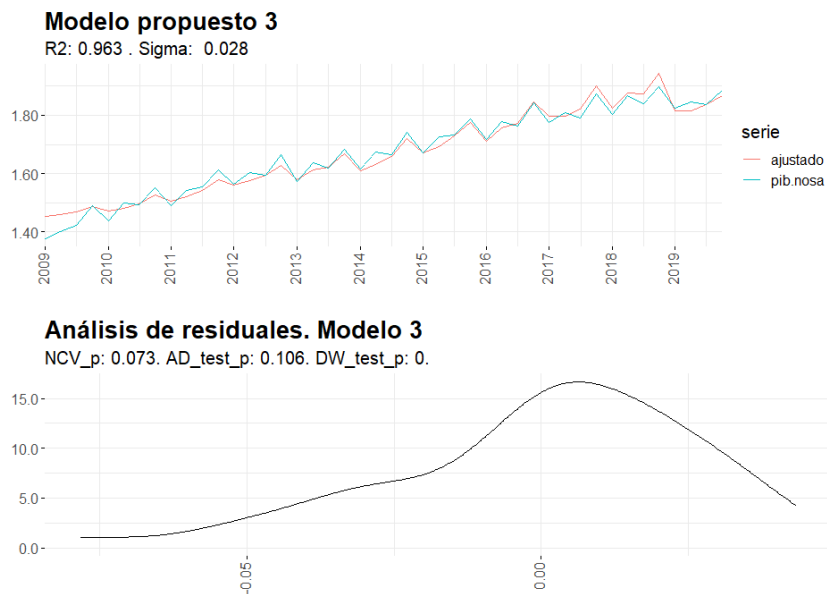


Figura 4.3: Modelo 3 de 3 el cual se acepta la prueba de Johansen de cointegración de las series.

4.0.4. Gráficos de predicciones

2017

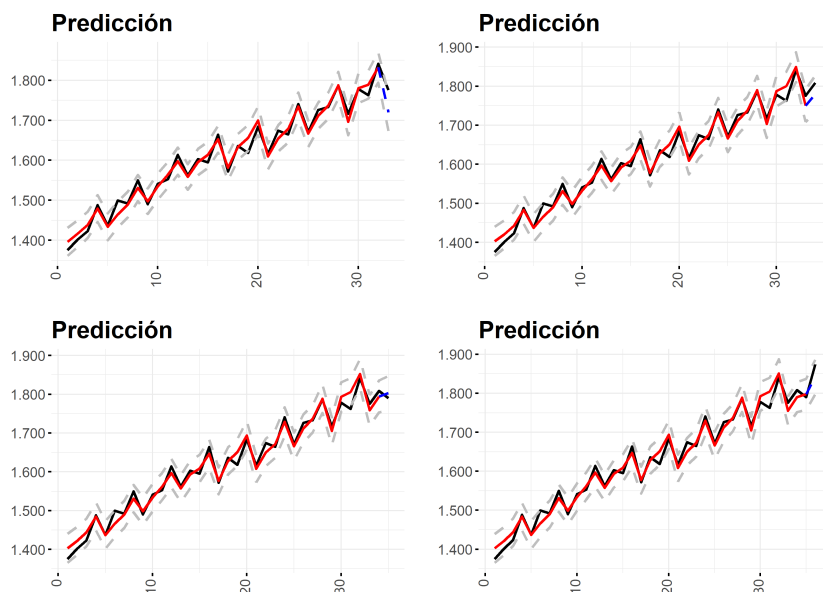


Figura 4.4: Predicciones 2017.

2018

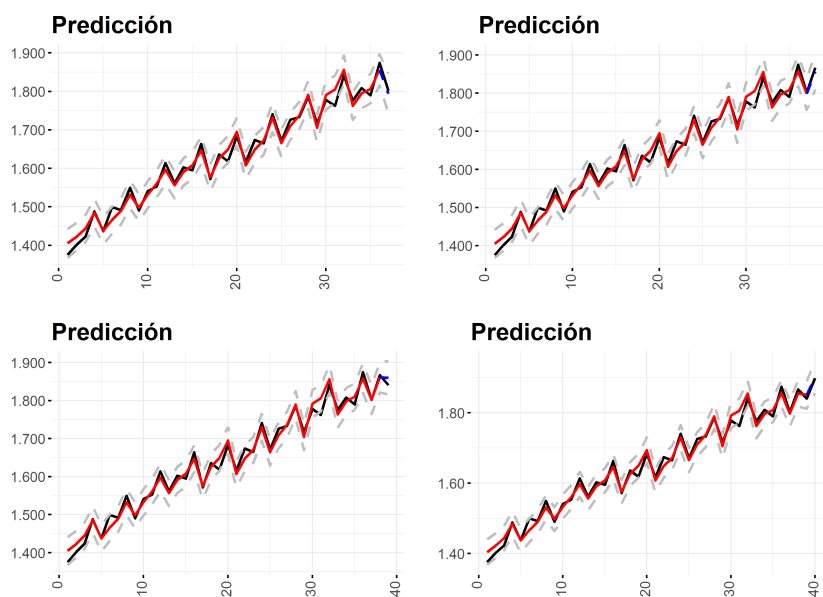


Figura 4.5: Predicciones 2018.

2019

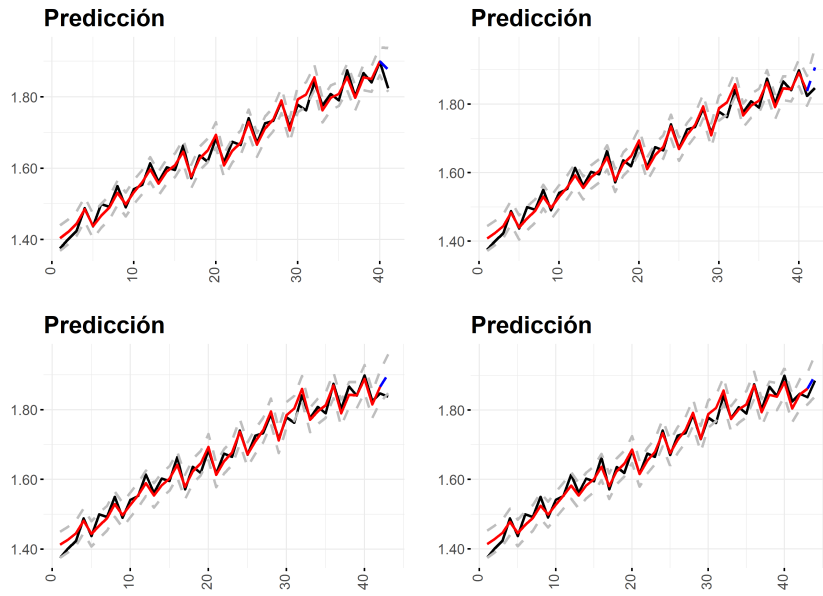


Figura 4.6: Predicciones 2019.

2020

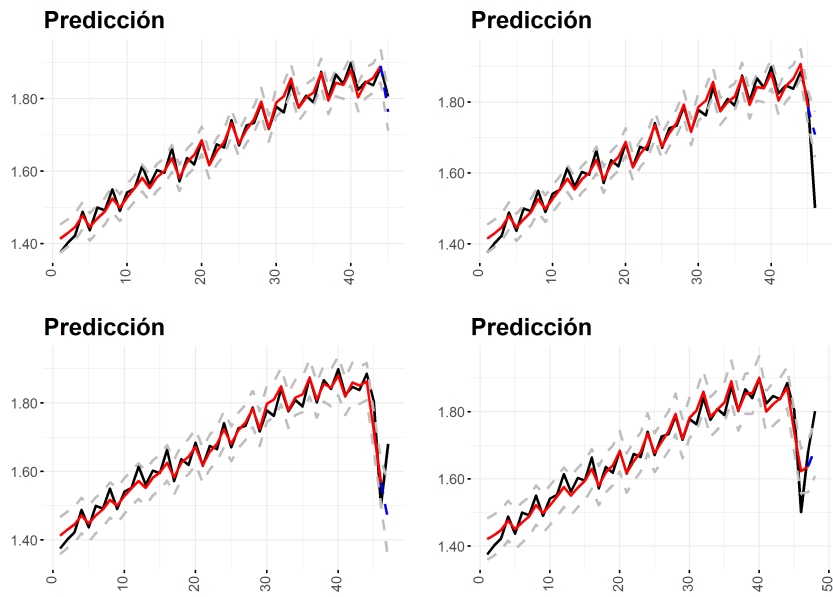


Figura 4.7: Predicciones 2020.

2021

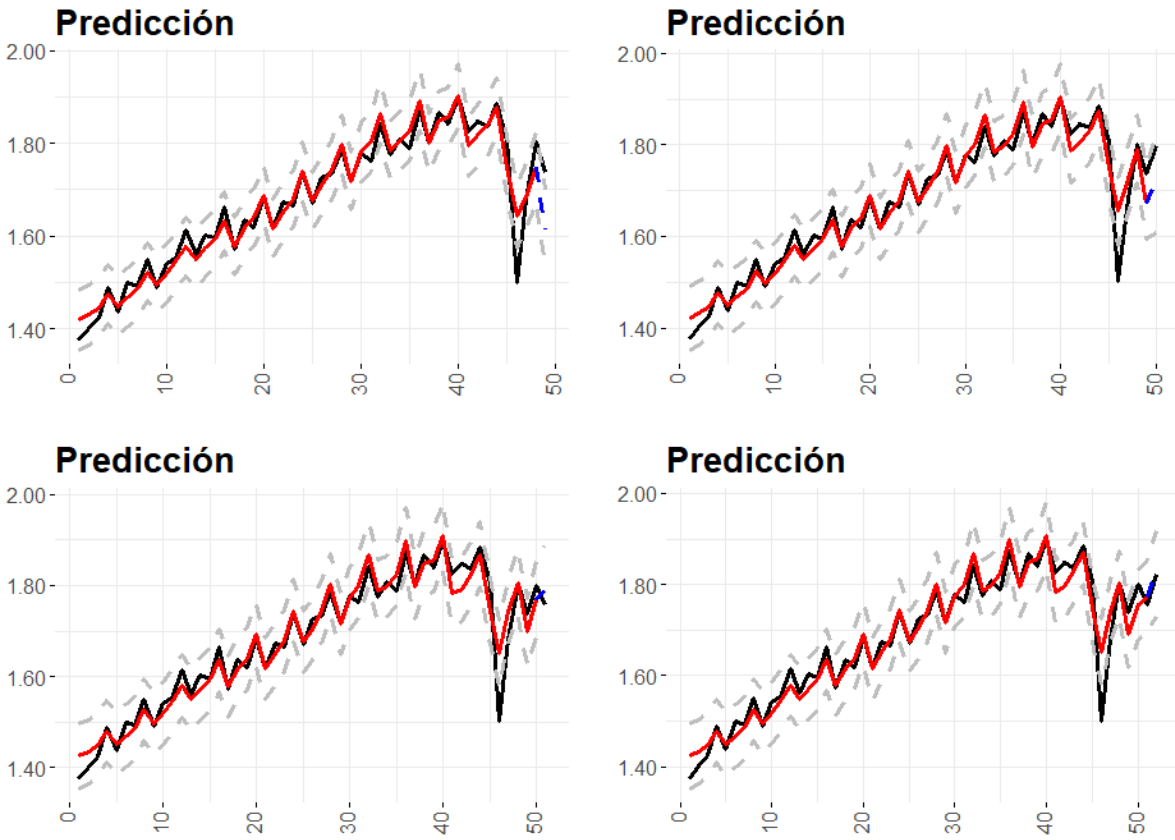


Figura 4.8: Predicciones 2021.

Bibliografía

- [1] FMI et. al. Sistema de Cuentas Nacionales 2008. Inf. téc. Fondo Monetario Internacional, 2016.
- [2] Valentina Aprigliano, Guerino Ardizzi y Libero Monteforte. “Using Payment System Data to Forecast Economic Activity”. En: vol. 15, issue 4 (2019), págs. 55-80.
- [3] W. Barnett, D. Leiva-Leon M. Chauvet y L. Su. Nowcasting Nom. GDP with the Credit-Card Augmented Divisia Monetary Agg. Inf. téc. University of Kansas, Department of Economics., 2016.
- [4] M. Carlsen y P. E. Storgaard. Dankort Payments as a Ind of Retail Sales in Denmark. Inf. téc. Danmarks Nationalbank, Copenhagen, 2010.
- [5] Secretaría de Hacienda y Crédito Público. “¿Quieres conocer la historia de la tarjeta de crédito?” En: (2018). URL: <https://www.gob.mx/shcp/articulos/quieres-conocer-la-historia-de-la-tarjeta-de-credito?idiom=es>.
- [6] Gerald P. Dwyer. The Johansen Tests for Cointegration. 2015.
- [7] Periódico El Economista. “Número de casos de Covid-19 en México al 31 de diciembre de 2020”. En: (2021). URL: <https://www.eleconomista.com.mx/politica/Numero-de-casos-de-Covid-19-en-Mexico-al-31-de-diciembre-de-2020-20201231-0032.html>.
- [8] P. S. Esteves. Are ATM/POS Data Relevant When Nowcasting Priv. Consumption? Inf. téc. Banco de Portugal, 2009.
- [9] Antonio Soriano Flores, ed. Regresión Lineal: Enfoque Bayesiano. UNAM-IIMAS.

- [10] J. W. Galbraith y G. Tkacz. Electronic Transactions as H.F. Ind of Eco Activity. Inf. téc. International Journal of Forecasting, 2007.
- [11] Francis Galton. “Regression Towards Mediocrity in Hereditary Stature.” En: Journal of the Anthropological Institute of Great Britain and Ireland (1886), págs. 246-263.
- [12] C. W. J. Granger y P. Newbold. “Spurious regressions in econometrics.” En: Vol. 2 Journal of Econometrics. (1974), págs. 111-120.
- [13] INEGI. Sistema de Cuentas Nacionales de México. PIB Trimestral. Inf. téc. INEGI, 2017.
- [14] Søren Johansen. “Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.” En: Econometric Theory Vol. 14, No. 4 (Aug., 1998) (1995), págs. 517-524.
- [15] Dirección General de los Sistemas de Pagos e Infraestructuras de Mercados. Informe Anual sobre las Infraestructuras de los Mercados Financieros 2020. Inf. téc. Banco de México, 2021.
- [16] T. Padoa-Schioppa, ed. Shaping the Payment System: A Central Bank’s Role. Bank of Korea’s Conference on Payment Systems. 2004.
- [17] Sudhakar Raju y Mahadevan Balakrishnan. “Nowcasting economic activity in India using payment systems data”. En: Volume 13 Number 1 Page 72 (2019), págs. 72-81.
- [18] J. D. Sargan y Alok Bhargava. “Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root Lies on the Unit Circle.” En: Vol. 51, No.3. Econometrica (1983).
- [19] J. D. Sargan y Alok Bhargava. “Testing residuals from least squares regressions for being generated by the Gaussian random walk.” En: Vol. 51, No.1. Econometrica (1983).
- [20] Comisión Nacional Bancaria y de Valores. Panorama Anual de Inclusión Financiera. Inf. téc. CNBV, 2021.
- [21] David Williams. Weighing the Odds. Cambridge University Press, 1991.

- [22] Campbell John Y. y Perron Pierre. “Pitfalls and Opportunities: What macroeconomists should know about unit roots.” En: National Bureau of economic research (1991).