

N.º 21/05

Documento de trabajo

# Patrones de Consumo de Efectivo vs Tarjeta en México: una aproximación Big Data

Jaime Oliver, Álvaro Ortiz, Tomasa Rodrigo,  
Saidé Salazar e Ignacio Tamarit

Mayo de 2021

# Patrones de Consumo de Efectivo vs Tarjeta en México: una aproximación Big Data

Jaime Oliver (Clarity AI), Alvaro Ortiz (BBVA Research), Tomasa Rodrigo (BBVA Research), Saidé Salazar (BBVA Research), Ignacio Tamarit (Clarity AI)<sup>1</sup>

Mayo de 2021

## Resumen

El estudio de los patrones de consumo en los hogares resulta relevante para el bienestar social, la elaboración de políticas y el análisis económico. Los análisis empíricos tradicionales del comportamiento de los consumidores se basan en encuestas acerca del consumo en el hogar, las cuales ofrecen una imagen incompleta y presentan algunos problemas cuando predomina la informalidad en la economía. Este documento propone una nueva metodología que combina datos de alta frecuencia sobre operaciones con tarjeta e información del punto de venta (TPV) provenientes de operaciones en efectivo registradas en supermercados. Estudiamos los cambios de los patrones de consumo en México en relación con las variaciones de ingresos, lo que incluye la evolución del consumo de bienes y el uso de distintos canales de pago (en efectivo frente al electrónico). En concreto, analizamos cómo distribuye la población sus compras con tarjeta y en efectivo, utilizando una gama de modelos econométricos y de aprendizaje automático. Aprovechamos los últimos avances en torno a la explicación de modelos según los valores de Shapley para profundizar en el mejor modelo obtenido, Random Forest, que alcanza puntuaciones de  $R^2$  superiores a 0,92. Los resultados muestran que las variables más relevantes para incrementar el gasto con tarjeta en relación al uso de efectivo son la variación de ingresos, la residencia en el centro urbano y las variables de inclusión financiera. Mientras que el nivel de ingresos produce efectos no lineales de gran importancia para ambos tipos de transacciones, la residencia en el centro urbano y la inclusión financiera influyen cada vez más las operaciones con tarjeta.

**Palabras clave:** patrones de consumo, efectivo, pagos electrónicos, Big Data, elasticidad del consumo, aprendizaje automático, valores de Shapley

**Clasificación JEL:** C32, D12, O17, O54.

---

1: Nos gustaría expresar nuestro más sincero agradecimiento a Marta Rivera Alba y Tatiana Dávila por su contribución al presente documento y toda su ayuda. Sin su colaboración, no habríamos logrado concluir este estudio.

## 1. Introducción

El análisis de los patrones de consumo individuales y en los hogares puede convertirse en una herramienta importante para mejorar la vida de las personas en situación de necesidad. El análisis detallado de los datos geográficos y temporales puede permitir identificar tanto la escasez de recursos como la existencia de excedentes. A su vez, esta información también puede orientar a los legisladores a la hora de abordar iniciativas públicas y privadas encaminadas a la consecución de los Objetivos de Desarrollo Sostenible (ODS) de la ONU de «hambre cero» y «fin de la pobreza». Un mayor desglose de los perfiles de consumo por región o género podría servir de ayuda para otros ODS, como la igualdad de género, el crecimiento económico o la educación de calidad.

Tradicionalmente, los patrones de consumo se analizan mediante encuestas dirigidas a los hogares. Si bien este enfoque cuenta con muchas ventajas, también presenta ciertas deficiencias: exige mucho tiempo y una gran cantidad de recursos; por lo tanto, su frecuencia de actualización es baja y puede resultar insuficiente para capturar eventos clave.

En este trabajo, proponemos el empleo de datos de alta definición en tiempo real para monitorizar el consumo. Combinamos dos conjuntos de datos: uno extenso sobre pagos con tarjeta, proveniente del Big Data de un banco, y otro extraído de una plataforma de punto de venta (TPV) integrada que permite que las pequeñas empresas realicen un seguimiento de sus operaciones, la mayoría de las cuales se realizan en efectivo. Usamos el primero para representar transacciones digitales y el segundo para aproximar el consumo de efectivo. Este novedoso enfoque nos permite analizar dos visiones complementarias de la economía, contando así con una comprensión integral de los patrones de consumo a través de distintos canales de pago. Además, el alto grado de granularidad de los datos, tanto en términos temporales como de resolución de operaciones, proporciona información valiosa para la elaboración de políticas.

Los patrones de consumo revelan las preferencias de la población entre las distintas categorías de productos y servicios. El uso de modelos econométricos y de aprendizaje automático permite clasificar las formas en las que los diferentes grupos socioeconómicos perciben cada categoría de consumo como bienes de primera necesidad o bienes superiores, y estimar si estos grupos gastan más o menos en una determinada categoría. Este análisis nos permitirá identificar las categorías de consumo relevantes en las que centrarse con el propósito de lograr un mayor impacto social. Por tanto, la dimensión temporal de estos datos de representación de consumo podría permitir el análisis de cambios en el consumo, tanto a nivel de categoría como de marca, en tiempo real, debido a cambios en el presupuesto disponible de un consumidor, o bien como consecuencia de eventos personales o sociales.

Nuestro estudio se centra en el caso concreto de México, donde nuestra metodología combinada puede añadir un valor significativo, debido a la importancia de la economía informal, así como a sus altos niveles de desigualdad (ver Anexo 1). La inclusión financiera en México es relativamente débil, ya que la mayoría de la población con bajo nivel de ingresos permanece sin servicios bancarios y una parte sustancial de las operaciones se realiza mediante el uso de efectivo. Normalmente, los bajos niveles de ingresos se asocian a niveles de educación inferiores, menores habilidades laborales, empresas de menor capitalización (negocios familiares) y trabajos temporales. Estas características estructurales de la economía favorecen la informalidad económica y constituyen un factor importante que determina el uso de ciertos canales de pago. Como consecuencia, de una población de 130 millones, únicamente el 38,7% de los mexicanos tiene abierta una cuenta en alguna institución financiera. Con el propósito de abordar este problema, el Gobierno mexicano implementó el programa CoDi<sup>2</sup> para promover el uso de los medios de pago electrónico. Los datos de alta resolución permiten analizar el impacto de este tipo de

---

2: Programa puesto en marcha por el Banco de México para promover el uso de los medios de pago electrónico y reemplazar las operaciones en efectivo en la economía del país, mediante la incorporación de una plataforma de pago a través de teléfonos móviles que usa la tecnología de los códigos QR. Consulte el Anexo 1 para obtener más información.

iniciativas y estudiar la heterogeneidad de los patrones de consumo entre regiones, sectores y características de los clientes.

El presente documento cuenta con la estructura siguiente: la sección 2 describe las fuentes de datos para las operaciones con tarjeta y en efectivo que se han empleado en el proyecto, así como las principales variables explicativas que representan el nivel de ingresos, la urbanización y la inclusión financiera. Asimismo, la referida sección describe un análisis de datos exploratorio.

La sección 3 muestra los resultados de un modelo lineal, tanto para operaciones con tarjeta como en efectivo, en función de los niveles de ingresos, urbanización, ubicación geográfica e inclusión financiera como variables de control. Una vez estimado el modelo, también estimamos la semielasticidad del consumo en relación con los ingresos para distintos bienes y medios de pago.

En la sección 4, exploramos la capacidad de los modelos de aprendizaje automático para capturar las no linealidades de los patrones de consumo. Analizamos algunos modelos alternativos, como la regresión lineal, el modelo «Logit», «Random Forest» y «SVM», y convertimos la relevancia de las variables explicativas usando los valores de Shapley.

La última sección contiene los comentarios acerca de los resultados principales del estudio.

## 2. Datos

### 2.1 Descripción de los datos

Para caracterizar los patrones de consumo de la población mexicana, combinamos la información de las compras realizadas en efectivo en supermercados con el gasto con tarjetas de BBVA. Los datos de efectivo se registraron mediante una sencilla aplicación de [Frogtek](#), una empresa social con ánimo de lucro dedicada al desarrollo de herramientas comerciales para pequeñas empresas en mercados emergentes. Frogtek diseña sus propias aplicaciones móviles para tiendas y clientes en el extremo inferior de la cadena de valor. También ha creado una plataforma en la nube con servicios para socios, como proveedores de pagos, investigadores de mercado y empresas de bienes de consumo empaquetados. Uno de sus proyectos principales es Tiendatek, una plataforma de TPV integrada que permite que las pequeñas empresas realicen un seguimiento de su inventario y sus ventas, registren transacciones, obtengan métricas útiles y realicen cobros mediante tarjetas de crédito.

La muestra de Frogtek incluye 62 millones de operaciones, desde 2016 hasta 2018, de 1.835 tiendas en 12 estados de México, con un total de casi 26.000 productos diferentes. Añadimos los datos para cubrir 13 categorías de consumo de la clasificación del consumo individual por finalidad (CCIF) (entre las que se incluyen la alimentación, la ingesta de bebidas alcohólicas, el transporte, el equipamiento y la asistencia sanitaria), mientras que el conjunto de datos original identifica 171 tipos de productos. Pudimos atribuir un nivel de ingresos a los consumidores que optan por pagar en efectivo a través de la renta media del municipio. Además, asignamos el consumo a consumidores individuales añadiendo los datos de frecuencia de consumo por categoría obtenidos del Instituto Nacional de Estadística y Geografía de México (INEGI).

Los datos relativos al gasto con tarjeta se obtienen de los datos sobre operaciones de BBVA. BBVA es la institución bancaria más importante de México, con una cuota de mercado del 20% y presencia en siete países<sup>3</sup>. Analizamos todas las operaciones realizadas con una tarjeta de crédito o débito de BBVA, además de la información de los PV de BBVA, entre 2017 y 2018. Esto incluye 3.200 millones de transacciones distribuidas en 32 estados y 1.987 municipios, y 344 tipos de categorías de consumo. Asimismo, utilizamos las características socioeconómicas de los clientes para dividir los datos según cinco niveles distintos de ingresos.

## 2.2 Análisis exploratorio de datos

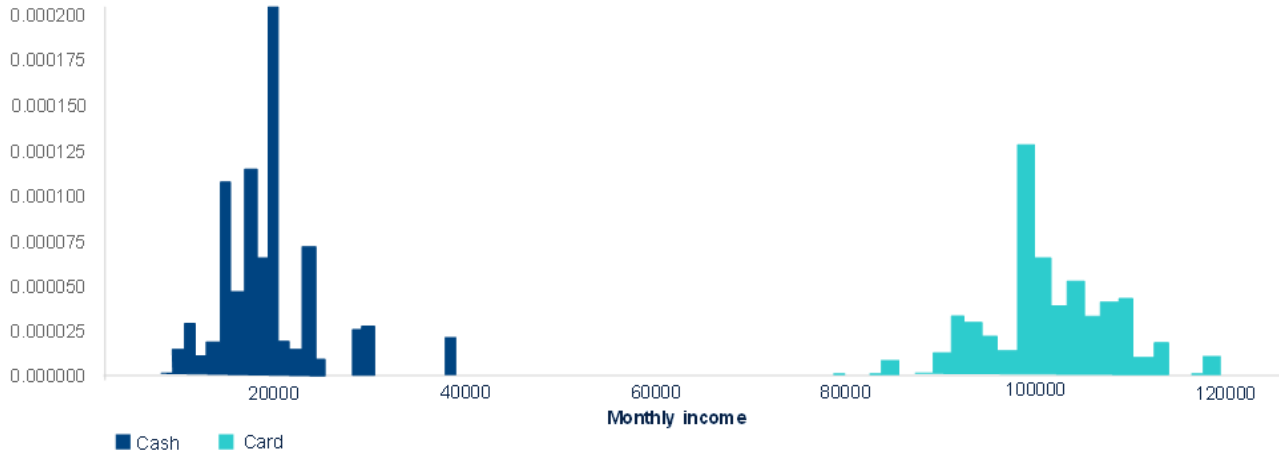
El análisis exploratorio de datos (EDA, por sus siglas en inglés) revela diferencias significativas entre las operaciones realizadas con tarjeta y en efectivo para distintos niveles de ingresos (ver Gráfica 1). Existe una mayor frecuencia de datos sobre operaciones con tarjeta en los niveles de ingresos más elevados; mientras que las transacciones en efectivo son más habituales en los niveles de ingresos inferiores. Por ello, la combinación de ambas fuentes de datos resulta fundamental para capturar sectores socioeconómicos complementarios.

La granularidad de los datos nos permite integrar datos de los TPV, así como datos bancarios. Añadimos datos con frecuencia mensual y nos enfocamos en 43 municipios, de los cuales existen datos en relación con el consumo a través de pago con tarjeta y en efectivo. En primer lugar, verificamos la existencia de sesgos. La Gráfica 2 compara la distribución de nuestra muestra de población con la demografía mexicana. No se observan sesgos poblacionales en cuanto al grado de urbanización de los municipios incluidos, como tampoco se contemplan en lo que concierne a los datos de pagos mediante el uso de tarjetas y en efectivo. A los efectos de esta comparación, consideramos cinco categorías de municipios<sup>4</sup> según el número de habitantes: rurales (menos de 5.000 habitantes), en transición (de 5.001 a 15.000 habitantes), semiurbanos (de 15.001 a 50.000 habitantes), urbanos (de 50.001 a 300.000 habitantes), semimetropolitanos (de 300.001 a 1 millón de habitantes) y metropolitanos (más de 1 millón de habitantes).

3: BBVA cuenta con sedes en España, Turquía, México, Perú, Colombia, Argentina y Venezuela.

4: Estas categorías se corresponden con la clasificación de municipios según el grado de urbanización que emplea la Comisión Nacional Bancaria y de Valores de México (CNBV) en su informe de inclusión financiera anual.

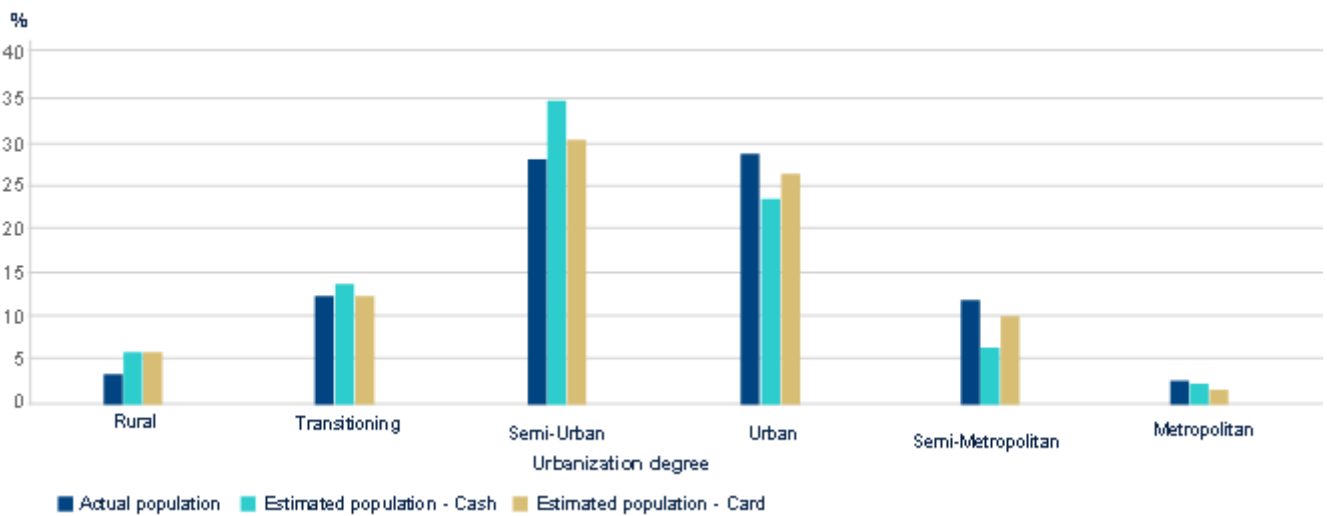
Gráfica 1. **DISTRIBUCIÓN DE INGRESOS SEGÚN DATOS DE EFECTIVO/TARJETAS**



Observamos que los datos de pagos mediante el uso de tarjetas se relacionan con los municipios con las rentas más elevadas, mientras que el uso de efectivo se vincula en mayor medida con los municipios con rentas inferiores. De esta forma, la naturaleza complementaria de los dos conjuntos de datos nos permite cubrir distintos segmentos socioeconómicos de la población.

Fuente: BBVA y Frogtek

Gráfica 2. **DISTRIBUCIÓN DEL GRADO DE URBANIZACIÓN SEGÚN DATOS DE EFECTIVO/TARJETAS**



Distribución de la población, tanto estimada como real, según el grado de urbanización. Observamos que ambos conjuntos de datos, efectivo y tarjetas, conforman una muestra representativa de la población real, si bien se aprecia una ligera tendencia de los datos de efectivo siendo más frecuentes en las zonas rurales, mientras que los pagos con tarjeta resultan más habituales en las zonas metropolitanas.

Fuente: CNBV, BBVA y Frogtek

Usamos los datos sobre pagos en efectivo y con tarjeta para elaborar un indicador de la asignación presupuestaria de cada individuo; es decir, el gasto en efectivo y con tarjeta por categoría y persona dividido por el nivel de ingresos. En el caso de los datos sobre efectivo, en línea con los informes oficiales del INEGI, asumimos que cada persona realiza seis compras al mes en una tienda que utiliza la aplicación Frogtek y suponemos que la renta media se corresponde con la del municipio de residencia (datos igualmente provenientes del INEGI). En lo que respecta a los datos acerca del uso de tarjetas, consideramos la cantidad total consumida dividida por el número

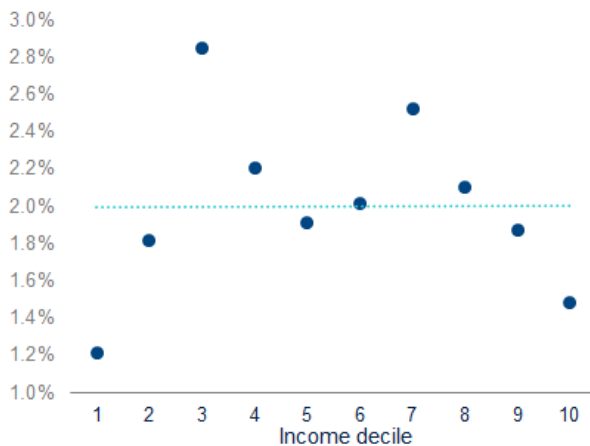
de clientes de BBVA en cada categoría y usamos el nivel de ingresos de cada cliente para calcular su asignación presupuestaria.

La evolución de las tendencias de pago con tarjeta y en efectivo está estrechamente relacionada con la evolución de la informalidad de la economía en México. En 2019, la economía informal representaba el 26,2% del PIB (ver Gráfica A2 en el Anexo 1) y alrededor del 56% de la población activa. Al mismo tiempo, la cantidad de billetes y monedas en circulación suponía el 62,7% del PIB y cerca del 95% de las compras cuyo importe equivalía a 500 pesos mexicanos o menos (25 USD o menos) se realizaron con pagos en efectivo. En lo que concierne a la división geográfica, apreciamos que los estados con menor renta muestran las mayores tasas de informalidad económica del país (ver Anexo 1).

Las decisiones individuales y de los hogares de consumir empleando tarjetas o efectivo, medidas como porcentaje del presupuesto total, dependen de múltiples factores. Un factor importante parece ser el nivel de ingresos. El análisis de la asignación presupuestaria por deciles de ingreso para el gasto en efectivo y con tarjeta muestra algunos patrones interesantes (Gráficas 3 y 4).

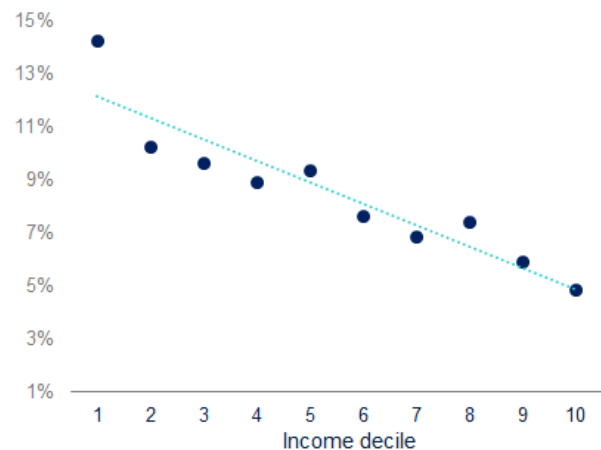
Si bien el gasto con tarjeta es menos sensible al decil de ingresos, encontramos una significativa relación negativa entre el gasto en efectivo y los ingresos, donde el decil de ingresos más alto tiene una asignación presupuestaria de gastos en efectivo de casi 10 puntos porcentuales por debajo del decil de ingresos más bajo (Gráfica 4). Los resultados del gasto con tarjeta, sin embargo, muestran una mayor asignación presupuestaria entre los deciles medios, pero se observan asignaciones similares en los extremos, tanto inferiores como superiores (Gráfica 3).

Gráfica 3. **DISTRIBUCIÓN DEL GASTO CON TARJETA POR DECIL DE INGRESOS**



Fuente: cálculos propios. BBVA y Clarity

Gráfica 4. **DISTRIBUCIÓN DEL GASTO EN EFECTIVO POR DECIL DE INGRESOS**



Fuente: cálculos propios. BBVA y Clarity

## 2.3 Análisis regional

Aprovechando los datos granulares por municipio, analizamos la evolución del gasto con tarjeta y en efectivo a nivel regional, y encontramos un patrón heterogéneo entre regiones y momentos temporales entre los 43 municipios analizados. En primer lugar, dentro de un mismo estado, los municipios con rentas elevadas suelen mostrar una correlación con un mayor aumento de los gastos con tarjeta a lo largo del tiempo. Por su parte, los municipios con rentas más reducidas y mayor densidad de población muestran una conexión con un mayor aumento de los gastos en efectivo. Estos hallazgos son coherentes con la correlación negativa, ampliamente documentada, entre el nivel de ingresos y la informalidad económica a nivel regional (ver Anexo 1). Las diferencias entre los municipios de un mismo estado ponen de manifiesto la distribución deficiente de la renta que también caracteriza al país (Gráficas 5 y 6).

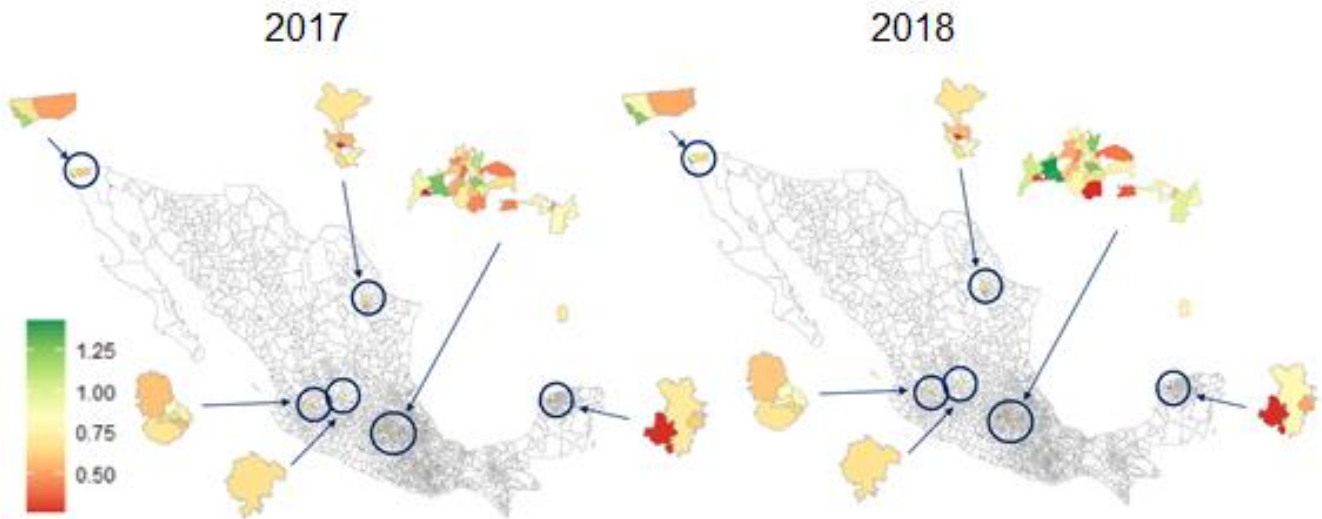
Encontramos patrones heterogéneos en las distintas regiones de la Ciudad de México: los municipios con menor renta per cápita muestran un mayor aumento del gasto en efectivo (Tláhuac, Tlalpan, Gustavo A. Madero), mientras que los más ricos registran el menor crecimiento en lo que respecta al uso de efectivo (p. ej., la demarcación territorial de Miguel Hidalgo) o incluso se observa una disminución (Iztacalco). En general, la mayoría de los municipios (a excepción de Iztacalco y Milpa Alta) muestran un aumento de la asignación presupuestaria en efectivo durante el periodo (2016 – 2018). Por el contrario, la asignación presupuestaria de gastos con tarjeta mostró un crecimiento prácticamente nulo en todos los municipios de Ciudad de México incluidos en este análisis durante el periodo 2017 – 2018. Globalmente, en la Ciudad de México, 44 de cada 100 trabajadores pertenecen al sector informal.

En Nuevo León, el tercer estado más rico del país, el municipio con la renta más elevada incluido en este estudio (San Nicolás de los Garza) mostró la asignación presupuestaria en efectivo más baja de la muestra estatal en 2018, mientras que uno de los municipios más pobres (Juárez) muestra el mayor uso de efectivo. En conjunto, en Nuevo León, 35 de cada 100 trabajadores pertenecen al sector informal. Como referencia, la tasa promedio de informalidad económica en las principales ciudades del país es del 41%, 13 p.p. por debajo de la cifra nacional (54%), lo que sugiere que la informalidad económica se concentra en zonas menos urbanizadas (ver Anexo 1 para obtener más información).

En Yucatán, el estado con menor renta de nuestra muestra, el municipio de Mérida (capital del susodicho estado) muestra un mayor aumento del gasto con tarjeta durante el periodo analizado en comparación con los municipios de Kanasín y Umán, que también registraron variaciones positivas sobre la asignación presupuestaria de gastos con tarjeta, aunque fueron menores. Resulta interesante la particularidad de Umán, que es el estado que registra el menor uso de efectivo de entre los incluidos en la muestra, justo después de San Nicolás de los Garza, en el estado de Nuevo León.

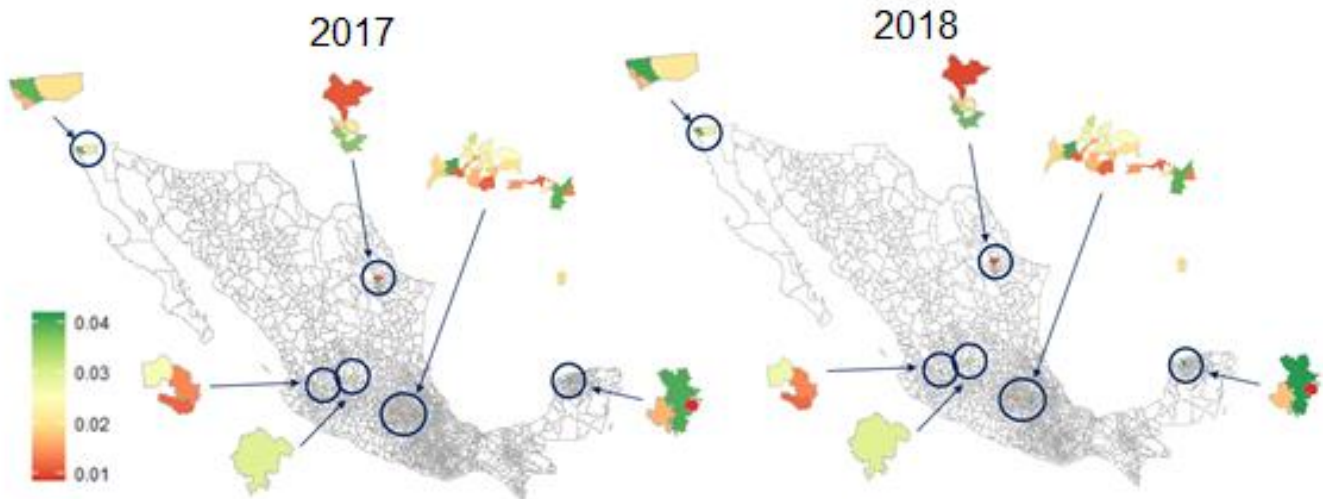


Gráfica 5. **ASIGNACIÓN PRESUPUESTARIA DE GASTOS EN EFECTIVO POR MUNICIPIO**



Fuente: cálculos propios. BBVA y Clarity

Gráfica 6. **ASIGNACIÓN PRESUPUESTARIA DE GASTOS CON TARJETA POR MUNICIPIO**



Fuente: cálculos propios. BBVA y Clarity

### 3. Modelo lineal de los patrones de consumo en México: tarjeta vs. efectivo

Una vez concluida la sección del análisis de datos exploratorio, formamos modelos estadísticos y de aprendizaje automático con el propósito de explicar los patrones de consumo en México. Esto nos brindará información complementaria acerca de los patrones de consumo y las operaciones de pago en México, además de cuantificar la relevancia de las variables.

#### 3.1 Modelo lineal de los patrones de consumo en México

A fin de analizar la relación entre ingresos y operaciones en efectivo y con tarjeta, modelamos la proporción de transacciones con tarjeta o en efectivo en relación con los ingresos  $\omega_{it} = \frac{C_{it}}{I_{it}}$  como una función de variables exógenas alternativas en la siguiente especificación de datos de panel

$$\omega_{it} = c + \beta_1 \log(I_{it}) + \beta_{iFD} X_{it}^{FD} + \beta_{iU} X_{it}^U + \beta_{iG} X_{it}^G + u_{it} \quad (1)$$

donde  $t$  es el índice de tiempo,  $i$  representa los municipios calculados según el promedio individual,  $C_{it}$  equivale al consumo en efectivo o mediante pagos con tarjeta, e  $I_{it}$  es el nivel de ingresos o renta. El modelo incluye una intersección, efectos fijos y un conjunto de variables explicativas: el efecto del nivel de ingresos o renta ( $I_{it}$ ) y un conjunto de variables de control, representativas de la inclusión financiera ( $X_{it}^{FD}$ ), el nivel de urbanización ( $X_{it}^U$ ) y la ubicación geográfica ( $X_{it}^G$ ).

El nivel de ingresos o renta se introduce en logaritmos  $\log(I_{it})$ , y al igual que la variable dependiente representa una parte del ingreso total  $\omega_{it}$ , el coeficiente  $\beta_1$  representa la semielasticidad del consumo en proporción al ingreso. Como una representación de la inclusión financiera ( $X_{it}^{FD}$ ), consideramos dos variables: contratos de depósito por 10.000 habitantes y contratos de crédito por 10.000 habitantes. La CNBV publica ambas variables, además de supervisar la evolución del uso de servicios financieros o la intensidad financiera por municipio. El grado de urbanización ( $X_{it}^U$ ) también es una variable de control importante, ya que existen razones en lo que se refiere a la demanda (nivel de ingresos o renta) y a la oferta (desarrollo bancario) que pueden afectar a los resultados. Por ello, presentamos variables ficticias categóricas que representan una zona metropolitana o rural. Por último, también añadimos variables geográficas ( $X_{it}^G$ ) para controlar las diferencias en el desarrollo económico entre municipios asociadas a la ubicación.

#### 3.2 Resultados

Los resultados de las estimaciones del modelo muestran algunas diferencias importantes entre las regresiones para el gasto en efectivo y mediante el uso de tarjetas. La semielasticidad de los ingresos muestra un efecto positivo significativo del nivel de ingresos sobre el gasto con tarjeta, lo que refleja que, conforme aumenta la renta, también aumenta la proporción de bienes consumidos mediante el uso de tarjetas (Tabla 1). También ocurre lo contrario en las regresiones para el gasto en efectivo, lo que revela que el gasto en efectivo disminuye con el nivel de ingresos. El efecto negativo sobre el consumo en efectivo es ligeramente mayor y más significativo que el que

se observa en las regresiones para el gasto mediante el uso de tarjetas. En términos cuantitativos, esto confirma las relaciones observadas en el análisis exploratorio de datos, donde la pendiente de efectivo es claramente negativa y significativa, mientras que los datos de las operaciones con tarjeta son menos claros y, de alguna forma, más planos.

La mayoría de las variables de control también son significativas, pero con signos diferentes en las regresiones para el gasto en efectivo y mediante el uso de tarjetas. El nivel de urbanización<sup>5</sup> (estas variables se incluyen con respecto a la zona rural) tiene un efecto positivo en el gasto con tarjeta, mientras que resulta negativo para el gasto en efectivo. La asimetría también se hace patente en las variables ficticias regionales del norte y el sur del país en relación con las zonas centrales. Existe una relación positiva entre el norte y el sur en relación con el centro en las regresiones para el gasto mediante el uso de tarjetas, mientras que el efecto en el caso del efectivo es tan significativo como negativo. Por último, la inclusión financiera tiene un efecto positivo tanto en el gasto en efectivo como mediante el uso de tarjetas, si bien es la menos significativa de todas las variables explicativas (Tablas 1 y 2).

Tabla 1. **RESULTADOS ESTIMADOS MEDIANTE EL USO DE UN MODELO LINEAL PARA EL GASTO CON TARJETA**

	Parámetro	Error estándar	Estado T	Valor P
Intersección	-0,0731	0,0302	-2,4253	0,0154*
Logaritmo (ingresos mensuales)	0,0071	0,0027	2,5725	0,0102*
Logaritmo (contratos de depósito por 10.000)	0,0021	0,0005	4,1337	0,0000***
Logaritmo (contratos de crédito por 10.000)	-0,0020	0,0006	-3,2251	0,0013***
Zona metropolitana	0,0178	0,0017	10,184	0,0000***
Zona urbana	0,0085	0,0017	5,0152	0,0000***
Zona norte	0,0052	0,0006	8,7008	0,0000***
Zona sur	0,0041	0,0007	5,9053	0,0000***

Resultados de la regresión del modelo de datos de panel lineal para la asignación presupuestaria de gastos con tarjeta ( $R^2 = 0,39$ ). Observamos una significativa elasticidad positiva de los ingresos en la demanda. Las variables de control (inclusión financiera, nivel de urbanización y ubicación geográfica) también tienen efectos significativos.

Fuente: cálculos propios. BBVA y Clarity

5: El nivel de urbanización se corresponde con las categorías de municipios rurales, en transición, semiurbanos, urbanos, semimetropolitanos y metropolitanos (fuente: Comisión Nacional Bancaria y de Valores de México).

Tabla 2. **RESULTADOS ESTIMADOS MEDIANTE EL USO DE UN MODELO LINEAL PARA EL GASTO EN EFECTIVO**

	Parámetro	Error estándar	Estado T	Valor P
Intersección	0,9288	0,0223	41,625	0,0000***
Logaritmo (ingresos mensuales)	-0,0907	0,0028	-32,843	0,0000***
Logaritmo (contratos de depósito por 10.000)	0,0024	0,0014	1,7246	0,0848
Logaritmo (contratos de crédito por 10.000)	0,0043	0,0018	2,3557	0,0186*
Zona metropolitana	-0,0056	0,0053	-1,0628	0,2880
Zona urbana	-0,0172	0,0051	-3,3542	0,0008***
Zona norte	-0,0017	0,0017	-1,0299	0,3032
Zona sur	-0,0125	0,0020	-6,3674	0,0000***

Resultados de la regresión del modelo de datos de panel lineal para la asignación presupuestaria de gastos en efectivo ( $R^2 = 0,53$ ). Observamos una significativa elasticidad negativa de los ingresos en la demanda. Las variables de control (inclusión financiera, nivel de urbanización y ubicación geográfica) también tienen efectos significativos.

Fuente: cálculos propios. BBVA y Clarity

### 3.3 Clasificación de los bienes normales y superiores mediante la semielasticidad de los ingresos

La teoría económica considera que un bien normal es aquel producto o servicio que los consumidores adquieren independientemente de su nivel de ingresos. Su elasticidad en proporción al ingreso es positiva, si bien inferior a la de los bienes superiores o de lujo, por lo que la proporción de gasto dedicada a estos bienes normales disminuye conforme aumenta la renta: su semielasticidad en relación con la demanda es negativa<sup>6</sup>. Por el contrario, los bienes superiores tienen una elasticidad positiva mayor en proporción a los ingresos que los bienes normales, ya que el consumo aumenta más que en el caso de estos últimos conforme sube el nivel de ingresos. A diferencia de los bienes normales, su semielasticidad en proporción al ingreso es positiva, ya que la asignación presupuestaria correspondiente a estos bienes aumenta cuando lo hace la renta.

La estimación de la semielasticidad en proporción al ingreso y en relación con la demanda permite identificar estos bienes normales y superiores. Gracias a la alta granularidad de los datos, podemos estimar el modelo por nivel de categoría de producto e identificar bienes normales y superiores en función del valor del coeficiente que representa la semielasticidad de la demanda de los distintos bienes. Encontramos que la mayoría de los bienes clasificados como «normales» (coeficiente de semielasticidad negativo) se pagaron en efectivo (Tabla 3). Entre estos podemos encontrar bienes como los productos de alimentación, las bebidas alcohólicas y el tabaco. Por otro lado, la mayoría de los bienes clasificados como «superiores» (coeficiente de semielasticidad positivo) se pagaron usando tarjetas. Los valores más elevados se registraron para los bienes relacionados con viajes, vestuario, restaurantes y educación.

6: Deaton, A.; Muellbauer, J.; otros. Economics and Consumer Behavior; Cambridge University Press: Nueva York, NY, 1980.

Tabla 3. SEMIELASTICIDAD DE LA DEMANDA DE BIENES NORMALES Y SUPERIORES

Bienes superiores				Bienes normales			
Categoría de consumo	Origen	Elasticidad de los ingresos	Valor P	Categoría de consumo	Origen	Elasticidad de los ingresos	Valor P
Otros	Tarjeta	0,4968	0.0000E+00	Hogar	Tarjeta	-0,0182	4.7010E-02
Viajes	Tarjeta	0,2675	1.7292E-09	Edificaciones	Tarjeta	-0,0591	6.6576E-08
Vestuario	Tarjeta	0,1767	0.0000E+00	Vestuario	Efectivo	-0,1139	4.9960E-13
Restaurantes	Tarjeta	0,1505	0.0000E+00	Equipos de transporte	Efectivo	-0,1757	0.0000E+00
Educación	Tarjeta	0,1011	2.7325E-08	Mantenimiento y reparaciones en vivienda	Efectivo	-0,5071	0.0000E+00
Otros servicios	Efectivo	0,0692	6.1209E-05	Alimentación	Efectivo	-0,622	0.0000E+00
Mascotas	Tarjeta	0,0285	5.5511E-15	Cuidado y efectos personales	Efectivo	-0,6622	0.0000E+00
Oficinas	Tarjeta	0,0275	2.3891E-04	Bebidas no alcohólicas	Efectivo	-0,6646	0.0000E+00
Calzado	Tarjeta	0,0206	1.9322E-03	Hogar, jardinería y mascotas	Efectivo	-0,6709	0.0000E+00
Productos de belleza	Tarjeta	0,0206	6.6613E-16	Alimentación	Tarjeta	-0,7721	0.0000E+00
Música	Tarjeta	0,0139	2.9060E-04	Bebidas alcohólicas	Efectivo	-1,5388	0.0000E+00
Joyería	Tarjeta	0,0087	1.4360E-09	Tabaco	Efectivo	-1,8128	0.0000E+00
Retiradas de efectivo	Tarjeta	0,0037	1.7764E-15				
Libros	Tarjeta	0,0033	6.3647E-04				

Categorías de consumo clasificadas como bienes superiores con una semielasticidad positiva de los ingresos en relación con la demanda. Estas categorías suponen un 37% de todas las categorías y la inmensa mayoría se corresponden con pagos con tarjeta. Fuente: cálculos propios. Categorías de consumo clasificadas como bienes normales con una semielasticidad negativa de los ingresos en relación con la demanda. Estas categorías suponen un 32% de todas las categorías y la inmensa mayoría se corresponden con pagos en efectivo. Fuente: BBVA y Clarity

## 4. Más allá de la linealidad para testar la relación no lineal de los datos

### 4.1 Estimación de modelos de aprendizaje automático

Los modelos lineales cuentan con la ventaja de la interpretabilidad. Sin embargo, la relación entre la asignación presupuestaria, los ingresos o la renta, y el resto de las variables de control podría ser no lineal y más compleja. Con el objetivo de poner a prueba esta hipótesis, ajustamos un conjunto de modelos que incluyen regresión lineal (regresión de mínimos cuadrados ordinarios u OLS, por sus siglas en inglés), árboles de decisiones, «Random Forest», «Gradient Boosting», «Support Vector Machines» y un modelo «Logit» para el que efectuamos la

binarización del gasto en efectivo y mediante el uso de tarjetas, equiparandolo a uno si la asignación presupuestaria es mayor que el promedio, o bien a cero en caso contrario.

Formamos los modelos usando el 80% de los datos y seleccionamos el mejor basado en errores de validación cruzada de cinco pliegues usando un esquema estándar de ajuste de hiper-parámetros. Utilizamos el 20% restante de los datos para probar el modelo y estimar el error de generalización. El mejor modelo seleccionado a partir del conjunto de formación se utiliza para predecir las asignaciones presupuestarias invisibles en el conjunto de prueba. Hacemos esto para cada uno de los conjuntos de datos sobre uso de efectivo y tarjetas, y registramos la precisión en el caso del modelo «Logit» (la proporción de pronósticos correctos en relación con el total de modelos alternativos) y el rendimiento ( $R^2$ ) de los modelos en sus respectivas muestras de prueba en el caso de los modelos de aprendizaje automático.

Los resultados de la Tabla 4 muestran el rendimiento de los distintos modelos. El primer resultado importante es que todos los modelos (lo que incluye el modelo «Logit», los árboles de decisiones, «Gradient Boosting» y «Random Forest») ofrecen mejores resultados que la referencia lineal. Esto ya es indicativo de que las especificaciones no lineales pueden mejorar nuestro modelo lineal. Una especificación tradicional en términos del modelo «Logit» (0,78 para tarjeta y 0,82 para efectivo) también goza de buena precisión.

Tabla 4. **PUNTUACIÓN DE RENDIMIENTO/PRECISIÓN DE LOS MODELOS PARA GASTO CON TARJETA Y EN EFECTIVO**

	«Random Forest»	«Gradient Boosting»	Árboles de decisión	Modelo «Logit»	Mínimos cuadrados ordinarios (OLS)
Gasto con tarjeta	0,93	0,83	0,70	0,78	0,39
Gasto en efectivo	0,92	0,78	0,75	0,82	0,53

Fuente: cálculos propios. BBVA y Clarity

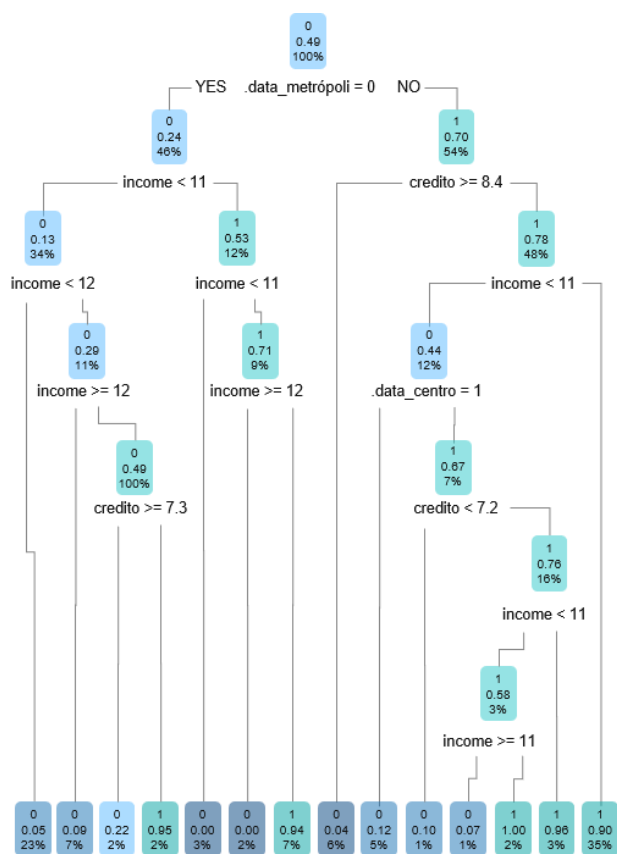
El mejor regresor da como resultado un «Random Forest», un modelo no lineal basado en la agregación de múltiples árboles de decisiones<sup>7</sup>. Sorprendentemente, su rendimiento es aproximadamente un 100% mejor que el de las referencias lineales, con lo que obtiene un  $R^2$  de 0,93 y 0,92 para los modelos de gasto con tarjeta y en efectivo, respectivamente. No cabe duda de que esto apunta a la existencia de no linealidades significativas.<sup>8</sup>

7: Friedman, J.; Hastie, T. y Tibshirani, R. (2001). The Elements of Statistical Learning (Vol. 1, N.º 10). Nueva York: Colección sobre Estadística de Springer

8: El modelo de bosques aleatorios («Random Forest») trata diversas relaciones no lineales difíciles de analizar usando enfoques tradicionales. Genera una gran cantidad de árboles de regresión, cada uno de ellos calibrado en una muestra de «bootstrap» de los datos. Cada nodo se divide mediante un subconjunto de predictores seleccionados al azar. A fin de predecir el valor de un nuevo punto de datos, los datos se ejecutan en cada uno de los árboles del bosque y cada árbol proporciona un valor. La predicción del modelo se calcula a continuación como el valor promedio sobre las predicciones de todos los árboles del bosque.

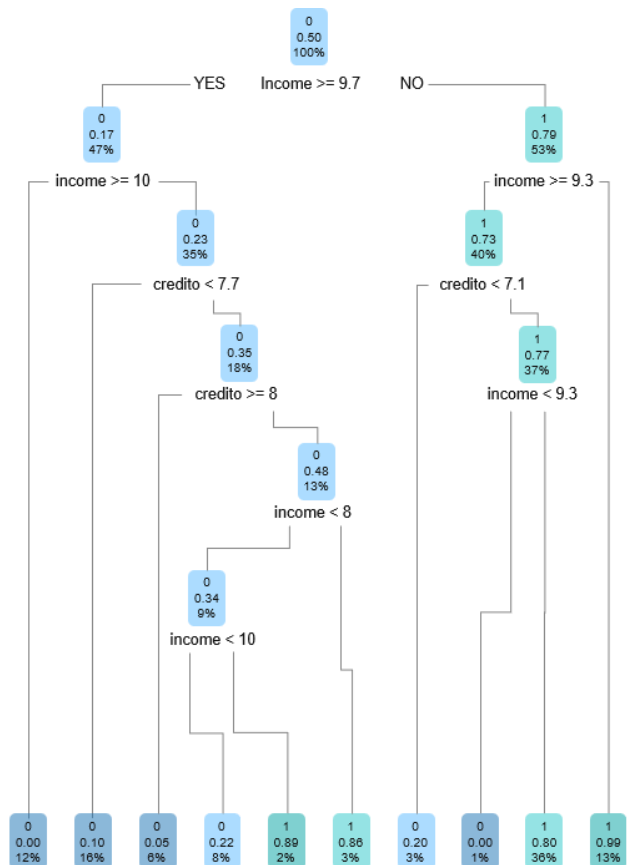
Así pues, con el propósito de obtener información sobre el tipo de no linealidades presentes en los datos, pasamos ahora a analizar un único árbol de decisiones. Ha de considerarse que, pese a que su rendimiento es menor que el del «Random Forest» (lo cual es habitual), sigue siendo mucho mayor que el de las referencias lineales<sup>9</sup>. El árbol de gasto con tarjeta (Gráfica 7) revela que la residencia en una zona metropolitana, la inclusión financiera y un elevado nivel de ingresos son las variables más importantes. Como se ha observado, el modelo es capaz de capturar las no linealidades en los niveles más altos de intensificación financiera (rama derecha del árbol) entre crédito y renta (niveles inferiores del árbol). En lo que concierne al gasto en efectivo (Gráfica 8), la variable más discriminante es el ingreso, tal como muestra el árbol de decisiones. En este caso, las relaciones no lineales se encuentran en los niveles de renta más bajos (ramas izquierdas del árbol) al interactuar con la intensificación financiera.

Gráfica 7. **ÁRBOL DE DECISIONES: GASTO CON TARJETA**



Fuente: cálculos propios. BBVA y Clarity

Gráfica 8. **ÁRBOL DE DECISIONES: GASTO EN EFECTIVO**



Fuente: cálculos propios. BBVA y Clarity

9: Debemos tener en cuenta que estos árboles de decisiones son tan solo un ejemplo ilustrativo de las relaciones entre las variables, pero los resultados finales del enfoque de bosques aleatorios («Random Forest») pueden variar, ya que se consideran múltiples combinaciones de árboles de decisiones. No obstante, encontramos que, aunque estos árboles solo tienen una profundidad de 6, representan el 80% (0,75 R<sup>2</sup>) y el 90% (0,86 R<sup>2</sup>) de la variabilidad explicada por los mejores modelos de tarjeta y efectivo, respectivamente, por lo que la información relativa a la interpretación derivada de este ejemplo se aproxima al resultado final.

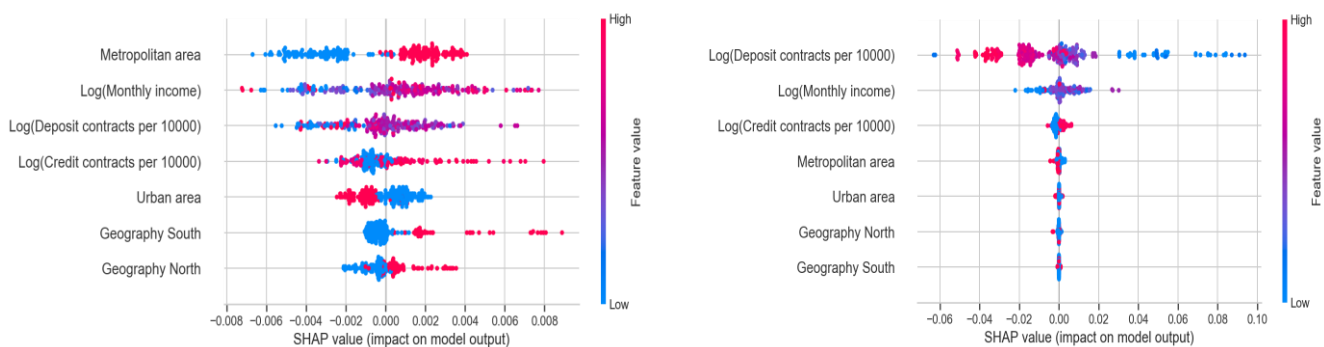
## 4.2 Identificación de valores de aprendizaje automático mediante los valores de Shapley

Si bien los algoritmos de aprendizaje automático resultan más adecuados a la hora de capturar la relación no lineal entre las variables, también tienen ciertas limitaciones. En particular, podría resultar difícil cuantificar la relación entre las variables, una de las principales razones por las que se los suele denominar modelos de «caja negra».

Para poder explicar mejor estos modelos, los valores SHAP<sup>10</sup> (inspirados en el concepto de la teoría de juegos de los valores de Shapley) asignan la contribución justa de las variables explicativas a una predicción mediante el promedio de sus contribuciones marginales. Es importante destacar que los valores SHAP capturan los factores que impulsan hacia arriba o hacia abajo una estimación determinada con respecto a una referencia. La Gráfica 9 muestra el valor SHAP (eje X) de cualquiera de las variables explicativas (eje Y). El eje X corresponde a los gastos en efectivo o con tarjeta.

Las variables se ordenan según su relevancia y el color representa cómo los valores altos (rojo) y bajos (azul) de la variable inciden en el modelo. En el caso de las operaciones con tarjeta (gráfica izquierda), los valores más altos de ingresos (en color rojo) se asocian con valores negativos de la variable dependiente (la asignación del gasto), lo que cuadra con una semielasticidad negativa. Mientras tanto, lo contrario ocurre en el modelo de gasto con tarjeta, donde la semielasticidad es positiva, pero menos clara, ya que encontramos un leve efecto positivo para la renta. En pocas palabras, vemos que existen pruebas sólidas de una relación no lineal entre los ingresos y la asignación presupuestaria capturada mediante técnicas de aprendizaje automático. Estos efectos son direccionalmente consistentes con sus contrapartes lineales, si bien el enfoque lineal arrojó resultados más moderados.

Gráfica 9. **DIAGRAMA RESUMEN DEL VALOR SHAP Y EL USO DE UN MODELO NO LINEAL PARA EL GASTO CON TARJETA Y EN EFECTIVO**



Fuente: cálculos propios. BBVA y Clarity

Una propiedad destacable de los valores SHAP es que son aditivos, en el sentido de que las contribuciones independientes de las variables son iguales a la variación total de la variable dependiente. Las Gráficas 10 y 11, que presentan la contribución de las variables para los modelos alternativos probados en este ejercicio, muestran

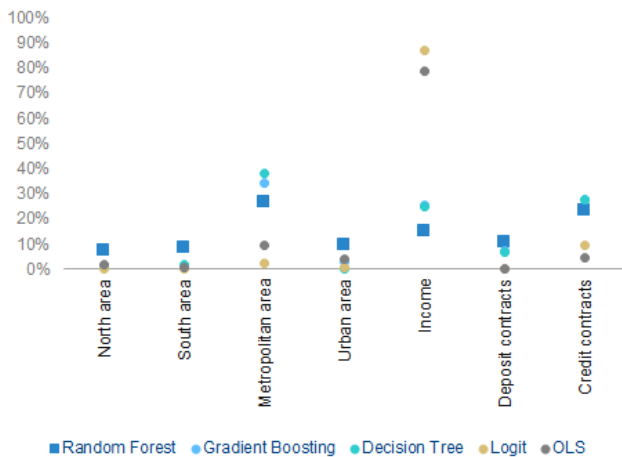
10: Lundberg, Scott y Su-In Lee. «A Unified Approach to Interpreting Model Predictions». Prepublicación de arXiv: 1705.07874 (2017).



una diferencia importante entre los modelos lineales y no lineales (la Tabla A1 que contiene los datos puede encontrarse en el Anexo 2).

Particularmente, en el caso del gasto con tarjeta, mientras que los modelos lineales y el modelo «Logit» muestran la relevancia de la variable del ingreso, los algoritmos no lineales capturan la influencia de vivir en zonas metropolitanas y las variables representativas de la intensificación financiera. En lo que respecta al gasto en efectivo, las diferencias entre los modelos de aproximación lineal y aprendizaje automático son igualmente significativas (Gráficas 12 y 13). Sin embargo, la relevancia del ingreso se mantiene tanto en aproximaciones lineales como en las alternativas de aprendizaje automático (aunque en menor grado), mientras que la inclusión financiera (contratos de crédito) y, en menor medida, la ubicación geográfica siguen siendo relevantes, pero menores que en el caso de las operaciones en efectivo (la Tabla A2 que contiene los datos puede encontrarse en el Anexo 2).

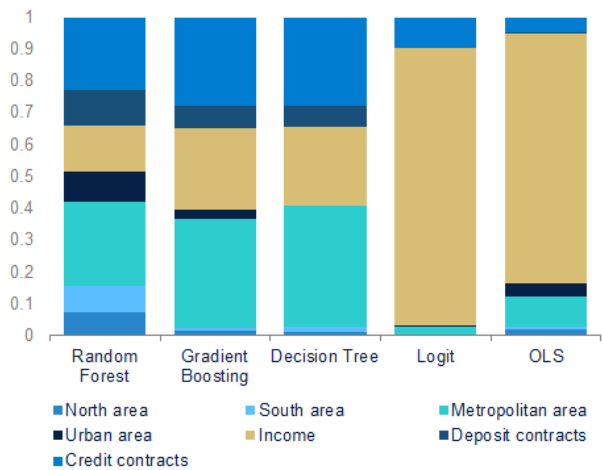
Gráfica 10. **COEFICIENTES PROPORCIONALES DE SHAP PARA MODELOS DE GASTO CON TARJETA POR VARIABLE**



Los modelos aparecen ordenados según su rendimiento, yendo del mejor (bosques aleatorios o «Random Forest») al peor (mínimos cuadrados ordinarios u OLS).

Fuente: cálculos propios. BBVA y Clarity

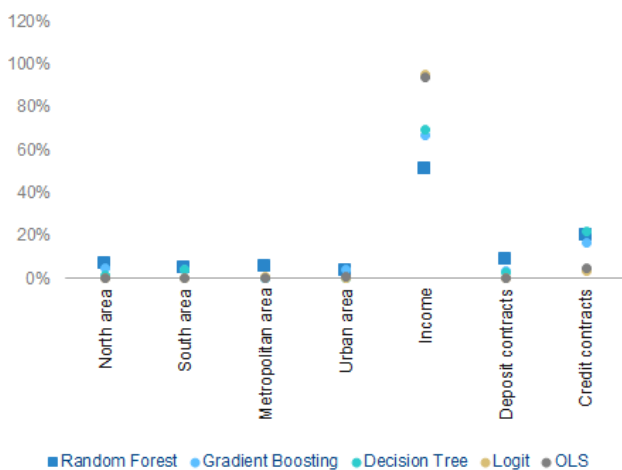
Gráfica 11. **COEFICIENTES PROPORCIONALES DE SHAP PARA MODELOS DE GASTO CON TARJETA POR MODELO**



BBVA y Clarity. Los modelos aparecen ordenados según su rendimiento, yendo del mejor (bosques aleatorios o «Random Forest») al peor (mínimos cuadrados ordinarios u OLS).

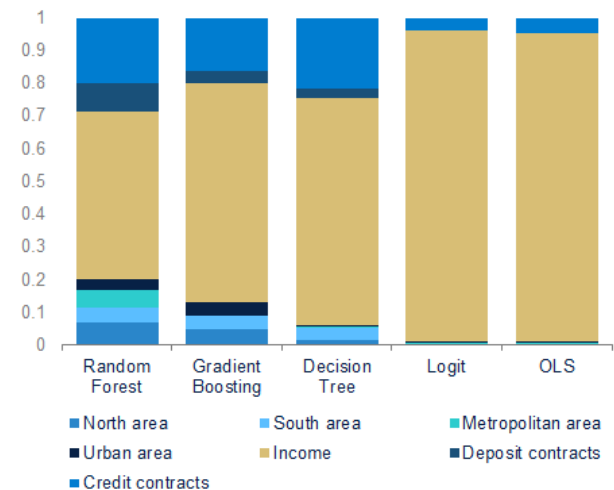
Fuente: cálculos propios. BBVA y Clarity

Gráfica 12. **COEFICIENTES PROPORCIONALES DE SHAP PARA MODELOS DE GASTO EN EFECTIVO POR VARIABLE**



BBVA y Clarity. Los modelos aparecen ordenados según su rendimiento, yendo del mejor (bosques aleatorios o «Random Forest») al peor (mínimos cuadrados ordinarios u OLS).  
Fuente: cálculos propios. BBVA y Clarity

Gráfica 13. **COEFICIENTES PROPORCIONALES DE SHAP PARA MODELOS DE GASTO EN EFECTIVO POR MODELO**



BBVA y Clarity. Los modelos aparecen ordenados según su rendimiento, yendo del mejor (bosques aleatorios o «Random Forest») al peor (mínimos cuadrados ordinarios u OLS).  
Fuente: cálculos propios. BBVA y Clarity

## Conclusiones

Los datos de consumo resultan cruciales para el bienestar social, la elaboración de políticas y el análisis económico. Sin embargo, los datos recopilados mediante las encuestas tradicionales adolecen de ciertas deficiencias sobradamente conocidas. En el presente documento, hemos propuesto un enfoque alternativo que emplea Big Data para rastrear el consumo combinando dos conjuntos de datos complementarios: uno que registra operaciones en efectivo y otro que hace lo propio con las transacciones con tarjeta y los pagos electrónicos. El conjunto de datos concerniente al consumo de efectivo proviene de recibos de TPV digitalizados, mientras que el que se emplea en relación con el consumo mediante tarjetas se deriva de datos transaccionales. Ambos conjuntos de datos proporcionan información valiosa para el análisis gracias a su alta granularidad temporal, geográfica y a nivel de producto.

Nuestro análisis de datos apunta a la existencia de diferencias significativas entre municipios pertenecientes a un mismo estado, vinculadas al nivel de renta y al tamaño de la economía informal en la región. Los municipios más pobres y con mayor densidad de población son los que más usan el efectivo como medio de pago o los que registran el mayor aumento en la asignación presupuestaria en efectivo. Por su parte, los municipios más ricos registran un menor gasto en efectivo y un mayor gasto con tarjeta.

El uso combinado de datos de operaciones con tarjeta y en efectivo ha demostrado ser una herramienta eficaz para caracterizar las necesidades de consumo. Nuestros análisis revelan que los ingresos tienen efectos lineales y no lineales significativos sobre el gasto en efectivo y con tarjeta. Cuanto mayor sea la renta, mayor será el grado

de bancarización y menor el gasto en efectivo, incluso controlando la inclusión financiera, el nivel de urbanización y la ubicación geográfica. Además, pudimos identificar las categorías de consumo que se conciben como bienes de primera necesidad o bienes superiores, y los métodos de pago preferidos para cada categoría. Como era de esperar, la mayoría de los bienes de primera necesidad detectados se pagan en efectivo. Esto cuadra con el hecho de que los supermercados son quienes venden este tipo de productos en nuestra muestra. Esta técnica de discriminación entre los dos tipos de consumo puede ser de gran valor, por ejemplo, a la hora de elaborar políticas de protección de sectores concretos.

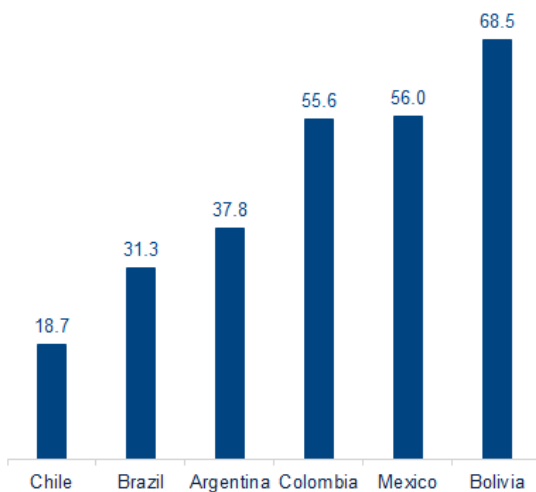
Asimismo, este estudio muestra la importancia del empleo del Big Data y las técnicas de ciencia de datos para capturar relaciones no lineales entre variables, mediante la obtención de información complementaria a enfoques más tradicionales. El análisis de datos de pago con tarjeta y en efectivo muestra diferentes patrones de consumo en la población mexicana según los niveles de ingresos, las zonas urbanas y rurales, y los niveles de inclusión financiera. No obstante, los modelos de aprendizaje automático muestran que los patrones de consumo podrían ser diferentes según las operaciones de pago. En particular, la influencia de los ingresos parece ser más importante en el caso de las transacciones en efectivo, mientras que los efectos de la urbanización y la inclusión financiera quedan fielmente capturados gracias a los modelos de aprendizaje automático.

## Anexo 1: Informalidad en la economía Mexicana

En los últimos años, el Banco de México ha puesto en marcha iniciativas de políticas públicas orientadas a fomentar el uso del dinero digital o los pagos electrónicos con el objetivo de reemplazar las operaciones en efectivo en la economía del país. Uno de los programas puestos en marcha como parte de estas iniciativas es CoDi, una plataforma de pago a través de teléfonos móviles que usa la tecnología de los códigos QR. La primera fase de CoDi dio comienzo el 30 de septiembre de 2019; supuso la adopción tecnológica de la plataforma por parte de la banca múltiple y la difusión de información entre la población para los usuarios que se iban inscribiendo en esta propuesta (Banxico, 2020). A diciembre de 2020, había 7,4 millones de cuentas validadas a través del programa, de las cuales BBVA validó un 64% (CoDi, 2021).

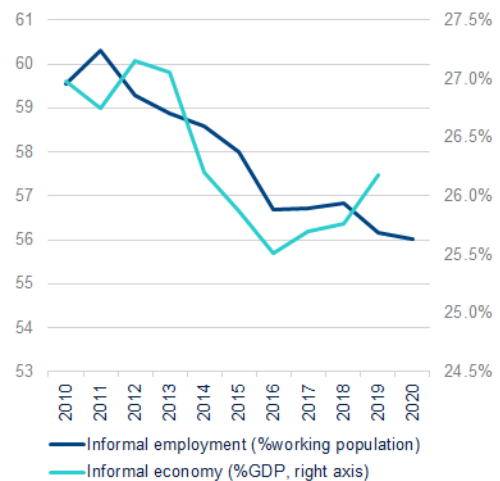
Los programas como CoDi cobran protagonismo en una economía como la de México, donde predomina el uso de efectivo y en la que 56 de cada 100 trabajadores pertenecen al mercado laboral informal. Esta elevada cifra coloca a México detrás de Chile, Argentina y Colombia en términos de tasa de informalidad económica (Gráfica A1). Según datos oficiales, en 2019 la economía informal del país llegó a representar el 26,2% del PIB (Gráfica A2). Aunque la tasa de informalidad económica ha disminuido durante la última década (del 59,5% en 2010 al 56,0% en 2020, en proporción a la población activa), ha mostrado un rendimiento mixto a lo largo de esos 10 años. Vale la pena prestar atención al periodo 2016 – 2018, ya que muestra un estancamiento en la tendencia a la baja de la informalidad económica, coincidiendo con la tendencia al alza del tamaño de la economía informal durante esos mismos años.

Gráfica A1. **INFORMALIDAD ECONÓMICA (% DE LA POBLACIÓN ACTIVA)**



Fuente: INEGI

Gráfica A2. **INFORMALIDAD LABORAL Y ECONOMÍA INFORMAL (% DE LA POBLACIÓN ACTIVA Y % DEL PIB, RESPECTIVAMENTE)**

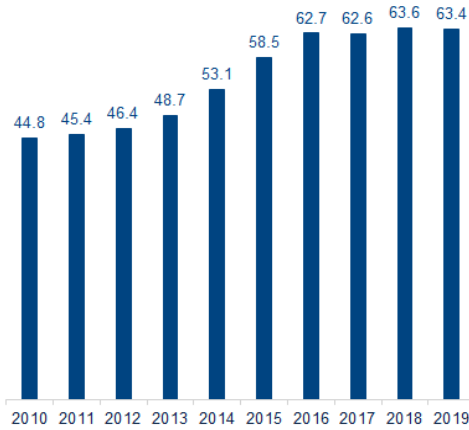


Fuente: INEGI

Junto con la intensificación de la economía informal durante el periodo 2016 – 2018, la cantidad de billetes y monedas en circulación llegó a suponer el 62,7% del PIB (desde el 58,5% registrado en 2015), lo que indica un mayor uso de efectivo. Según la Encuesta Nacional de Inclusión Financiera (ENIF) 2018, más del 90% de las transacciones comerciales (venta minorista, alquileres, transporte público, servicios públicos y privados) se abonaron en efectivo durante ese año. Los mismos datos revelan que casi el 95% de las compras por un importe

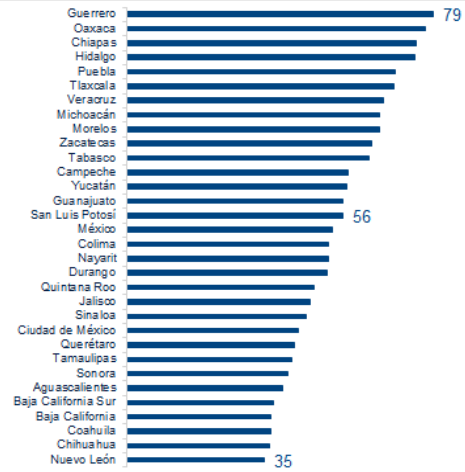
de 500 MXN o menos (25 USD o menos) se realizaron con pagos en efectivo, mientras que un 87% de las compras por un importe de 501 MXN o más emplearon el efectivo como método de pago principal.

Gráfica A3. **BILLETES Y MONEDAS EN CIRCULACIÓN (% DEL PIB)**



Fuente: Banxico

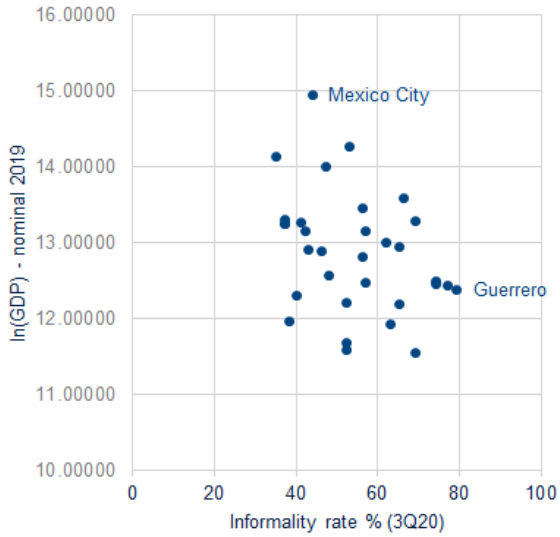
Gráfica A4. **INFORMALIDAD ECONÓMICA POR ESTADO (% DE LA POBLACIÓN ACTIVA)**



Fuente: INEGI

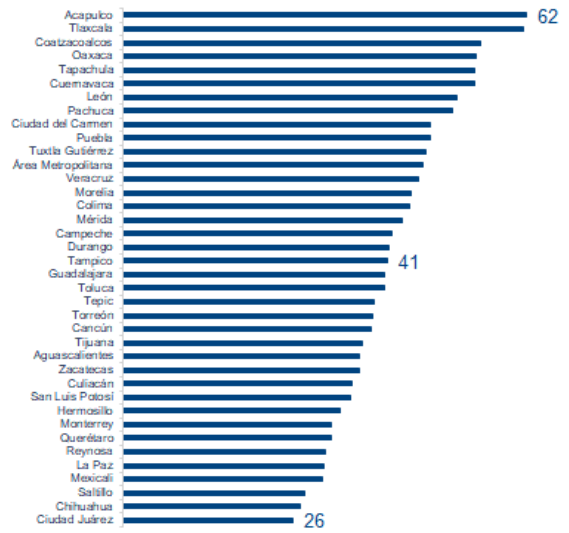
Los estados con menor renta muestran las mayores tasas de informalidad económica del país. El estado de Guerrero informa de que el 79% de su población activa trabaja en régimen de informalidad laboral, seguido de Oaxaca (77%) y Chiapas (74%). En el extremo opuesto se sitúan Nuevo León y Chihuahua, con tasas de informalidad económica del 35% y del 37%, respectivamente. Puede observarse un patrón geográfico idéntico entre las principales ciudades del país, siendo los centros urbanos con menores ingresos los que registran los niveles más elevados de informalidad económica (Acapulco: 62%; Tlaxcala: 61%). En general, la tasa promedio de informalidad económica en las principales ciudades del país es del 41%, 13 p.p. por debajo de la cifra nacional (54%), lo que apunta a que la informalidad económica se concentra en zonas menos urbanizadas.

Gráfica A5. **INFORMALIDAD ECONÓMICA VS. PIB POR ESTADO (% Y LOGARITMO NATURAL, RESPECTIVAMENTE)**



Fuente: INEGI

Gráfica A6. **INFORMALIDAD ECONÓMICA POR PRINCIPALES CIUDADES (% DE LA POBLACIÓN ACTIVA)**



Fuente: INEGI

## Anexo 2: Coeficientes proporcionales de SHAP

Tabla A1. **COEFICIENTES PROPORCIONALES DE SHAP PARA MODELOS DE GASTO CON TARJETA**

	Bosques aleatorios («Random Forest»)	Potenciación del gradiente	Árboles de decisiones	Modelo «Logit»	Mínimos cuadrados ordinarios (OLS)
Zona norte	0,0749	0,0142	0,0128	0,002	0,019
Zona sur	0,082	0,0102	0,0158	0,0001	0,0074
Zona metropolitana	0,265	0,341	0,3792	0,0238	0,0969
Zona urbana	0,0925	0,0296	0	0,006	0,0419
Ingresos	0,1473	0,257	0,2469	0,8703	0,7858
Contratos de depósito	0,1082	0,0705	0,0686	0,0009	0,0024
Contratos de crédito	0,2301	0,2776	0,2767	0,0969	0,0467

Fuente: cálculos propios. BBVA y Clarity

Tabla A2. **COEFICIENTES PROPORCIONALES DE SHAP PARA MODELOS DE GASTO EN EFECTIVO**

	Bosques aleatorios («Random Forest»)	Potenciación del gradiente	Árboles de decisiones	Modelo «Logit»	Mínimos cuadrados ordinarios (OLS)
Zona norte	0,0704	0,0474	0,0149	0,0022	0,0008
Zona sur	0,0448	0,0404	0,0394	0,0001	0,0023
Zona metropolitana	0,0534	0,0031	0,0024	0,0065	0,003
Zona urbana	0,0316	0,0393	0,0022	0,001	0,0066
Ingresos	0,5118	0,6685	0,6941	0,9529	0,9392
Contratos de depósito	0,0886	0,0372	0,0305	0,0002	0,0026
Contratos de crédito	0,1994	0,1641	0,2165	0,037	0,0456

Fuente: cálculos propios. BBVA y Clarity

## Documentos de Trabajo

### 2021

21/05 **Jaime Oliver, Álvaro Ortiz, Tomasa Rodrigo Saidé Salazar e Ignacio Tamarit:**

**ESP/** Patrones de Consumo de Efectivo vs Tarjeta en México: una aproximación Big Data.

**ING /** Cash Vs Card Consumption Patterns in Mexico: A Machine Learning Approach.

21/04 **Ángel de la Fuente:** La financiación autonómica en 2020: una primera aproximación y una propuesta de cara a 2021.

21/03 **Ángel de la Fuente:** Las finanzas autonómicas en 2020 y entre 2003 y 2020.

21/02 **Joxe Mari Barrutiabengoa, J. Julián Cubero and Rodolfo Méndez-Marcano:** Output-side GHG Emissions Intensity: A consistent international indicator.

21/01 **Ángel de la Fuente y Pep Ruiz:** Series largas de VAB y empleo regional por sectores, 1955-2019 Actualización de *RegData-Sect* hasta 2019.

### 2020

20/17 **Amparo Castelló-Climent and R. Doménech:** Human Capital and Income Inequality Revisited.

20/16 **J.E. Boscá, R. Doménech, J. Ferri, J.R. García and C. Ulloa:** The Stabilizing Effects of Economic Policies in Spain in Times of COVID-19.

20/15 **Ángel de la Fuente:** La evolución de la financiación de las comunidades autónomas de régimen común, 2002-2018.

20/14 **Ángel de la Fuente:** El impacto de la crisis del Covid sobre el PIB de las CCAA en 2020: una primera aproximación.

20/13 **Ali B. Barlas, Seda Guler Mert, Álvaro Ortiz and Tomasa Rodrigo:** Investment in Real Time and High Definition: A Big Data Approach.

20/12 **Félix Lores, Pep Ruiz, Angie Suárez y Alfonso Ugarte:** Modelo de precios de la vivienda en España. Una perspectiva regional.

20/11 **Ángel de la Fuente:** Series largas de algunos agregados económicos y demográficos regionales: Actualización de *RegData* hasta 2019.

20/10 **Ángel de la Fuente:** La liquidación de 2018 del sistema de financiación de las comunidades autónomas de régimen común.

20/09 **Lucía Pacheco Rodríguez and Pablo Urbiola Ortún:** From FinTech to BigTech: an evolving regulatory response.

20/08 **Federico D. Forte:** Network Topology of the Argentine Interbank Money Market.

20/07 **Ángel de la Fuente:** Las finanzas autonómicas en 2019 y entre 2003 y 2019.



20/06 **Vasco M. Carvalho, Juan R. Garcia, Stephen Hansen, Álvaro Ortiz, Tomasa Rodrigo, José V. Rodríguez Mora and Pep Ruiz:** Tracking the COVID-19 Crisis with High-Resolution Transaction Data.

20/05 **Jinyue Dong and Le Xia:** Forecasting modeling for China's inflation.

20/04 **Le Xia:** Lessons from China's past banking bailouts.

20/03 **Ángel de la Fuente y Pep Ruiz:** Series largas de VAB y empleo regional por sectores, 1955-2018. RegData\_Sect FEDEA-BBVA (v5.0\_1955-2018).

20/02 **Luis Antonio Espinosa y Juan José Li Ng:**

**ESP/** El riesgo del sargazo para la economía y turismo de Quintana Roo y México.

**ING /** The risk of sargassum to the economy and tourism of Quintana Roo and Mexico.

20/01 **Ángel de la Fuente:** La dinámica territorial de la renta en España, 1955-2018. Los determinantes directos de la renta relativa: productividad, ocupación y demografía.

**HAGA CLIC PARA ACCEDER A LOS DOCUMENTOS DE TRABAJO PUBLICADOS EN**  
Español e Inglés

## AVISO LEGAL

El presente documento no constituye una "Recomendación de Inversión" según lo definido en el artículo 3.1 (34) y (35) del Reglamento (UE) 596/2014 del Parlamento Europeo y del Consejo sobre abuso de mercado ("MAR"). En particular, el presente documento no constituye un "Informe de Inversiones" ni una "Comunicación Publicitaria" a los efectos del artículo 36 del Reglamento Delegado (UE) 2017/565 de la Comisión de 25 de abril de 2016 por el que se completa la Directiva 2014/65/UE del Parlamento Europeo y del Consejo en lo relativo a los requisitos organizativos y las condiciones de funcionamiento de las empresas de servicios de inversión ("MiFID II").

Los lectores deben ser conscientes de que en ningún caso deben tomar este documento como base para tomar sus decisiones de inversión y que las personas o entidades que potencialmente les puedan ofrecer productos de inversión serán las obligadas legalmente a proporcionarles toda la información que necesiten para esta toma de decisión.

El presente documento, elaborado por el Departamento de BBVA Research, tiene carácter divulgativo y contiene datos u opiniones referidas a la fecha del mismo, de elaboración propia o procedentes o basadas en fuentes que consideramos fiables, sin que hayan sido objeto de verificación independiente por BBVA. BBVA, por tanto, no ofrece garantía, expresa o implícita, en cuanto a su precisión, integridad o corrección.

El contenido de este documento está sujeto a cambios sin previo aviso en función, por ejemplo, del contexto económico o las fluctuaciones del mercado. BBVA no asume compromiso alguno de actualizar dicho contenido o comunicar esos cambios.

BBVA no asume responsabilidad alguna por cualquier pérdida, directa o indirecta, que pudiera resultar del uso de este documento o de su contenido.

Ni el presente documento, ni su contenido, constituyen una oferta, invitación o solicitud para adquirir, desinvertir u obtener interés alguno en activos o instrumentos financieros, ni pueden servir de base para ningún contrato, compromiso o decisión de ningún tipo.

El contenido del presente documento está protegido por la legislación de propiedad intelectual. Queda expresamente prohibida su reproducción, transformación, distribución, comunicación pública, puesta a disposición, extracción, reutilización, reenvío o la utilización de cualquier naturaleza, por cualquier medio o procedimiento, salvo en los casos en que esté legalmente permitido o sea autorizado expresamente por BBVA.

### INTERESADOS DIRIGIRSE A:

BBVA Research: Calle Azul, 4. Edificio La Vela – 4ª y 5ª planta. 28050 Madrid (España).  
Tel.: +34 91 374 60 00 y +34 91 537 70 00 / Fax: +34 91 374 30 25  
bbvaresearch@bbva.com www.bbvaresearch.com

