

---

# Un vistazo a las transacciones desde el análisis de datos funcionales

## Viewing transactions using functional data analysis

Carlos Gómez<sup>a</sup>  
carlosgomez@usantotomas.edu.co

Omar Montejó<sup>b</sup>  
omarmontejó@usantotomas.edu.co

Anyi Moreno<sup>c</sup>  
anyimoreno@usantotomas.edu.co

---

### Resumen

El análisis de datos funcionales ha comenzado a tener gran relevancia gracias a su utilidad en la reducción de dimensionalidad, por su implementación de métodos longitudinales y multivaridos, donde las observaciones pueden ser tratadas como funciones en un intervalo cerrado. En este trabajo se presenta un análisis de las operaciones que realizan los clientes de las entidades financieras por medio de los canales transaccionales, clasificándolos en presenciales y virtuales, con el fin de identificar características y comportamiento de los clientes, tanto en su número como el monto que transan, desde un enfoque funcional. Se realiza un estudio descriptivo de los datos funcionales, contrastando sus resultados con las variables Producto Interno Bruto e Inflación, se aplica un cluster por medio del algoritmo de  $k$ -medias que permite identificar agrupaciones entre las entidades consideradas dentro del estudio, se realiza una comparación de medias tanto para el número y el monto de transacciones en los canales transaccionales mencionados, identificando punto de cambio en la media funcional. Finalizando el análisis se evidencia un fuerte cambio en el comportamiento de los datos considerados, esto debido posiblemente a la pandemia, lo que hizo que las transacciones por medio de canales presenciales disminuyeran, fortaleciendo por tanto las transacciones por medio de canales virtuales.

**Palabras clave:** Colombia; Entidades financieras; Transacciones; Datos funcionales; Análisis de datos funcionales.

### Abstract

---

<sup>a</sup>Universidad de Santo Tomás, Bogotá, Colombia, Maestría en Estadística Aplicada

<sup>b</sup>Universidad de Santo Tomás, Bogotá, Colombia, Maestría en Estadística Aplicada

<sup>c</sup>Universidad de Santo Tomás, Bogotá, Colombia, Maestría en Estadística Aplicada

Functional data analysis has begun to gain great importance due to its usefulness in dimensionality reduction, due to its implementation of longitudinal and multivariate methods, where observations can be treated as functions on a closed interval. This paper presents an analysis of the operations carried out by customers of financial institutions through transactional channels, classifying them into face-to-face and virtual, in order to identify characteristics and behavior of customers, both in number and amount. that trade, from a functional approach. A descriptive study of the functional data is carried out, contrasting its results with the variables Gross Domestic Product and Inflation, a cluster is applied through the k-means algorithm that allows identifying groupings between the entities considered within the study, a comparison is made of means for both the number and amount of transactions in the mentioned transactional channels, identifying a point of change in the functional mean. At the end of the analysis, a strong change in the behavior of the data considered is evident, possibly due to the pandemic, which caused transactions through face-to-face channels to decrease, thus strengthening transactions through virtual channels.

**Keywords:** Colombia; Financial entities; transactions; Functional data; Functional data analysis.

## 1. Introducción

El presente artículo abordará dos temas de interés relacionados con el comportamiento de las transacciones, en términos de monto y número, que reportan las entidades vigiladas por la Superintendencia Financiera de Colombia (SFC), el primero es analizar las transacciones frente al comportamiento del Producto Interno Bruto y la Inflación, estas como principales indicadores económicos, a fin de observar si estas variables tienen relación es su tendencias, el segundo tema esta relacionado con el comportamiento de los clientes y consumidores financieros frente a la tendencia del uso de canales no presenciales (virtuales), definido como las transacciones que se realizan por medios remotos y que cursan mediante los canales banca móvil, ACH , Internet, débito automático y audio respuesta, y los presenciales, definidos como el encuentro físico con el cliente mediante el uso de un instrumento de pago, como por ejemplo tarjetas débito, talonarios cheques, entre otros, para lo cual cursan mediante los canales oficinas, cajeros automáticos y datáfonos, con el propósito de establecer si el comportamiento de los clientes y consumidores financieros han cambiado su hábito de disposición de los recursos monetarios.

Para lograr los objetivos se utilizará la metodología de análisis de datos funcionales (FDA), considerando el grupo de entidades que reportan al Ente de Control, a fin de reducir su dimensionalidad sin perder información, por ende, explicar de manera sencilla el comportamiento de las transacciones como una sola agrupación, en este sentido, realizaremos análisis de diferencias de medias y armónicos a fin de observar tendencias, mediante análisis de cluster establecer grupos de entidades comunes

conforme a su transaccionalidad.

## Bases de datos utilizadas en el análisis

- **Transacciones:** La SFC, mediante Circular Externa 014 expedida el 17 de abril de 2008, adoptó el formato número 444, denominado Transacciones a través de los canales de distribución, donde imparte instrucciones a sus vigiladas tendientes a que reporten de forma semestral las transacciones que cursan en los canales de distribución, tales como: las oficinas de atención al público, los cajeros automáticos, el sistema de audio respuesta, Internet, telefonía móvil, débito automático, datáfonos y ACH . Los tipos de entidades obligados a reportar en el formato son: los establecimientos bancarios, las compañías de financiamiento, las cooperativas financieras, las administradoras de sistemas de pago de bajo valor, las sociedades especializadas en depósitos y pagos electrónicos SEDPE y el Banco de la República. Del formato referido se extrajo la siguiente información:
  1. Entidad vigilada, nombre de la entidad que debe reportar la información a la SFC, este parámetro fue anonimizado a fin de referirnos a las industrias que reportan la información.
  2. Número total de transacción, cantidad de transacciones que se realizan en el período de reporte.
  3. Monto total de transacciones, valor en pesos del monto de transacciones que se realizan en el período de reporte.
  4. Nombre de la unidad de captura, la columna contiene los canales de distribución de las oficinas de atención al público, los cajeros automáticos, el sistema de audio respuesta, Internet, telefonía móvil, débito automático, datáfonos y ACH. Ahora bien, para el análisis de hábitos transaccionales se construyó cuatro grupos de variables por monto y número de transacciones, así:
    - a) Monto de transacciones virtuales: banca móvil, ACH, Internet, débito automático y audio respuesta.
    - b) Número de transacciones virtuales: banca móvil, ACH, Internet, débito automático y audio respuesta.
    - c) Monto de transacciones presenciales: oficinas, cajeros automáticos y datáfonos.
    - d) Número de transacciones presenciales: oficinas, cajeros automáticos y datáfonos.
- **Inflación:** Para la información asociada a la inflación, se consideró los datos de 12 meses corridos mensuales, desde enero de 2008 hasta diciembre de 2020.
- **Producto Interno Bruto PIB:** Los datos tomados para el PIB son trimestrales a precios corrientes y corregidos de efectos estacionales y de calendario, de enero de 2008 hasta diciembre de 2020.

## 2. Marco Teórico

### 2.1. Análisis de datos funcionales

El análisis de datos funcionales implica el estudio de variables aleatorias las cuales toman un conjunto de valores en un espacio funcional, comprendido como un conjunto de funciones que definen un fenómeno específico.

Esta teoría aparece como una necesidad de la sociedad para comprender a profundidad las grandes cantidades de información que se manejan en el mundo actual, por lo cual dentro del tema surge la necesidad de construir nociones que ayuden a comprender los datos, además de producir estructuras matemáticas que logren una adquisición de herramientas para obtener el análisis óptimo de los datos tratados.

Un dato funcional puede comprenderse como una función  $f(t)$  finita que pertenece a un intervalo continuo. Las observaciones han de ser analizada por medio de su forma funcional lo cual implica definir el espacio de las funciones cuadrado integrable en la ecuación (tal), lo cual permite aproximar los registros de los datos funcionales.

$$L^2(T) = f : R \rightarrow R \mid \int_a^b f(t)^2 dt < \infty \quad (1)$$

con el producto interno definido como

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt \quad (2)$$

#### 2.1.1. Representación en B-Spline

Para modelar los datos muestreados como datos funcionales se aproximan por medio de una combinación lineal de puntos de control  $P_i$  y funciones base B-spline  $N_{i,k}(t)$ :

$$r(t) = \sum_{i=0}^n P_i N_{i,k}(t) \quad n \geq k - 1, \quad t \in [t_{k-1}, t_{n+1}] \quad (3)$$

Esta representación permite el análisis de los datos por medio de medidas estadísticas estructuradas especialmente desde la teoría de análisis funcional.

### 2.1.2. Estadísticos descriptivos en análisis funcional

Dado el conjunto de datos funcionales  $f_1, f_2, \dots, f_3$  definidos en el intervalo  $[a, b]$ . Las funciones descriptiva de la media esta dada por:

$$f(\bar{t}) = \frac{1}{n} \sum_{j=1}^n f_j(t) \quad (4)$$

### 2.1.3. Análisis de conglomerados sobre datos funcionales

El método explícito en este margen usa el algoritmo k-medias funcional, el cual estima una segmentación de las muestras considerando el uso de componentes principales permitiendo la mínima variabilidad entre los individuos de cada grupo, de forma que se logren grupos con características únicas.

### 2.1.4. Componentes principales

El propósito del análisis en componentes principales es conseguir la máxima información dentro de una combinación lineal de funciones, logrando obtener una base de menor dimensión, por tanto halla los auto-valores y auto funciones de la función de covarianza  $cov(f(t_1), \bar{f}(t_2))$ .

El análisis de componentes principales multivariado funcional se puede reducir según Jacques (Jacques 2013) al usual PCA de la matriz  $\frac{1}{\sqrt{n-1}}AW^{1/2}$  donde  $A = (I_n - I_n(1/n, \dots, 1/n))\bar{A}$ , con  $\bar{A}$  es una matriz de coeficientes de expansión y  $W = \int_0^T \Phi(t) * Phi(t)'$  con  $T$  como el espacio donde se ubican los datos funcionales y  $\Phi$  los componentes basicos.

$$\frac{1}{\sqrt{n-1}}AW^{1/2}Au_j = \lambda_j u_j \quad (5)$$

### 2.1.5. Conglomerados por el método de k-medias

En este proyecto se utilizara el algoritmo de componentes principales funcionales por k-medias, lo cual nos lleva a minimizar la ecuación 6 dada por Yamamoto (Yamamoto 2014).

$$g(U, V) = \frac{1}{n} \sum_n \sum_{i=1}^{k=1} u_{ik} \|p_v x_i - p_v \bar{x}_k\|^2 \quad (6)$$

donde  $U = u_{ik}$ ,  $V = v_t$  funciones que conforman una base ortonormal del subespacio de proyección,  $p_v$  es un operador de proyeccion ortogonal y  $\bar{x}_k$  son los centroides de cada conglomerado.

## 2.2. Inferencia para la media de dos Muestras

Cuando se tienen dos muestras con medias  $\mu_1$  y  $\mu_2$ , el estimador puntual que permite determinar la diferencia entre  $\mu_1$  y  $\mu_2$  está dado por

$$\bar{X}_1 - \bar{X}_2$$

En el caso funcional, se tienen dos muestras funcionales  $X_1, X_2, \dots, X_N, X_1^*, X_2^*, \dots, X_M^*$ , se supone que satisfacen los siguiente modelos:

$$\begin{aligned} x_i(t) &= \mu(t) + \varepsilon_i(t) & 1 \leq i \leq N \\ x_i^*(t) &= \mu^*(t) + \varepsilon_i^*(t) & 1 \leq i \leq M \end{aligned}$$

El contraste de hipótesis a realizar es:

$$H_o : \mu = \mu^* \text{ en el espacio } L^2$$

$$H_a : \mu \neq \mu^*$$

Se tiene que:

$$\begin{aligned} \bar{X}_N &= \frac{1}{N} \sum_{i=1}^N X_i(t) \\ \bar{X}_M &= \frac{1}{M} \sum_{i=1}^M X_i^*(t) \end{aligned}$$

son estimadores insesgados de  $\mu(t)$  y  $\mu^*(t)$ , respectivamente. Se rechaza la hipótesis nula, de igualdad de medias, si la siguiente expresión se hace muy grande.

$$U_{N,M} = \frac{NM}{N+M} \int_0^1 \left( \bar{X}_N(t) - \bar{X}_M^*(t) \right)^2 dt$$

Teniendo en cuenta los métodos y teoremas planteados por Horváth, Kokoszka (2012) que establecen la convergencia asintótica de  $U_{N,M}$  bajo  $H_o$  y considerando los valores y funciones propios del operador  $Z = (1 - \theta)C + \theta C^*$ , donde  $C$  y  $C^*$  son covarianzas desconocidas, los valores propios del operador  $Z$  cumple con  $\tau_1 > \tau_2 > \dots > \tau_d > \tau_{d+1}$ . Además de considerar que las dos muestras son independientes, que  $\varepsilon_1, \dots, \varepsilon_N$  son independientes e idénticamente distribuidos con  $E\varepsilon_1(t) = 0$  y  $E\|\varepsilon_1\|^4 < \infty$  también que  $\varepsilon_1^*, \dots, \varepsilon_N^*$  son independientes e idénticamente distribuidos con  $E\varepsilon_1^*(t) = 0$  y  $E\|\varepsilon_1^*\|^4 < \infty$  y  $\frac{N}{N+M} \rightarrow \theta$  con  $0 \leq \theta \leq 1$ . Se obtienen estadísticos para determinar la igualdad de medias:

$$T_{N,M}^{(1)} = \frac{NM}{N+M} \sum_{k=1}^d \frac{\hat{a}_k^2}{\hat{\tau}_k} \quad (7)$$

$$T_{N,M}^{(2)} = \frac{NM}{N+M} \sum_{k=1}^d \hat{a}_k^2$$

para estimar  $\hat{\tau}_k$  se emplean los valores propios de la función de varianza empírica. Bajo la hipótesis nula y los supuestos anteriormente expuestos se tiene:

$$\begin{aligned} T_{N,M}^{(1)} &\xrightarrow{d} \chi^2(d) \\ T_{N,M}^{(2)} &\xrightarrow{d} \sum_{k=1}^d \tau_k N_k^2 \end{aligned} \quad (8)$$

donde  $\chi^2(d)$  es una variable aleatoria Chi-Cuadrado con  $d$  grados de libertad y  $N_1, N_2, \dots, N_d$  son variables aleatorias normal estándar.

### 2.3. Cambios en la Media Funcional

Es interesante estudiar el cambio de comportamiento que presenta la media en un conjunto de observaciones funcionales  $X_i$ , donde se asume como hipótesis nula que la media es constante contra la hipótesis alternativa donde se establece que la media es cambiante. Se asume que  $X_i = \mu + Y_i$ , donde  $E(Y_i) = 0$  como se esta trabajando en el espacio  $L^2$  se tiene que  $E(X_i) = \mu(t)$  con  $t \in [0, 1]$ , con lo cual la media sera considerada diferente de cero. En los datos funcionales un tipo de cambio sencillo y usual es denominado *Cambio Abrupto* entre una curva y otra, para identificar dichos cambios se emplea el siguiente modelo:

$$\begin{aligned} X_i &= \mu_1 + Y_i \quad 1 \leq i \leq k^* \\ X_i &= \mu_2 + Y_i \quad k^* \leq i \leq N \end{aligned}$$

donde  $k^*$  es un punto de cambio desconocido.

## 3. Metodología

El conjunto de datos a utilizar hace referencia al número de transacciones y al monto en canales presenciales y virtuales en Colombia, los cuales por sus grandes valores se escalaran dividiendo por  $10^6$ , por ende cada dato en referencia a estas variables se interpretara en términos de millones, se cuenta con 30 entidades con canales de carácter presencial y 29 de carácter virtual, los registros se realizan de manera semestral desde el primer periodo del 2008 y finales de 2020, por lo cual se cuenta con 26 periodos de tiempo (semestres), algunos de los sujetos (entidades) presentan datos faltantes en algunos periodos, por lo cual inicialmente se imputan los datos mediante un bosque aleatorio entrenado en los valores observados de la matriz de datos para predecir los valores perdidos, para esto se emplea la función

*missForest* de la librería en R con su mismo nombre. Luego, para hacer la conversión de la data a datos funcionales, se emplea una *Base Spline* ya que los datos no presentan ciclos para plantear una Base de Fourier.

Para la base *B-spline* se toman 9 elementos en la base y un orden de 4, construyendo polinomios de grado 3 en cada uno de los nodos, se emplea *smooth.basis* de la librería *fda* para generar el suavizamiento de los datos y lograr trabajar con estos desde un enfoque funcional, determinando posteriormente algunas medidas descriptivas, media, mediana, primera y segunda derivada de los datos, gráficos estadísticos Box plot funcional, Bag-plot y Hdr plot funcional que ayudan a la detección de curvas atípicas. atípicos, tanto para el número de transacciones como para el monto de las entidades de canales virtuales y presenciales como de los totales.

Prosiguiendo con el análisis se realiza un cluster para ACP, con lo cual se espera identificar dentro del conjunto de datos entidades de comportamientos similares, permitiendo reconocer comportamientos que acaten conductas parecidas, esto se ejecutara por medio del análisis de los estadísticos de cada cluster.

Finalmente se plantea una comparación de medias entre los canales presenciales y virtuales respecto al número de transacciones y los montos realizados durante los semestres del 2008 al 2020, además de identificar lo puntos de cambio en la media funcional.

## 4. Análisis y resultados

### 4.1. Análisis de las transacciones frente al comportamiento del Producto Interno Bruto (PIB) y la Inflación.

#### Datos originales y suavizamiento de las transacciones

En las gráficas que se muestran a continuación, relacionadas con el monto y número de transacciones de las series originales extraídas del formato 444 se observa que existe un grupo de entidades que sobresalen, además situación que se acentúa con el transcurso del tiempo.

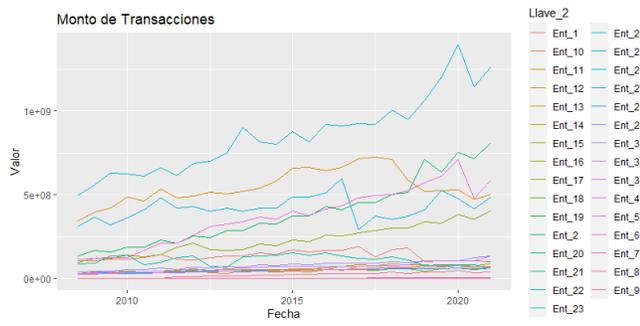


Figura 1: Monto de transacciones.

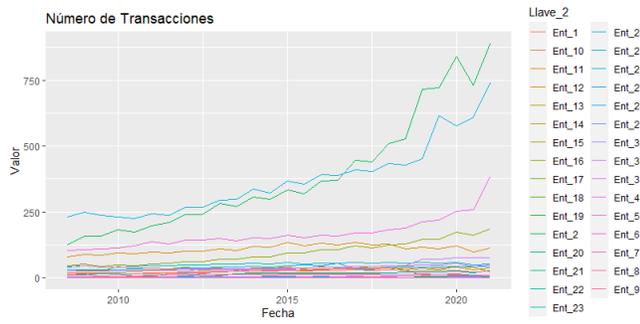


Figura 2: Número de transacciones.

Ahora bien, mediante el análisis de datos funcionales, se planteó ajustar los montos y número de transacciones, con una base B-spline con 9 elementos suavizados en la base, la cual como se observan en las gráficas se ajustan a los a los datos originales.

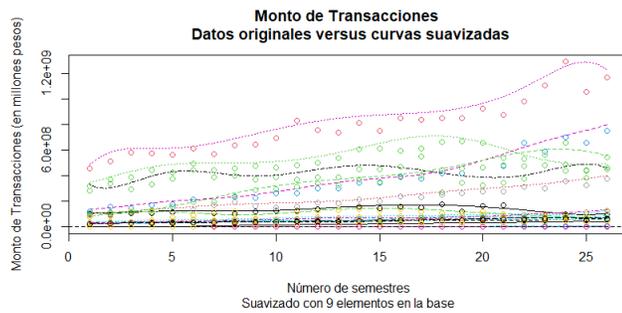


Figura 3: Comparativo curvas datos originales y suavizadas.

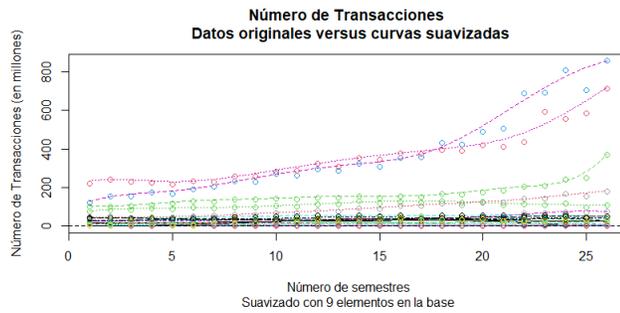


Figura 4: Comparativo curvas datos originales y suavizadas.

**Datos originales y suavizamiento de la inflación y el PIB.**

Así mismo, los datos de Inflación y PIB se ajustaron mediante datos funcional, aplicando al primero una base de Fourier con 30 elementos, considerando que la inflación presenta ciclos en su comportamiento, respecto al PIB, se aplicó una base B-spline con 51 elementos y una regularización que minimiza el criterio de validación cruzada generalizada, el ajuste aplicado recoge el comportamiento de las variables.

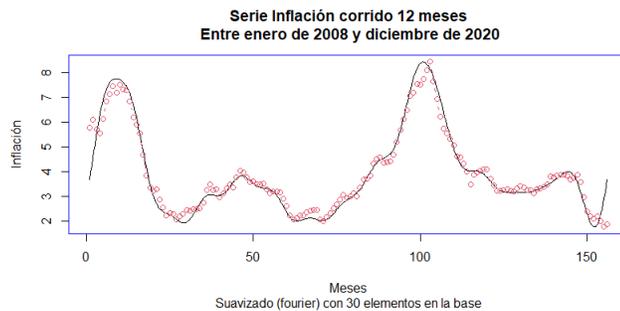


Figura 5: Comparativo curvas datos originales y suavizadas.

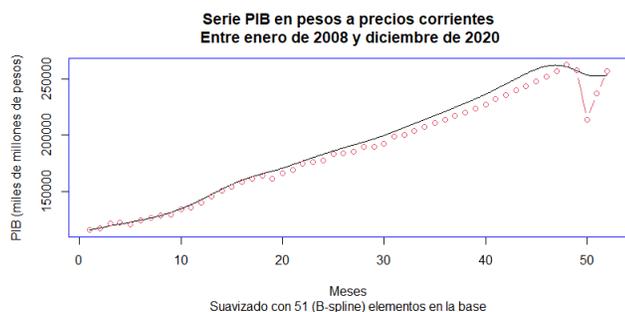


Figura 6: Comparativo curvas datos originales y suavizadas.

### Armónicos de las curvas suavizadas de las transacciones.

El análisis armónico de las curvas suavizadas del número de transacciones se observa que en la primera y segunda derivada, las entidades no son uniformes las variaciones en conjunto respecto a los cambios en el tiempo, resalta para un pequeño grupo, sus variaciones son más fuertes, como se aprecia en las siguientes gráficas.

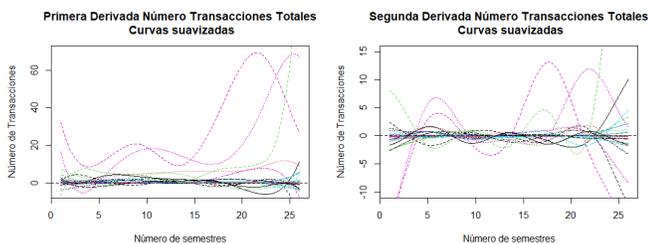


Figura 7: Primera y segunda derivada sobre las curvas suavizadas.

### Análisis de la media funcional y armónicos del número de transacciones.

La curva de la media funcional del número de transacciones resume el comportamiento de todas las curvas de las entidades, la cual muestra una tendencia creciente, con dos períodos de desaceleración marcados, el primero entre 2008 y 2009, el segundo entre el 2015 y 2016, posteriormente se observa una leve desaceleración en el primer semestre 2019, la anterior situación, no se detecta mediante el análisis con los armónicos de las curvas suavizadas, ya que cada entidad presenta un comportamiento disperso que impide observarlo, ahora bien, el comportamiento de media funcional, al contrastarlo con las curvas suavizadas del PIB y la inflación, no se observa relación alguna, como se muestra en la figura 5 y 6.

Teniendo en cuenta los impactos que ha generado la pandemia en Colombia y el mundo, el número de transacciones no se ve afectado para el año 2020, todo lo

contrario, muestra una senda de crecimiento, comenzando a partir del segundo semestre del del 2019.

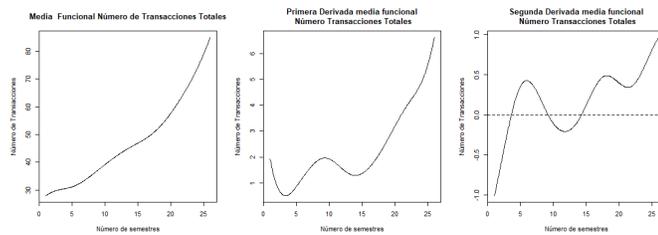


Figura 8: Media funcional número de transacciones.

### Análisis de la media funcional y armónicos del número de transacciones.

La media funcional del monto de transacciones recoge el comportamiento de todas las entidades, en él se observa que tiene una tendencia creciente hasta antes del inicio de la pandemia, en el primer semestre del 2020, es el único punto donde decrece los montos transados, precedido de una desaceleración que parecía cambiar de tendencia que se gestaba en el 2019, además presenta dos periodos muy fuertes de desaceleración, el primero entre 2008 y 2009 y el segundo entre 2015 y 2016, estos cambios son más pronunciados que los observados en el número de transacciones, como se refleja en las gráficas de la primera y segunda derivada del monto y número de operaciones (fig.9.).

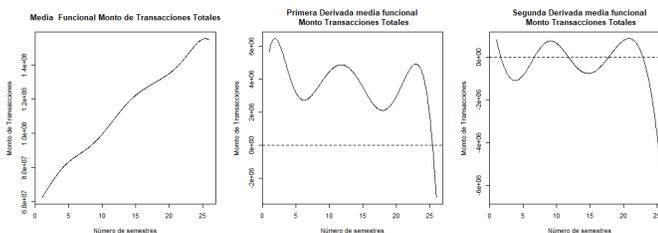


Figura 9: Media funcional monto de transacciones.

### Análisis de la media funcional y armónicos del monto de transacciones con contraste con al inflación y PIB funcional.

Contrastando el comportamiento de la media funcional y sus armónicos del monto de transacciones con los armónicos del PIB, se observa que coinciden con el declive del 2019 producto de la pandemia, el resto de la línea del PIB presenta situación que no explica el comportamiento del monto de las transacciones, mientras para el monto transado en el 2009 y 2010 se estaba es desaceleración, el PIB muestra una senda creciente sostenida desde 2009 hasta 2012 precedido de una leve desaceleración en el 2013, a partir del 2015 comienza una tendencia sostenida de

crecimiento hasta finales de 2018, se debe considerar que es posible que el PIB se vea afectado producto de la desestacionalización al momento de comparar con el monto transado, en todo caso, se debe validar esta situación con otros tipos de análisis y metodología, a fin de confirmar que las dos variables no tienen relación. Respecto de la inflación, variable económica que estima aumento generalizado y sostenido del nivel general de precios de una economía en el tiempo, por ende en los periodos de inflación alta supone alto nivel de rotación de flujos monetarios, no parece tener relación con el comportamiento de los montos transados, donde la inflación expresada en términos anuales para cada mes oscila, caso contrario que el monto de las transacciones que tiene una tendencia creciente.

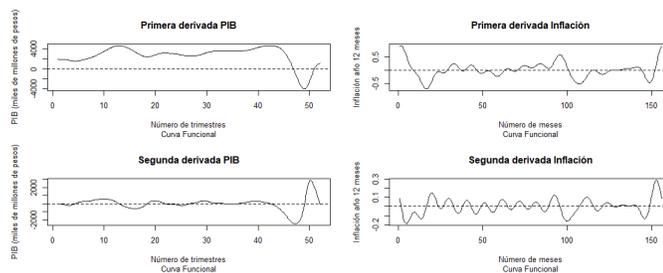


Figura 10: Media funcional.

### Análisis de clúters de las transacciones por monto y número.

Se plantea hacer un análisis de clústers, ya que las curvas suavizadas no muestran como se comportan las entidades a los largo del tiempo. EL resultado del modelo sugiere conformar cuatro grupo, ya las entidades son heterogéneo entre los grupos y homogénea en el grupo, tal como se muestra en la siguiente gráfica.

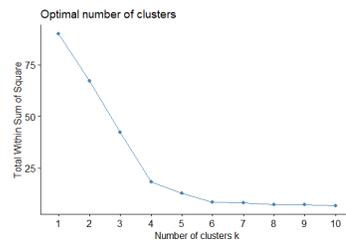


Figura 11: Tamaño de los grupos.

Como resultado de la agrupación tanto por monto como en grupo tenemos: Grupo 1: conformado por 4 bancos de tamaño medio y el grupo de bancos pequeños. Grupo 2: conformado por las redes de pago de bajo valor. Grupo 3: conformado por los banco pequeños, compañías de financiamiento, cooperativas y las entidades liquidadas. Grupo 4: conformado por los dos bancos más transaccionales del sistema financiero

En las gráficas que se muestran a continuación, relacionadas con el número de transacciones, se observa que para el grupo 1 viene con tendencia creciente hasta el 2018 y luego presenta un bache en su comportamiento, quedando en senda de crecimiento, como dato particular no se ha visto afectados por la pandemia. en el grupo dos, bienes con crecimientos moderados hasta 2017, a partir de este momento se acelera el crecimiento, sin verse afectado por la pandemia, el grupo tres tiene tendencia creciente con unos leves desaceleraciones en algunos momentos del tiempo, para el 2020 la pandemia afectó la cantidad de transacciones decreciendo y finalmente el grupo cuatro presenta un incremento importante en el número de transacciones a partir del año 2018, tal situación no ha sido afectada por la pandemia.

Las gráficas que siguen muestra el comportamiento de los grupos respecto a los montos transados, en tal sentido, el grupo uno entre el 2016 y 2018 muestra una fuerte decrecimiento, cambiando la tendencia en el 2020 sin verse afectados por la pandemia, el grupo dos tiene una tendencia creciente con leves desaceleraciones en el 2010 y 2017, el tercer grupo tiene baches de decrecimiento en los años 2010 y 2018, siendo impactado por el covid para el 2020 y el cuarto grupo la pandemia afecto su tendencia de crecimiento para el 2020 rompiendo la curva de crecimiento.

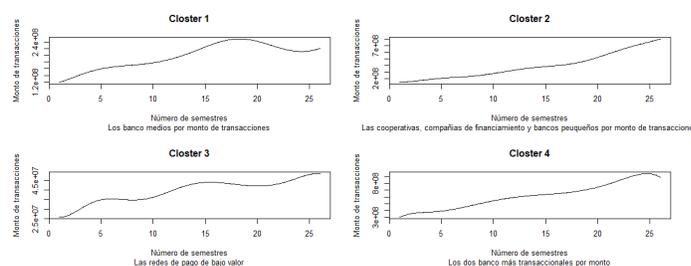


Figura 12: Grupos por monto de transacción.

## 4.2. Canales presenciales

Para el análisis del número de transacciones y el monto en los canales de carácter presencial se cuenta con 30 entidades, con los 26 registros de muestra, cabe resaltar que la data organizada de manera matricial presentaba valores faltantes que fueron imputados mediante un bosque aleatorio tomando los datos que se tienen para predecir los valores que faltan utilizando la librería *missForest*. En las curvas de las gráficas originales y suavizadas, gráficas 13 y 14, se evidencia un comportamiento homogéneo en la mayoría de las entidades en el número de transacciones, sin embargo, se resaltan algunas entidades con un comportamiento diferente, donde el número de transacciones mientras que en el monto de las transacciones, gráficas 15 y 16, se observa un comportamiento heterogéneo entre un gran número de entidades.

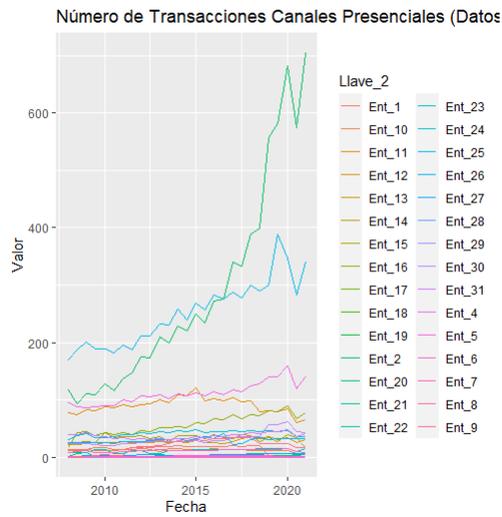


Figura 13: Número de transacciones canales presenciales 2008-2020

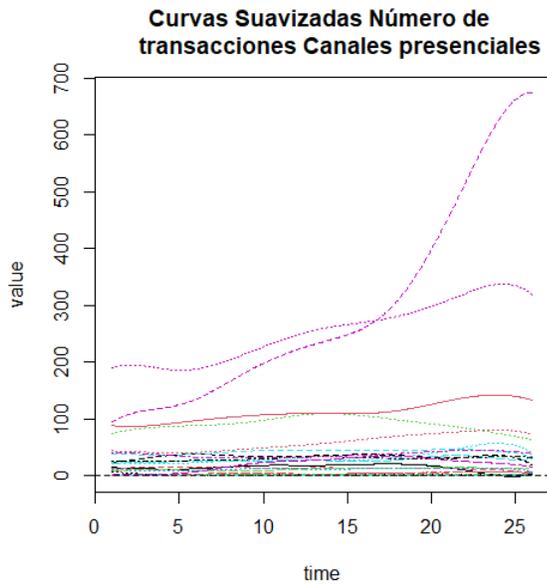


Figura 14: Curvas Suavizadas para Número de transacciones canales presenciales 2008-2020

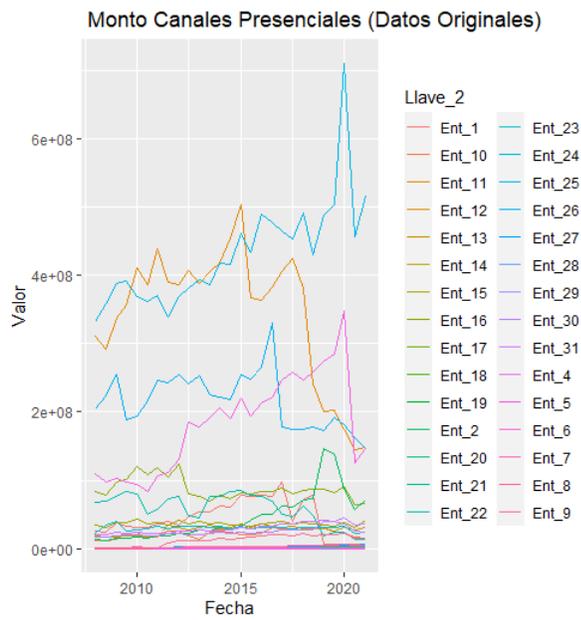


Figura 15: Monto de transacciones canales presenciales 2008-2020 (Datos originales)

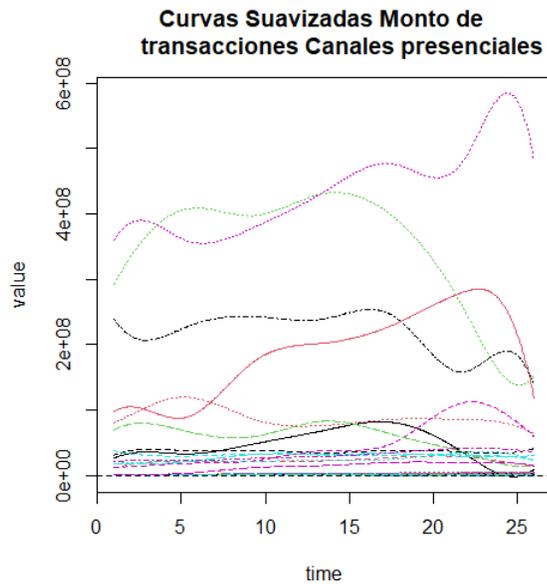


Figura 16: Curvas Suavizadas para el Monto de transacciones canales presenciales 2008-2020

La media funcional del número de transacciones y el monto de transacciones por medio de los canales de carácter presencial se puede observar en la gráfica 17, el promedio del número de transacciones presenta una tendencia creciente desde el 2008 hasta el 2020, en los primeros 5 semestres, a mediados del 2010 el promedio del número de transacciones esta alrededor del los 25 millones, en los próximos 4 años se observa que el promedio del número de transacciones por canales presenciales aumentaron aproximándose en 10 millones, logrando el mayor auge entre el semestre 20 y 25, es decir, a inicios del 2017 y 2020, alcanzando un promedio máximo en el número de transacciones cerca a 60 millones, en el segundo semestre del 2020 se observa que el promedio del número de transacciones en canales presenciales comienza a decrecer esto puede deberse a las medidas restrictivas y a todos los efectos generados por la pandemia del Covid-19.

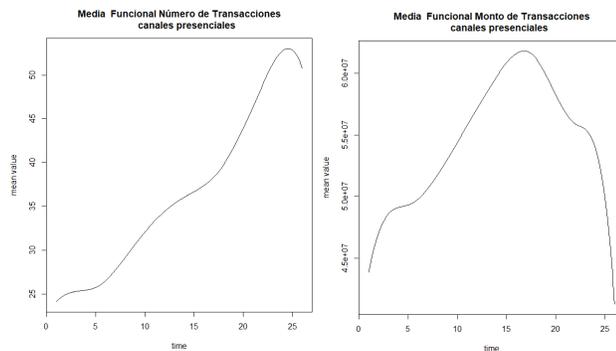


Figura 17: Media funcional del Número de Transacciones y monto de transacciones Canales Presenciales

El promedio del monto de las transacciones presenta un comportamiento creciente hasta mediados del 2015, en los primeros 5 semestres a partir de enero de 2008 se presenta un crecimiento lento pasando de un monto promedio de transacciones de  $4.5e+07$  a  $5.0e+07$  millones de pesos, entre 2010 y 2015 se observa un crecimiento acelerado en el monto promedio de transacciones alcanzando su punto máximo a mediados del 2015, luego el promedio del monto de las transacciones comienza a decrecer lentamente hasta mediados del segundo semestre de 2018, posterior a esto se evidencia un decrecimiento acelerado del promedio del monto de transferencia, siendo notorio en el segundo semestre del año 2020.

Por otro lado, puesto que existen entidades que acaparan gran parte de los canales presenciales, estas tienden a tener un comportamiento atípico tanto en el número de transacciones y el monto de las mismas, esto corroborar en las gráficas de las curvas con los datos originales y los suavizados, en análisis de datos funcionales se suelen emplear gráficos estadísticos como el *Box Plot*, *Bag plot* y *HDR plot* para identificar estas curvas atípicas. En las gráficas 18 y 19, se pueden observar el Bag Plot y HDR Plot para el número de transacciones y el monto de transacciones, respectivamente. Según el Bag Plot, para el número de transacciones se tienen 17 curvas atípicas y para el monto de transacciones 8 curvas atípicas, para el HDR plot la cantidad de curvas atípicas se reducen a 2 en ambos casos, coincidiendo en la curva 26.

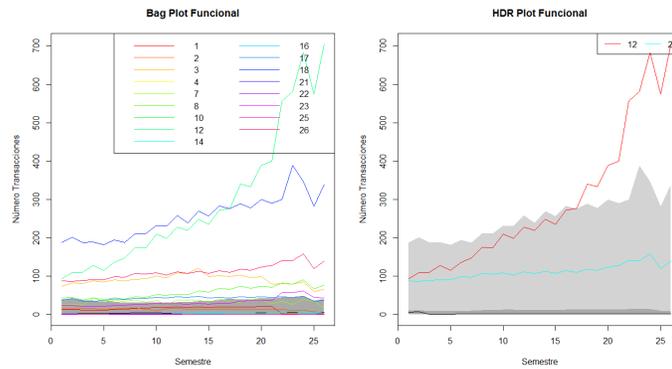


Figura 18: Bag Plot y HDR Plot funcional para Número de Transacciones Canales Presenciales

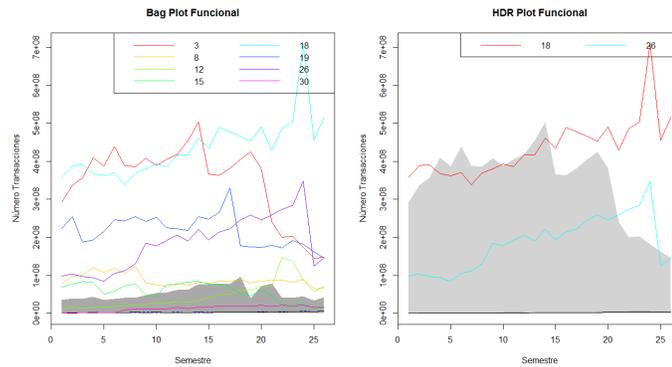


Figura 19: Bag Plot y HDR Plot funcional para Monto de Transacciones Canales Presenciales

### 4.3. Canales virtuales

Continuando se realiza el análisis de los datos, considerando únicamente las entidades que manejan entidades virtuales. En las curvas de las gráficas originales y suavizadas, gráficas 20 y 21, se evidencia un comportamiento homogéneo en la mayoría de las entidades, sobretodo en las que manejan baja cantidad de transacciones, sin embargo, se resaltan algunas entidades con un comportamiento diferente, donde el número de transacciones es alto, mientras que en el monto de las transacciones, gráficas 22 y 23, se observa un comportamiento similar.

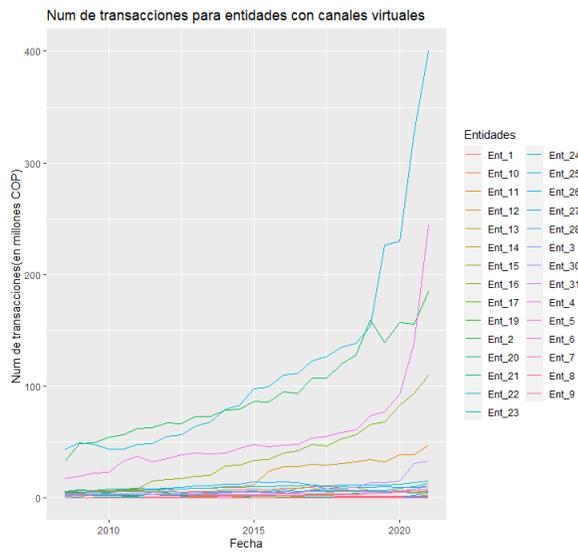


Figura 20: Número de transacciones canales virtuales 2008-2020

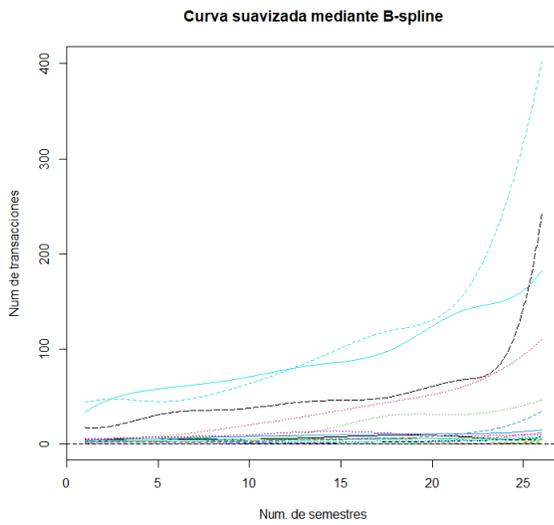


Figura 21: Curvas suavizadas para número de transacciones canales virtuales 2008-2020

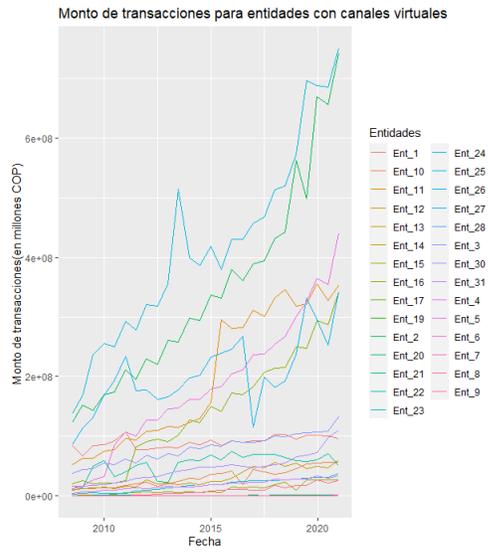


Figura 22: Monto de transacciones canales virtuales 2008-2020

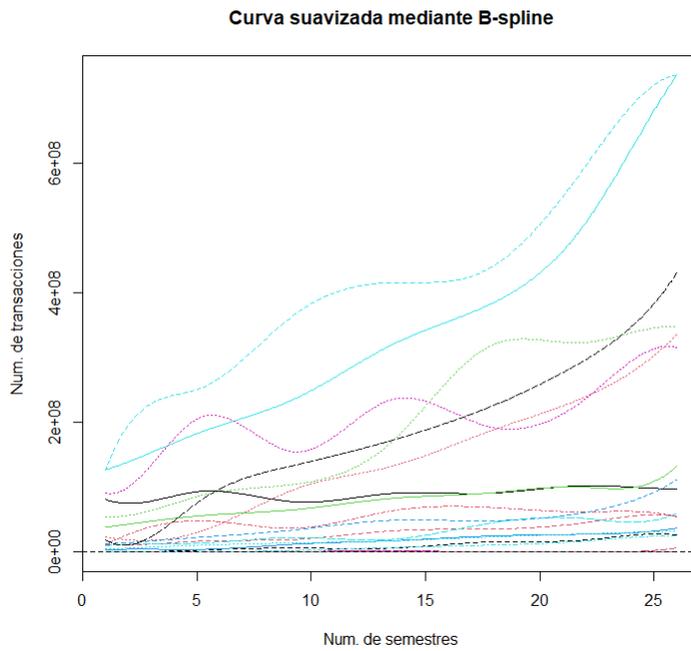


Figura 23: Curvas suavizadas monto de transacciones canales virtuales 2008-2020

Se puede ver un comportamiento uniforme en los primeros 20 semestres y un aumento después del semestre 20 lo cual puede ser dado por el fenómeno del covid-19.

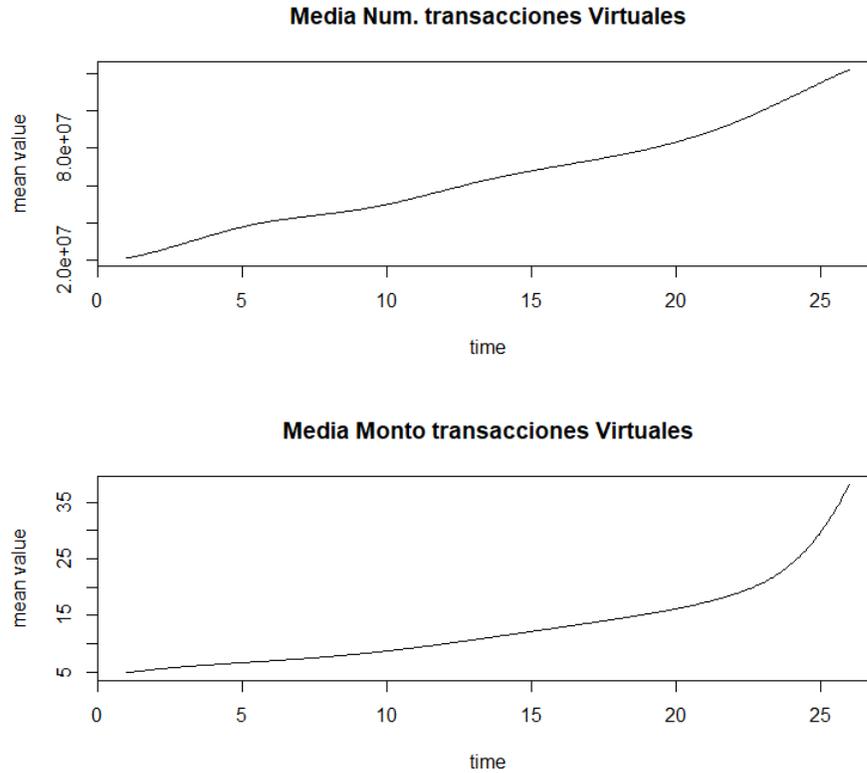


Figura 24: Promedio de numero de transacciones y monto de transacciones para entidades virtuales 2008-2020 (Datos originales)

El promedio del numero de las transacciones presenta un comportamiento creciente y no se muestra ninguna variación en este comportamiento que se mantiene a lo largo de los años. De comportamiento similar se puede ver la gráfica 24, promedio del monto de transacciones hechas en las entidades estudiadas, aun así esta crece lentamente en los primeros 22 semestres desde el 2008 y cambia su velocidad de cambio después del semestre 22, igualmente por un comportamiento acelerado.

igualmente que en el análisis de entidades que consideran canales presenciales, existen entidades que acaparan gran parte de los canales presenciales, estas tienden a tener un comportamiento atípico tanto en el número de transacciones y el monto de las mismas, esto se corrobora en las gráficas de las curvas con los datos originales y los suavizados. En las gráficas 25 y 26, se pueden observar el Bag Plot y HDR

Plot para el número de transacciones y el monto de transacciones, respectivamente. Según el Bag Plot, para el número de transacciones se tienen 5 curvas atípicas, igualmente para el monto de transacciones, para el HDR plot la cantidad de curvas atípicas se reducen a 2 en ambos casos, coincidiendo en la curva 25.

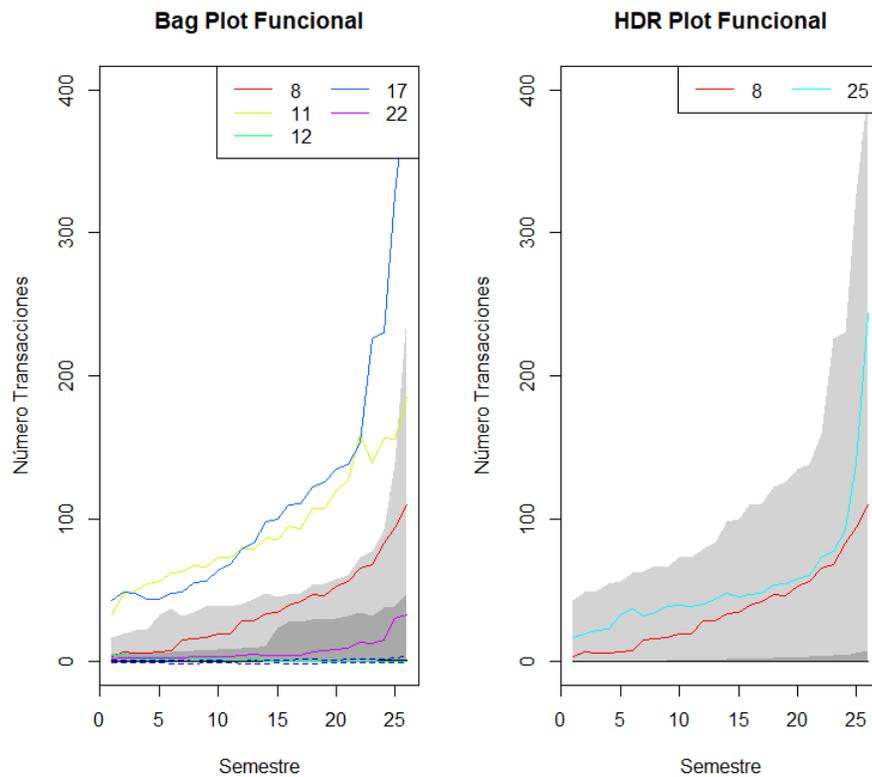


Figura 25: Bag Plot y HDR Plot funcional para Número de Transacciones canales Virtuales

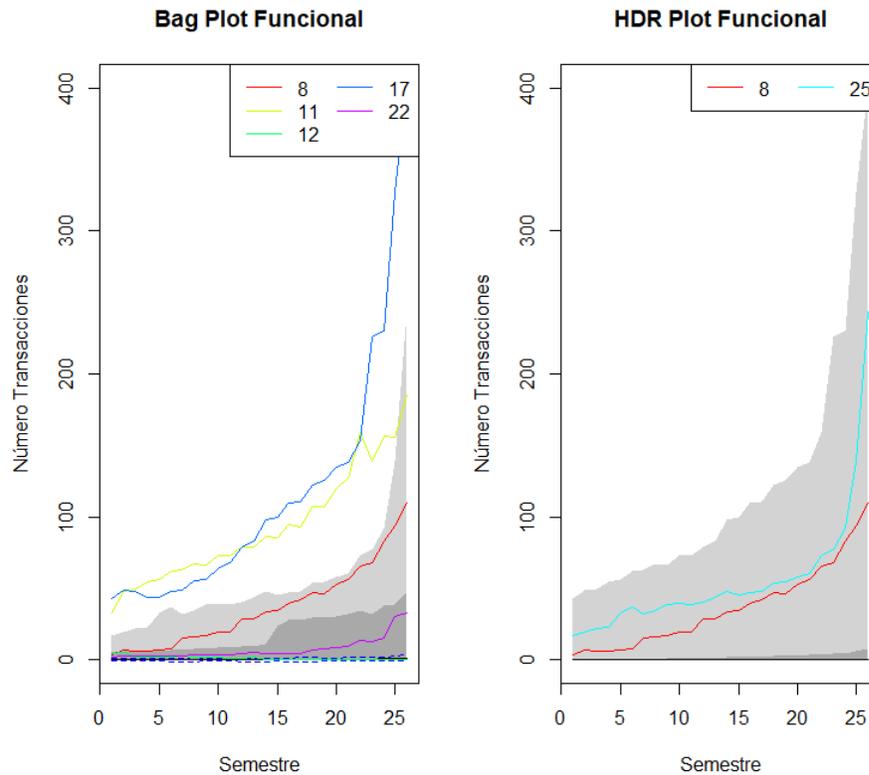


Figura 26: Bag Plot y HDR Plot funcional para monto de transacciones Canales virtuales

#### 4.4. Comparación de Medias

En esta sección se presenta el interés por hacer una distinción entre los canales de modalidad presencial y los de modalidad virtual, lo que se busca es comparar las medias del número de transacciones y del monto de las transacciones de los canales presenciales y virtuales. Para esto se emplea la expresión del estadístico T de la ecuación 7 y el primer resultado de la ecuación 8 de la sección 2.2. Para este caso la primera muestra corresponde a las canales presenciales, por lo tanto,  $N = 30$  y la segunda muestra es la referente a las canales virtuales  $M = 29$ . estimando  $\hat{a}_k^2$  y  $\hat{\tau}_k$  a partir de los valores propios de la matriz de covarianzas empírica, utilizando la función *fts.cov.structure* de la librería *freedom.fda* de R.

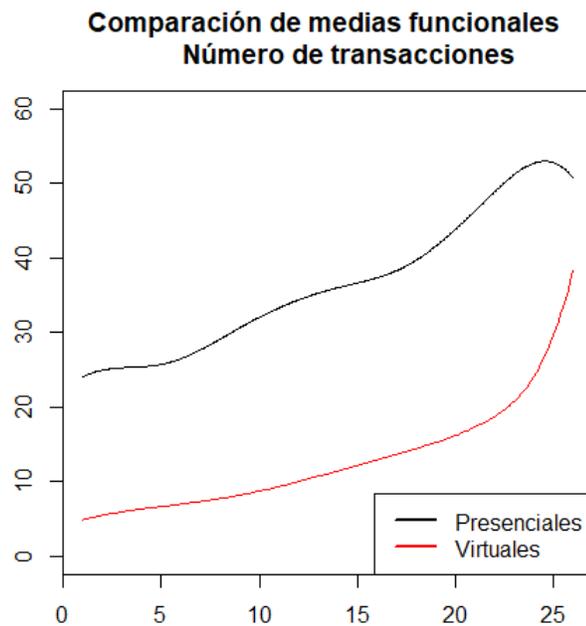


Figura 27: Diferencia de Medias Número de transacciones Canales Presenciales vs Canales Virtuales

Para aplicar el estadístico  $T$  de la ecuación 7, se hace un truncamiento en  $d = 4$ , teniendo como precedente que los datos fueron suavizados con 9 elementos en la base, el valor del estadístico  $T$  es de 11.11075, según la expresión 8, este resultado se puede contrastar con una Chi-Cuadrado  $\chi^2(4)$  con 5 grados de libertad, obteniendo 9.487729. Por tanto,  $T_{N,M}^{(1)} > \chi^2(d)$ , lo que implica que se rechaza la hipótesis nula  $H_o$  de igualdad de medias del números de transacciones de los canales presenciales y los virtuales. El estadístico  $T$  es mayor a un percentil de una Chi-cuadrado, por consiguiente, a un nivel de significancia del 5%, el número de transacciones en los canales presenciales durante el 2008 y 2020 son estadísticamente diferentes al número de transacciones realizados en los canales virtuales durante el mismo periodo. Dicho resultado se valida con la prueba *ANOVA*, la cual reporta un  $p$ -valor de 0.134, que rechaza la hipótesis nula.

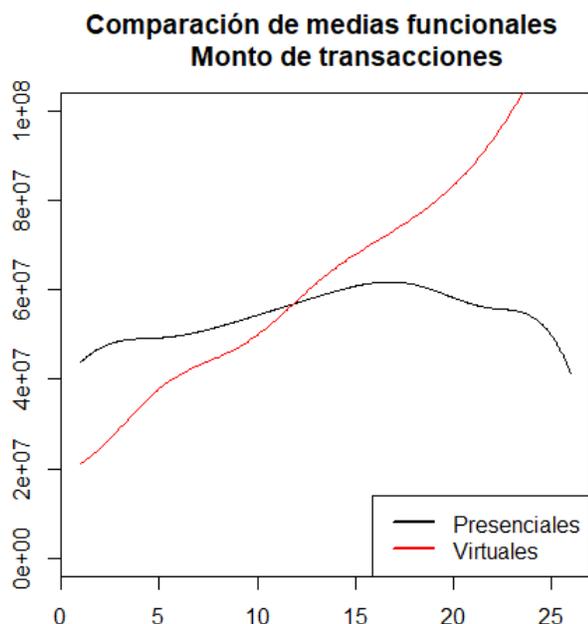


Figura 28: Diferencia de Medias Monto de transacciones Canales Presenciales vs Canales Virtuales

En cuanto la media del monto de transacciones, al aplicar el estadístico  $T$  y hacer el truncamiento en  $d = 4$ , se obtiene 13.15571 y se contrasta con una Chi-Cuadrado con 4 grados de libertad obteniendo 9.487729, lo que implica rechazar  $H_0$ , por lo tanto, las medias del monto de transacciones en canales presenciales y virtuales son estadísticamente diferentes bajo un nivel de significancia del 5%.

#### 4.5. Puntos de cambio en la media funcional

Determinar el punto de cambio de la media funcional resulta interesante, ya que permite detectar el punto en el tiempo en el cual la media sufre un cambio, en la gráfica 29, se muestra el comportamiento de la media del número de transacciones para el total, tomando en cuenta el número de transacciones tanto de canales presenciales como virtuales, para el total de transacciones se presenta un cambio en la media en el semestre 18, es decir, en el segundo semestre del 2016, después del punto de cambio, el promedio del número de transacciones aumento significativamente en comparación con el promedio del número de transacciones registradas en años anteriores, para los canales presenciales dicho cambio se da en el semestre 11, es decir, a inicio del año 2013 y para los canales virtuales se presenta en el

semestre 17, inicios del año 2016.

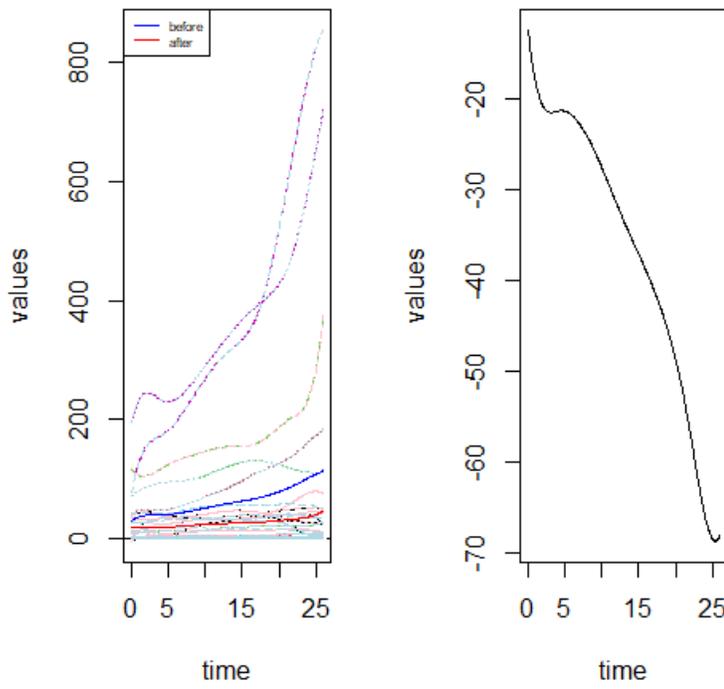


Figura 29: Punto de Cambio en la media funcional del Total de Número de Transacciones

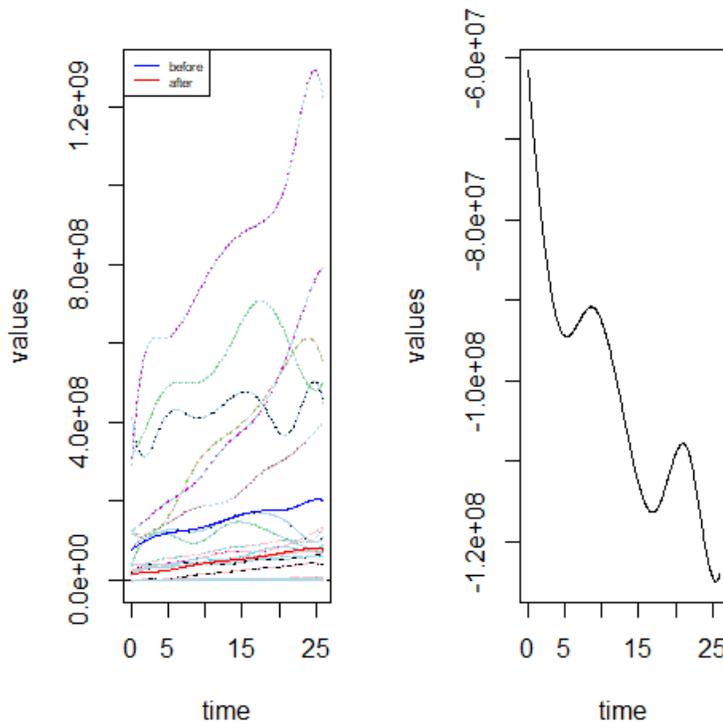


Figura 30: Punto de Cambio en la media funcional del Monto Total de Transacciones

En la media del monto total de transacciones, tanto presenciales como virtuales, se da un cambio en el semestre 19, es decir, en el primer semestre del 2017, mostrando un incremento notorio en los años posteriores al 2017 en la media del monto. La media del monto de transacciones para los canales presenciales experimenta también para el primer semestre del 2017, mientras que los canales virtuales lo presentan en el semestre 18, segundo semestre del 2016. En conclusión entre el 2016 y 2017 se presentaron puntos de cambios en la media del monto de transacciones en las entidades ya sea en sus canales presenciales y virtuales.

## 5. Conclusiones

Se evidencia mediante la metodología aplicada que los hábitos de los clientes en los últimos semestres analizados dentro de las entidades financieras han cambiado, ya que es más representativo la transaccionalidad virtual que la presencial, esto puede

ser debido en gran parte en el cambio abrupto que ha provocado la pandemia lo cual ha afectado la transaccionalidad de las entidades, a algunas de forma importante.

Los montos y número de transacciones al parecer no tienen relación causal con el PIB e inflación, ya que sus cambios no son consistentes, habrá que considerar en próximos trabajos otras técnicas a efectos de corroborar o desvirtuar la conclusión, es posible para este caso considerar desde la teoría funcional la regresión lineal o probit.

Las entidades que manejan grandes cantidades de transacciones en Colombia no se han visto afectadas por el fenómeno de la pandemia, en cambio parecen haber sido capaces de adaptarse al cambio y continuar con el mismo comportamiento que llevaban antes de la pandemia.

**Recibido: mayo**

**Aceptado: junio**

## Referencias

- [1] Carrillo Alexis and Garatejo Olga (2016) Análisis de datos funcionales aplicado en electroencefalogramas: Agrupamiento por K-medias funcional. Fundación Universitaria Los Libertadores.
- [2] J.O Ramsay et. al (2009) Functional data analysis with r and matlab. ISBN 978-0-387-98184-0. editorial Springer. ed 2009a.
- [3] Banca de las Oportunidades, [https://bancadelasoportunidades.gov.co/sites/default/files/2018-02/PDF\%202011\\_0.pdf](https://bancadelasoportunidades.gov.co/sites/default/files/2018-02/PDF\%202011_0.pdf) , Reporte de Inclusión Financiera, 2021
- [4] Banca de las Oportunidades, [https://bancadelasoportunidades.gov.co/sites/default/files/2020-07/Informe\\_RIF\\_2019.pdf](https://bancadelasoportunidades.gov.co/sites/default/files/2020-07/Informe_RIF_2019.pdf), 2021
- [5] Banca de las Oportunidades, <https://bancadelasoportunidades.gov.co/sites/default/files/2021-04/CNC\%20AD\%202021\%20AHORA\%20BCO.pdf>, 2021
- [6] Circular Externa 014 de 2008 expedida por la Superintendencia Financiera de Colombia, <https://www.superfinanciera.gov.co/inicio/normativa/normativa-general/circulares-externas-cartas-circulares-y-resoluciones-desde-el-ano-/circulares-externas/-20146>, 2021
- [7] Jacques, J. Preda, C. (2013), 'Functional data clustering: a survey', *Advances in Data Analysis and Classification* 8(3), 231–255
- [8] Yamamoto, M. Terada, Y. (2014), 'Functional factorial k-means analysis', *Computational Statistics Data Analysis* 79, 133–148

- [9] Horváth, L., Kokoszka, P. (2012). Inference for functional data with applications (Vol. 200). Springer Science Business Media. 65-68.