

## CAPÍTULO VII

## Transacciones financieras y economía en tiempo real y alta definición: el consumo español

Alvaro Ortiz Vidal-Abarca\*  
Tomaso Rodrigo

Las transacciones financieras que ocurren de manera natural en la vida diaria constituyen una rica fuente de información para aproximar el comportamiento de una economía en tiempo real y en alta definición. La información anonimizada, procesada y agregada adecuadamente nos permite aproximar las cuentas nacionales tanto por el lado de la demanda, de la oferta y de las rentas a una frecuencia mucho más elevada a la oficial y con un grado de detalle muy superior al incluido normalmente en las encuestas. En este capítulo presentamos cómo utilizar los datos de transacciones financieras para el análisis del consumo español. El uso del *big data* nos ayuda a complementar a las estadísticas oficiales y abre una ventana de oportunidades en el diseño y ejecución de políticas inteligentes.

*Palabras clave:* *big data*, consumo, cuentas nacionales.

---

\* Agradecemos los comentarios y sugerencias de Manuel Arellano, Scott Baker, V.V. Chari, Giancarlo Corsetti, Diane Coyle, Lukas Freund, Fatih Guvenen, Kyle Herkenhoff, Christopher Kurz, Francois Lafond, Rocio Madera, Clara Martínez Toledano, Mariacristina De Nardi, Fabrizio Perri, Christopher Rauh, Amir Sufi, Gianluca Violante, Michael E. Waugh y Nicolas Woloszko. También damos las gracias a los participantes en seminarios y conferencias del Banco de Canadá; el Banco de Italia; el CEPR Workshop on New Consumption Data; la OCDE; la 2022 European Micro-Macro-Midwest Conference; la Oficina de Estadísticas Nacionales del Reino Unido; el INE; el Banco de la Reserva Federal de Minneapolis; Banco de España; Riksbank-Federal Reserve Board-Bank of Canada-Bank of Italy Conference on Non-traditional Data, Machine Learning and Natural Language Processing in Macroeconomics; y US National Academies of Sciences, Engineering and Medicine's Panel on An Integrated System of U. S. Household Income, Wealth, and Consumption Data and Statistics to Inform Policy and Research. Agradecemos el apoyo y ayuda de Jose Miguel Leiva y Sirenia Vazquez en el proyecto y en general al equipo de BBVA Research por sus comentarios y sugerencias.

## 1. INTRODUCCIÓN

Desde hace décadas la inteligencia artificial ha venido desarrollándose de manera vertiginosa. La aceleración de esta tendencia no es casual, y es el resultado de la convergencia en el tiempo de avances paralelos en diferentes frentes. El primero de ellos, y quizás el más obvio, es el aumento de la capacidad de procesamiento de la información cuya velocidad es exponencial y vertiginosa. Esto ha reforzado un segundo avance, el rápido desarrollo de nuevos algoritmos y modelos que nos permiten día a día mejorar las reglas de decisión, acceder a nuevos datos y elaborar modelos para tratar fenómenos cada vez más complejos. Por último, y no menos importante, las dos han permitido un aumento rápido tanto de la digitalización de la información como de su procesamiento o conversión en información estructurada.

La cantidad de datos digitalizados ha aumentado a una velocidad asombrosa. Hoy en día, prácticamente el total de la información disponible en internet está digitalizado mientras que la capacidad de generar nueva información de los diferentes dispositivos aumenta segundo a segundo. Lo que entendemos por datos también está cambiando y un porcentaje creciente de la nueva información que se genera lo hace de manera no estructurada.

La Comisión Económica para Europa de las Naciones Unidas (UNECE) ha clasificado la información *big data* según sea de origen humano, generada por máquinas o procesada. Entre la información de origen humano destaca la información no estructurada en forma de textos, imágenes o vídeos (redes sociales, blogs, documentos personales, correo electrónico, mensajes de texto, búsquedas en internet, fotografías, audio y vídeo...) y que hoy en día podemos convertir en datos gracias a los avances en la capacidad de procesamiento y el desarrollo de los algoritmos de procesamiento de lenguaje natural (NLP), así como o algoritmos de redes neuronales utilizados para el tratamiento de imágenes y vídeos. El desarrollo de la tecnología y los numerosos dispositivos y sensores existentes en las calles o en las casas han permitido generar una ingente información proveniente de las máquinas. Por último, un tercer grupo de información es aquella registrada y procesada por empresas públicas y privadas. Entre las primeras, cabe destacar los datos generados por parte del sector público en forma de registros públicos. Entre los datos privados, una de las fuentes de información que se ha desarrollado más rápidamente, y principal objeto de análisis de este capítulo, son los datos agregados y anonimizados que provienen de las transacciones financieras.

El funcionamiento de los sistemas de pago modernos y de las instituciones financieras genera un registro completo de las transacciones diarias que se genera en tiempo real. Cada compra, cada débito, cada transferencia que realizamos deja una huella digital que termina registrándose como si de un gran libro de contabilidad se tratara. Esta gran cantidad de datos naturales y no estructurados sobre las transacciones, junto con los ricos metadatos asociados, está cada vez más al alcance de los investigadores y promete reconfigurar la medición económica. Cabe destacar que estos datos financieros, a diferencia de otros datos en tiempo real y alta definición, como los generados por la señal emitida a partir de terminales móviles o los generados de las aplicaciones de movilidad, nos permiten monitorizar actividad económica a través de movimientos financieros reales (como las compras) abarcando todos los sectores

de actividad, convirtiéndose así en una valiosa fuente de información para monitorizar la actividad.

Por diferentes motivos, esto no ha pasado desapercibido para la academia, las agencias nacionales de estadística y los responsables de la política económica, que vienen reafirmando desde hace unos años que los datos de transacciones financieras desempeñarán un papel cada vez más importante en la contabilidad nacional del siglo XXI (véase, por ejemplo, Lewis, 2020 y Ehrlich *et al.*, 2022).

En este capítulo mostramos cómo podemos utilizar los datos de transacciones financieras para replicar las cuentas nacionales. Nos centraremos en el caso del consumo de los hogares dada su relevancia (55 % del PIB). Además, mostraremos cómo podemos utilizar esta información para el desarrollo de indicadores de actividad en tiempo real y, dada su elevada granularidad, en alta definición.

En el primer apartado describimos brevemente la literatura reciente sobre el uso de los datos de transacciones financieras en tiempo real. Nos centraremos en la descripción de las diferentes datos financieros que han sido utilizados por los trabajos recientes, así como las ventajas e inconvenientes de utilizar este tipo de información frente al uso de los indicadores tradicionalmente utilizados por las agencias estadísticas.

En un segundo apartado describimos cómo podemos utilizar los datos financieros de pagos de los individuos a las empresas para replicar el consumo. Aparte de detallar los diferentes medios de pago, realizamos una descripción del proceso de limpieza y procesamiento de los datos para alcanzar una medición completa del consumo, en lugar de analizar solamente los datos de tarjetas de crédito y de débito como la mayoría de los trabajos existentes.

Aunque una de las propiedades más destacadas de este tipo de transacciones es la posibilidad de registrar la actividad económica en tiempo real, no menos importante es el componente de alta definición que presentan los datos de transacciones. En el tercer apartado mostramos cómo la alta granularidad de estos datos se está traduciendo en un aumento notable de las posibilidades que ofrece para el análisis económico, sobre todo en el desarrollo de cuentas distributivas del consumo e ingresos.

La alta definición de los datos, abre la puerta al uso de indicadores *big data* para el desarrollo de políticas económicas inteligentes. La alta granularidad de las transacciones supone una ventaja notable para los *policy-makers* a la hora de desarrollar políticas inteligentes. El hecho de que podamos acceder a la información de los individuos y empresas en alta definición permite enfocar y dirigir las políticas económicas allá donde más se necesitan o donde pueden lograr un mayor impacto en términos de bienestar.

Por último, y a modo de conclusión, introducimos un conjunto de reflexiones analizando las ventajas de la información proveniente de las transacciones financieras, aventurándonos a señalar algunos de los aspectos que vayan a tener más importancia en los próximos años.

## 2. TRANSACCIONES FINANCIERAS EN TIEMPO REAL: LITERATURA RECIENTE

La medición detallada del ciclo económico y sus componentes (consumo privado, inversión) ha constituido desde hace décadas un ingrediente clave en diferentes campos tanto de análisis como de política económica. Para ello, los economistas hemos utilizado tradicionalmente la información elaborada por organismos nacionales de estadística, diferentes ministerios o bancos centrales en forma de indicadores de diversa índole (PIB, consumo, producción industrial, desempleo, etc.) para evaluar el estado del ciclo económico. En términos generales, esta información tiene como base la información que proviene de encuestas a una muestra representativa de individuos o empresas sobre cuestiones de diferente índole (decisiones de compra, decisiones de inversión, . . .).

La información que proviene de las encuestas tiene ventajas y desventajas. Entre las ventajas, las encuestas están diseñadas para ser consistentes con el sistema de cuentas nacionales y ser representativas de la población. No obstante, también existen inconvenientes como la baja frecuencia de los datos, el número limitado de los encuestados, los retrasos en la publicación de estadísticas clave como el PIB, la baja granularidad de la información en algunos casos y las frecuentes revisiones de la información. Además, existen problemas de carácter logístico como los elevados costes y la cada vez menor participación en las encuestas.

Recientemente los investigadores y analistas han recurrido a los datos de las transacciones financieras para complementar la información tradicional. A pesar de que se les etiqueta también como “No tradicionales”, la información de las transacciones financieras lleva con nosotros mucho tiempo. La relación entre las transacciones financieras y la actividad económica tampoco es nueva y forma parte ya de trabajos seminales de grandes autores como Keynes, Fisher y muchos otros. No ha sido, por tanto, extraño que estos datos que “ocurren naturalmente” (Bean, 2016) comenzaran a explotarse masivamente a medida que la capacidad de proceso de los computadores y el desarrollo de nuevos algoritmos lo fueran permitiendo.

Pero la información que proviene de las transacciones financieras tiene también ventajas y desventajas. Entre las ventajas podemos destacar las siguientes:

- Al provenir de registros, y no de encuestas, la información puede ser igual de precisa si la información es suficientemente representativa y está debidamente procesada.
- El número de registros supera con creces normalmente al de los encuestados y la información posee un elevado grado de granularidad o alta definición.
- La información se procesa casi en tiempo real y suele ser de carácter diario. Esto supone una ventaja notable para los responsables de política económica, pues su respuesta a las condiciones económicas puede ser mucho más inmediata. De hecho, la necesidad de reaccionar rápidamente a las cambiantes condiciones económicas tras la crisis del COVID-19 ha potenciado, sin duda, el uso de los datos de transacciones financieras en la evaluación del estado de la actividad económica.

- Las transacciones están geolocalizadas, por lo que el potencial para el análisis a nivel desagregado es muy elevado.
- Por último, el coste de este tipo de información es mucho más bajo para las oficinas estadísticas. Esto es especialmente relevante para el desarrollo de estadísticas en los países en vías de desarrollo.

Sin embargo, también existen inconvenientes a la hora de utilizar este tipo de información para el seguimiento de la actividad económica:

- Aunque numerosos, los datos no recogen una muestra de la población sino de los clientes de la entidad financiera. Esto es muy relevante, pues la existencia de sesgos puede limitar la precisión de la información al evaluar el ciclo económico.
- En la mayoría de los casos, las bases de datos de transacciones de las instituciones financieras no fueron diseñadas para el análisis económico o para ser consistentes con las cuentas nacionales. Existen, por ejemplo, transacciones con tarjetas de crédito que no constituyen un gasto de consumo privado (pago de impuestos o compra de bitcoins con tarjetas, por ejemplo) o bien corresponden a una exportación de bienes y servicios (como el gasto con tarjetas de un no residente en España). Tener acceso a los datos brutos de transacciones financieras no garantiza que vayan a poder ser utilizados inmediatamente para el análisis. En términos generales hay un trabajo exhaustivo de procesamiento del dato antes de poder utilizarlo para el análisis.
- El tratamiento de los datos financieros tiene que cumplir en todo momento con la regulación y preservar la privacidad de los individuos.
- El coste de procesamiento del dato para las entidades que proporcionan los datos es normalmente elevado.

Aunque el procesamiento de estos datos no es sencillo y no está exento de costes, las numerosas ventajas en su utilización han llevado a analistas e investigadores a trabajar y mejorar continuamente su calidad. En este sentido, es clave que analistas e instituciones privadas y públicas continúen trabajando para garantizar la consistencia de esta información con las estadísticas oficiales. El desarrollo de un procesamiento de los datos consistente con los principios clave de las estadísticas oficiales (representatividad, estabilidad, consistencia...) es clave para el futuro de esta información como herramienta de análisis económico. Sólo así se puede garantizar la robustez y la confianza en este tipo de información para poder complementar con éxito a las fuentes de información tradicionales provista por los organismos oficiales.

Aunque en términos generales nos referimos a estos datos como transacciones financieras, su naturaleza y su utilización puede ser muy distinta. Baker y Kueng (2022) resumen bien la evolución en el uso de los datos de transacciones financieras que se ha producido en los últimos años en tres grandes grupos:

- Una primera oleada de trabajos comenzó a explotar los datos de las cuentas de los intermediarios bursátiles o *brokers*. A partir de ellos, algunos investigadores comenzaron a explorar y entender mejor como los inversores tomaban sus decisiones. A través de sus análisis encontraron diferencias por género o como los inversores cometían errores sistemáticos que penalizaban la tasa de retorno de las inversiones.
- A partir de principios de la década de 2010, comienzan a surgir trabajos que explotan datos detallados de las transacciones de las cuentas bancarias. Estos datos son suministrados bien por entidades bancarias individuales, bancos *online* o asociaciones de bancos. Dado que los bancos ofrecen una amplia gama de servicios a los clientes (cuentas corrientes, tarjetas de débito y crédito, hipotecas, préstamos al consumo...), estos datos ofrecen la posibilidad de observar no sólo los flujos completos de gastos e ingresos, sino también varios tipos de activos, el comportamiento de los préstamos y otra información demográfica. Como veremos más tarde, algunos de estos datos como los de transacciones de tarjetas resultaron ser muy relevantes en 2020 durante la epidemia del COVID-19 y los confinamientos.
- Una opción más limitada ha sido obtener acceso a los datos de las transacciones directamente de las empresas de tarjetas de crédito (véase, por ejemplo, Einav y Klenow, 2021) o de los emisores de tarjetas de crédito (por ejemplo, Gathergood *et al.*, 2021). El aumento de las aplicaciones de pago, como Apple Pay, Venmo y AliPay, ha creado nuevas vías para observar directamente las transacciones de gasto de los hogares en grandes sectores de la población de muchos países.
- Los análisis de la ciencia del comportamiento (*Behavioural Economics*) se han visto beneficiados del trabajo pionero de un ecosistema emergente de productos y aplicaciones *fintech*. Entre los datos que proveen estas *startups* se encuentran los que agregan y hacen un seguimiento de las distintas cuentas financieras de un cliente, y los que se centran más en un objetivo final concreto, como aumentar las tasas de ahorro o ayudar a pagar las deudas.

Pero sin duda alguna, la mejor demostración de las ventajas de estos datos ha sido su uso durante la crisis sanitaria generada por el COVID-19 que arrasó la economía mundial durante el primer trimestre de 2020. La pandemia y los confinamientos que se implementaron en muchos países para combatirla generaron una crisis económica sin precedentes. La crisis provocó una recesión súbita, sin precedentes en términos de magnitud y velocidad de propagación. Dada la incertidumbre del evento, los datos tradicionales que habían sido utilizados en anteriores ocasiones no lograban proporcionar una visión rápida y certera de lo que estaba ocurriendo.

La rapidez con la que se produjo la crisis durante los días del final de primer y principios del segundo trimestre de 2020 generó una elevada incertidumbre sobre la magnitud de la crisis y las políticas que se debían implementar para combatirla. En este sentido, la lentitud y desfase con la que se publican los indicadores económicos tradicionales no ayudaban a la toma de decisiones. Ante ello, muchos investigadores comenzaron a hacer uso de los

datos de transacciones financieras para analizar en tiempo real lo que estaba ocurriendo en la economía.

La mayoría de estos trabajos se centraron en los datos de transacciones de tarjetas para elaborar un seguimiento en tiempo real de la magnitud de la crisis. La respuesta fue relativamente rápida y algunos documentos de trabajo se publicaron solo un par de meses después del estallido del COVID-19. La mayoría de estos trabajos se enfocaron en países desarrollados como Andersen *et al.* (2020a) para Dinamarca y Suecia; Chetty *et al.* (2020) y Cox *et al.* (2020) para los Estados Unidos; Chen *et al.* (2021) para China; Carvalho *et al.* (2021) y Aspachs (2021) para España; Chronopoulos, Lukas, Wilson (2020) y Hacioglu, Kanzig, Surico (2020) para el Reino Unido, y Bounie, Camara, Galbraith (2020) en el caso de Francia. Sólo un par de estudios han extendido este trabajo empírico a las economías emergentes, como el caso de Carvalho *et al.* (2021) que utilizó información de tarjetas de crédito y débito como *proxy* del consumo en Turquía, México, Colombia, Perú y Argentina, o Chen (2021) que utilizó datos de consumo *online* para China.

En nuestro caso (Carvalho *et al.*, 2021) pronto comenzamos a descubrir algunas de las características de la crisis y de los efectos de los confinamientos en el consumo en los diferentes países analizados: una caída sin precedentes del consumo media del 40-45 % consecuencia de los confinamientos, con una elevada propagación en los distintos países, con mayor intensidad en aquellos sectores sujetos a las restricciones (como por ejemplo servicios), con un mayor efecto relativo en los segmentos de renta más elevados pero que, afortunadamente, fueron seguidos de una recuperación también sin precedentes.

### 3. RE-CONSTRUYENDO EL CONSUMO A TRAVÉS DE LOS MEDIOS DE PAGO

Para la construcción de nuestros indicadores de consumo en tiempo real y alta definición nos basamos en el universo de operaciones financieras realizadas por clientes particulares de BBVA en España en el periodo 2015T2 a 2021T4. A diferencia de otros trabajos previos, como destacamos previamente, vamos más allá del uso de datos de transacciones de tarjetas de crédito y débito y consideramos todos los medios de pago a través de los cuales los individuos pueden consumir, capturando de forma más precisa la dinámica del consumo total y entendiendo la evolución desigual de cada componente. Concretamente, medimos el consumo realizado con tarjeta de crédito y débito en puntos de venta (TPV) físicos y virtuales (comercio electrónico), que suponen un 26,5 % y un 8,5 % del consumo total respectivamente; el consumo realizado a través de domiciliaciones bancarias, comúnmente utilizadas para el pago de suministros del hogar como agua, electricidad, gas y comunicaciones (teléfono, red wifi, televisión a la carta, etc.), con un peso del 20 % del consumo total; el consumo a través de transferencias monetarias, ligado sobre todo a la compra de bienes duraderos como vehículos (5 % del consumo total); el pago de alquileres, que se trata del mayor componente del consumo (cercano al 25 % del total); y finalmente el efectivo, aproximado a través de las retiradas de efectivo en los cajeros BBVA y ventanilla, que representa el 15 % del total.

Los indicadores contruidos a partir de estos datos siguen los principios de contabilidad nacional para replicar las estadísticas oficiales del consumo agregado en las cuentas nacionales elaboradas por el INE. Para ello es importante entender los métodos utilizados por el INE, que están avalados y en línea con el sistema de cuentas nacionales de Eurostat. La información base en el caso del INE proviene de la *Encuesta de Presupuestos Familiares (EPF)*, que mide el consumo individual y de los hogares en los distintos grupos demográficos y tipos de productos y servicios. En esta encuesta, se define un conjunto de 2.275 secciones censales en función del tamaño del municipio, el empleo, la edad, la educación y otras características socioeconómicas. Dentro de estas secciones, se seleccionan aleatoriamente diez viviendas para participar en la encuesta durante dos años, con una rotación escalonada en la que la mitad de la muestra se sustituye cada año. Los hogares registran sus gastos durante un periodo de dos semanas donde a cada compra se le asigna una clasificación basada en el sistema de cinco dígitos de la clasificación individual de consumo por finalidad (*COICOP*, por sus siglas en inglés: *Classification of Individual Consumption by Purpose*). Los pagos recurrentes se estiman por el importe de la última factura emitida. En el caso de los hogares que son propietarios de sus viviendas, el INE imputa el valor de consumo de los servicios de la vivienda utilizando información sobre el tamaño de la casa y los precios locales de alquiler, además de las estimaciones subjetivas de los encuestados.

Para convertir los datos de gasto de BBVA en una encuesta de consumo hay dos temas esenciales a tratar. El primero es la definición de la muestra: tenemos que asegurar que el conjunto de clientes de BBVA es representativo de la población nacional. El segundo es que no todo el gasto de los consumidores equivale al consumo. El gasto en productos fuera del sistema de clasificación *COICOP* no se considera consumo en cuentas nacionales. Sin embargo, los datos de gasto del banco no siguen esta clasificación para limpiar las observaciones no asociadas a consumo, por lo que es necesario hacer una asignación de dicho gasto a *COICOP* para desestimar todo lo que no se considera consumo. Otro aspecto relevante al que le tenemos que prestar atención es a la estimación de los servicios de vivienda, que son una parte importante del consumo total (25 %), pero no se observan directamente en los datos de gasto de quienes son propietarios y ocupan sus viviendas.

### **3.1. De clientes BBVA a población representativa**

El conjunto de clientes particulares de BBVA que realizan al menos una transacción relacionada con el consumo entre 2015 T2 y 2021 T4 es de 10.270.041, lo que supone un 26 % del total de la población adulta España en 2021, que asciende a 39.177.710. Muchos de estos clientes gastan con poca frecuencia a través de nuestro sistema bancario. Para asegurar una vinculación de los clientes que nos permita seguir mejor su trazabilidad, definimos un panel de clientes activos que realizan al menos diez transacciones relacionadas con consumo en cada trimestre. De esta forma, nos quedamos con 1.827.866, excluyendo autónomos cuyas transacciones podrían reflejar insumos de producción en lugar de consumo. Además, este panel de clientes activos durante todo el periodo de estudio nos asegura que cualquier creci-

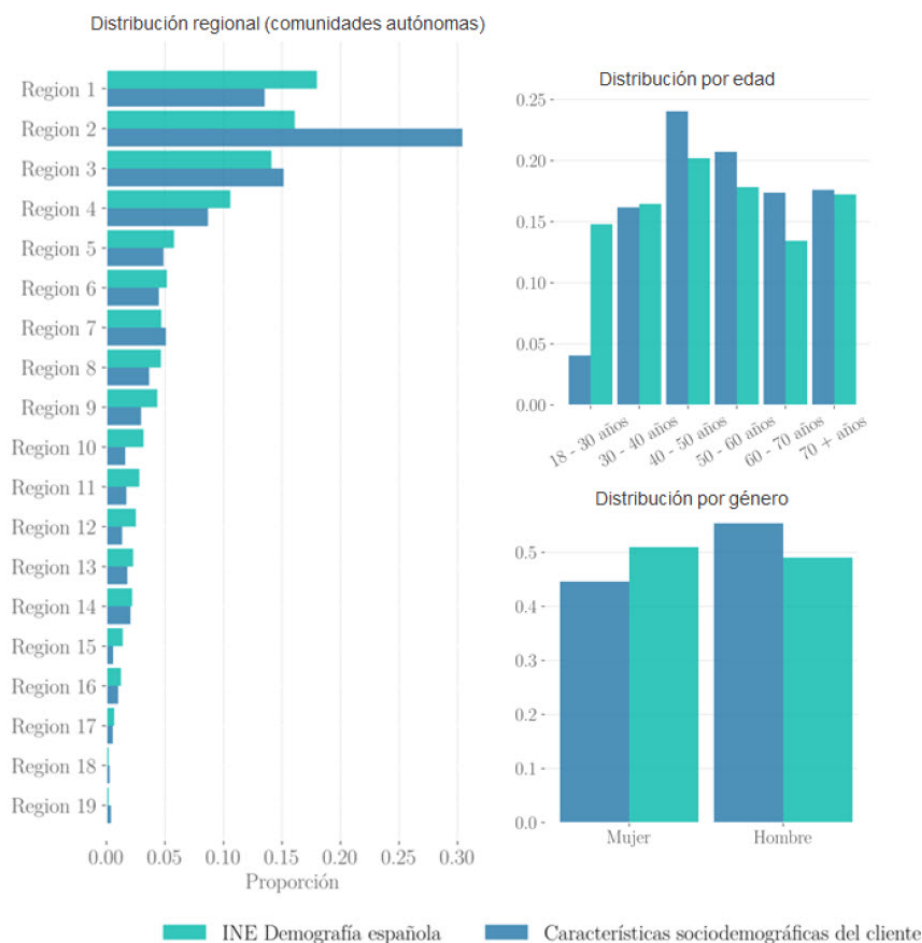


miento observado en el gasto agregado está impulsado por el aumento del gasto de nuestros clientes y no por un aumento de la cuota de mercado de BBVA.

Sobre este conjunto de clientes activos, estudiamos la distribución de su ubicación geográfica, de su residencia, edad, sexo e ingresos del distrito, comparándolos con los de todos los españoles registrados en el censo (figura 1). Para cada cliente de BBVA se observan datos de edad, género y domicilio. Los gráficos de barras comparan la distribución de las características de los clientes activos con los datos del censo español en 2018. Para construir la distribución de la renta del distrito, utilizamos la información que proporciona el INE sobre la renta

Figura 1.

### Distribución sociodemográfica de los clientes activos



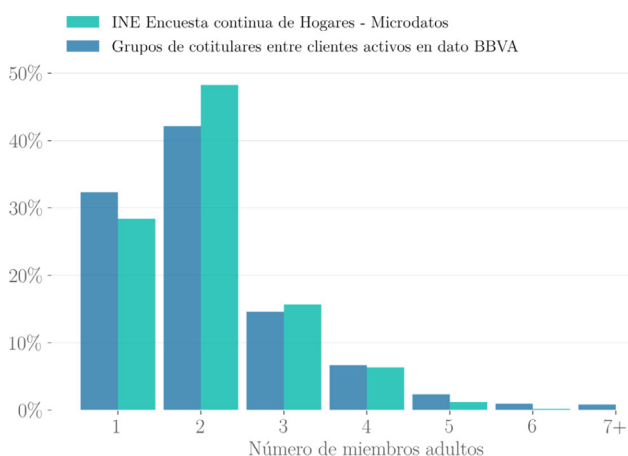
media de los residentes de cada sección censal en España (36.581 en total)<sup>1</sup>. Sobre la base de la renta media, agrupamos las secciones censales en quintiles dentro de cada una de las 52 provincias españolas y trazamos la distribución de la población española en cada quintil. A continuación, asignamos los clientes activos a las secciones censales en función del código postal y trazamos la distribución entre los quintiles.

En general las distribuciones están claramente relacionadas, pero existen discrepancias relevantes a solventar: por ejemplo, los clientes activos de BBVA están sobrerrepresentados en una región concreta, entre los hombres y entre las personas de mediana edad. También es más probable que vivan en distritos con mayores ingresos. Es por ello, que necesitamos convertir esta muestra en representativa reponderando el consumo de cada grupo como explicamos en el apartado 3.2.3.

La estructura familiar de los clientes activos es importante para determinar su consumo, pero no se observa directamente en los datos. Es por ello que la inferimos vinculando a cada cliente activo con el conjunto de otros clientes BBVA (estas relaciones se establecen sobre información tokenizada, sin poder identificar la identidad del cliente) que hayan firmado conjuntamente un contrato financiero (por ejemplo, que sean cotitulares de una cuenta bancaria, responsables conjuntos de un préstamo, etc.) en cualquier momento de la muestra y que residan en el mismo código postal al final de la muestra. Esto crea una estimación inicial del número de personas en el hogar de cada cliente activo además de él mismo. En estos casos, se unen los clientes en un solo hogar generando 1.589.280 grupos de hogares.

Figura 2.

### Tamaño del hogar: dato BBVA vs. dato oficial



<sup>1</sup> Esta información existe para todas las comunidades autónomas excepto para el País Vasco y Navarra. Ceuta y Melilla son pequeños enclaves que no están subdivididos en secciones censales.

En los casos en los que un cliente activo no está emparejado con ningún otro cliente de BBVA, pero figura como casado, suponemos que reside con otra persona, por ejemplo, un cónyuge. Por último, BBVA registra para cada cliente el número de adultos dependientes en el hogar. Si después de los pasos anteriores un cliente activo se agrupa con menos personas de las que aparecen como adultos dependientes, registramos el número de miembros adicionales del hogar igual al número de adultos dependientes. La *figura 2* compara la distribución resultante del tamaño de los hogares según nuestro procedimiento de agrupación con los datos oficiales. En general encontramos bastantes similitudes entre ambos grupos.

En resumen, contamos con una muestra de 1.827.866 clientes activos que forman 1.589.280 hogares, cuyo gasto en consumo observamos entre 2015 y 2021. En el caso de la *Encuesta de Presupuestos Familiares* del INE, contaron con 40.285 adultos y 19.170 hogares. De esta forma, nuestros datos incrementan significativamente el número de individuos cuyo consumo se puede registrar, lo que a su vez permite un corte mucho más fino de los datos. Además, a diferencia de nuestros datos, no se puede hacer un seguimiento del gasto a lo largo del tiempo de los individuos con los datos oficiales, dado que ningún hogar participa en la *EPF* durante más de dos años.

### 3.2. Del gasto al consumo

El siguiente reto es convertir los datos de gasto individual en datos de consumo. Para el consumo no relacionado con la vivienda, utilizamos los datos de transacciones para clasificar las compras individuales en dos grupos, las relacionadas con el consumo y los que no lo están. En el caso de que correspondan a datos de consumo, les asignamos una categoría de consumo de acuerdo a la clasificación *COICOP*. Por otro lado, para el consumo de vivienda, estimamos un modelo de regresión simple que predice los pagos de alquiler observados de las características de los hogares con objeto de imputar los pagos mensuales de vivienda de todos los clientes activos.

#### 3.2.1. Consumo no relacionado con la vivienda

De acuerdo a la definición de gasto en consumo final de cuentas nacionales, eliminamos las transacciones bancarias que no se ajustan a esta definición como por ejemplo las transferencias a instituciones de inversión o los pagos de impuestos y clasificamos cada transacción en categorías de consumo de acuerdo a la clasificación *COICOP* a nivel de dos dígitos (*cuadro 1*).

Hay tres modos de transacción distintos en los datos –gastos con tarjeta, adeudos en cuenta y transferencias bancarias no periódicas– y cada uno tiene una estructura distinta para categorizarlos. En el caso de las tarjetas nos basamos en el código de cliente comercial (MCC)

## Cuadro 1.

**Categorías de consumo COICOP (dos dígitos)**

<i>Categoría</i>	<i>Descripción</i>
01	Alimentos y bebidas no alcohólicas
02	Bebidas alcohólicas, tabaco y narcóticos
03	Ropa y calzado
04	Vivienda, agua, electricidad, gas y otros combustibles
05	Mobiliario, equipamiento doméstico y mantenimiento rutinario del hogar
06	Salud
07	Transporte
08	Comunicaciones
09	Ocio y cultura
10	Educación
11	Restaurantes y hoteles
12	Bienes y servicios diversos

o identificador del comercio, que es un sistema estandarizado para clasificar las actividades comerciales. Hay un total de 835 MCC, lo que permite desagregar con precisión las categorías de gasto y asignarlas a una categoría *COICOP*, con dos excepciones. En primer lugar, hay MCC que se refieren al consumo genérico. El ejemplo más destacado es la retirada de efectivo en cajeros automáticos. En segundo lugar, un número limitado de MCC se refiere a las ventas de minoristas multiproducto, como los supermercados. En estos casos, utilizamos las estadísticas publicadas sobre la distribución de las ventas en las categorías *COICOP* por sectores para imputar el gasto en estos establecimientos.

Para las operaciones de adeudo en cuenta, principalmente pago de facturas de suministros de energía, telecomunicaciones, seguros... , utilizamos la misma aproximación. Nos basamos en la empresa que recibe el pago para asignarla a una categoría *COICOP* a través de su identificación fiscal o código CNAE. Las transferencias ordinarias son el medio de pago más complicado de clasificar. Si identificamos a la contrapartida como una empresa, clasificamos las transferencias utilizando, como en el caso de adeudos, la información fiscal y pública de la empresa. El **cuadro 2** presenta el número y el volumen de las transacciones realizadas por

## Cuadro 2.

**Volumen de datos de consumo de clientes activos (todo el período)**

<i>Categoría de gasto</i>	<i>Volumen de transacciones</i>	<i>Número de transacciones</i>
Transacciones presenciales con tarjeta	60.319 millones de euros	1.772 millones
Transacciones <i>online</i> con tarjeta	11.858 millones de euros	313 millones
Adeudos en cuenta	66.036 millones de euros	752 millones
Retiradas de efectivo	64.592 millones de euros	359 millones
Transferencias excl. renta	11.148 millones de euros	15 millones

los clientes activos de nuestra muestra que clasificamos como relacionadas con el consumo, desglosadas por tipo de transacción.

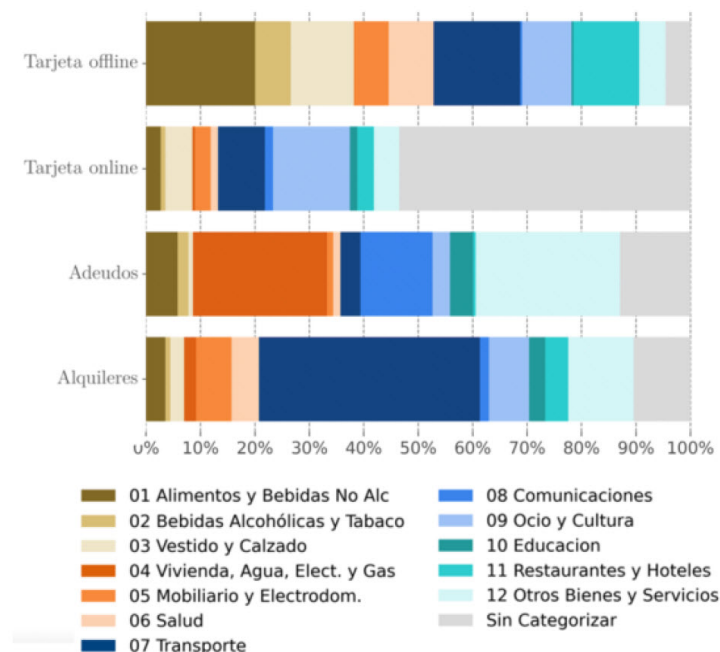
El valor total del gasto es de aproximadamente 200.000 millones de euros, que engloban 3.000 millones de transacciones totales. Aunque las transacciones con tarjeta constituyen la gran mayoría del total de transacciones, su valor total es comparable al de los adeudos en cuenta.

La **figura 3** muestra la distribución de las categorías de consumo según el medio de pago. Se observa una importante heterogeneidad entre ellos. El gasto en alimentación constituye una parte sustancial del gasto con tarjeta presencial. El transporte representa casi la mitad de las transferencias no periódicas, mientras que los pagos de suministros se realizan principalmente mediante adeudos en cuenta.

Por último, excluimos de la muestra a los clientes activos cuyo consumo no relacionado con la vivienda es elevado en relación con los ingresos medios de su sección censal, eliminando así valores atípicos de la distribución.

Figura 3.

### Distribución del consumo por COICOP por medio de pago



### 3.2.2. Consumo relacionado con la vivienda

Para la medición del consumo en vivienda, el primer paso es identificar todos los pagos de alquiler utilizando los términos "alquiler" o "renta" en castellano y otros idiomas regionales. Excluimos las transacciones que incluyen además términos que indican que el pago del alquiler es para un bien que no es la vivienda, como un garaje, una plaza de aparcamiento o un coche. También imponemos un valor mínimo de 100 euros para que una transacción se considere alquiler.

La unidad natural de análisis del consumo relativa a la vivienda es el hogar, por lo que buscamos los pagos realizados por todos los individuos que componen los hogares, sean o no clientes activos. A continuación, sumamos todos los pagos de alquiler a nivel de hogar y mes, obteniendo 437.307 hogares que tienen al menos un pago de alquiler. Para evitar el ruido derivado de los hogares con pocas observaciones mensuales de alquiler, nos centramos en los hogares con pagos de alquiler no ausentes en, al menos, 70 de los 81 meses totales de nuestra muestra, quedándonos con 32.127 hogares.

Para el modelo de estimación del alquiler, utilizamos los ingresos (que representan indirectamente la calidad de la vivienda), los pagos de suministros (que representan indirectamente el tamaño de la vivienda) y la ubicación geográfica. Para los ingresos nos basamos en los datos BBVA sobre los ingresos mensuales procedentes de los salarios, las prestaciones públicas y las pensiones, y calculamos los ingresos medios semestrales de los hogares. Los pagos de suministros se calculan a partir de la tabla de adeudos en cuenta. Solamente mantenemos en la muestra de estimación los hogares que tienen al menos un mes de pagos de suministros e ingresos observados. Esto reduce el número de hogares objeto de estudio a 16.977.

Para la localización geográfica, buscamos definir unidades geográficas que estén lo suficientemente pobladas de hogares como para poder estimar los efectos fijos de forma fiable. Por último, realizamos una regresión de los pagos mensuales de alquiler por hogar sobre los efectos fijos de la unidad geográfica, los ingresos y los pagos de suministros mediante mínimos cuadrados ordinarios. Cuando no se dispone de información sobre los ingresos o los suministros para elaborar el registro de un mes determinado, utilizamos la media del hogar en todos los meses.

El **cuadro 3** muestra los resultados. Aunque es sencillo, el modelo explica el 40 % de la variación de los pagos de alquiler, y ambas covariables continuas son altamente significativas y contribuyen a un elevado  $R^2$  dentro de la región. Los coeficientes estimados implican que un cambio de una desviación estándar en los ingresos desplaza los pagos de alquiler en 70 euros al mes, es decir, un 0,28 del IQR de la distribución general de los pagos de alquiler. El impacto de los suministros es más moderado, con un cambio de una desviación estándar que desplaza el alquiler en sólo 21 euros.

A continuación, utilizamos nuestra regresión de alquileres estimada para imputar los alquileres mensuales a todos los hogares. Cuando un hogar se encuentra fuera de las unidades

## Cuadro 3.

**Regresión modelo de alquiler**

<i>Variable</i>	<i>Modelo</i>	<i>Conjunto en estudio</i>
Gasto en suministros al hogar	0,0884 (0,0008)	
Ingresos	0,0362 (0,0011)	
Número de hogares	16.977	15.512
Número de observaciones	1.134.735	15.512
R <sup>2</sup>	0,3911	
R <sup>2</sup> ajustado	0,3765	
Dentro de R <sup>2</sup>	0,1200	
MSE raíz	204.614,4	221,64

geográficas definidas para la muestra de estimación, lo asignamos a la unidad más cercana basándonos en la distancia del centroide. Cuando no se dispone de información sobre los ingresos o suministros para elaborar el registro de un mes determinado, utilizamos la media del hogar en todos los meses. Si un hogar no tiene registros de suministros o ingresos, imputamos la media de la unidad geográfica.

Para elaborar una primera estimación de la precisión fuera de la muestra, consideramos los 15.512 hogares para los que observamos entre 50 y 70 pagos mensuales de alquiler y calculamos el error cuadrático medio del alquiler imputado con respecto al alquiler real para un mes elegido al azar para cada hogar. El error cuadrático medio sólo aumenta ligeramente en comparación con la muestra de estimación, lo que sugiere que nuestro modelo de alquiler, aunque sencillo, se generaliza bien fuera de la muestra. Los promedios también se ajustan bien: el alquiler medio real es de 551 euros y el alquiler medio imputado es de 538 euros.

**3.2.3. Ponderación de hogares y demográfica**

El proceso descrito da lugar a una medida de consumo no relacionado con la vivienda  $c_i^{NH}$  para cada cliente activo  $i$ , y a una medida de consumo de servicios de vivienda definida a nivel del hogar  $h$ . El último paso para producir medidas de consumo agregadas es ajustar estos totales en función de la vivienda y la estructura demográfica. Es importante tener en cuenta la estructura de los hogares porque una parte del gasto de cada cliente activo se realiza potencialmente en nombre de otros. Por otro lado, como no tabulamos el gasto de los clientes no activos (excepto en vivienda), estamos perdiendo la parte de su gasto que beneficia a clientes activos. Para equilibrar estos efectos, adoptamos el siguiente esquema de ponderación.  $A(i)(O(i))$  es el conjunto de clientes activos (otros) del hogar de  $i$ . Así, el consumo ponderado del hogar es:

$$c_i = \frac{\sum_{j \in A(i)} c_i^{NH} + c_{h(i)}^H}{|A(i)| + 0,5|O(i)|}$$

Supongamos primero que un hogar está formado sólo por clientes activos. En ese caso, la fórmula anterior agrega el gasto de todos los miembros y lo divide en partes iguales. Si el hogar también contiene miembros no activos, aplicamos una ponderación adicional a la baja que trata a cada cliente no activo como 0,5 de un cliente activo. Los clientes no activos comparten el consumo de los clientes activos, pero también generan potencialmente gasto de consumo fuera del universo BBVA. La ponderación a la baja de  $0,5|O(i)|$  tiene en cuenta estas fuerzas contrapuestas.

De esta forma, agregamos el gasto individual a nivel de género ( $g$ ), grupo de edad ( $a$ ) y quintil de ingresos del distrito ( $q$ ). Dado que los quintiles de distrito se forman por separado para cada región, esta última variable garantiza también la representatividad regional. Así, se define  $c_{t;g,a,q}$  como la suma de acuerdo a la fórmula anterior de todos los clientes activos en la celda ( $g, a, q$ ) calculada en el momento  $t$ . Dependiendo del contexto,  $t$  puede ser anual, trimestral, mensual, etc. Para agregar entre celdas en cada periodo de tiempo  $t$ , tenemos en cuenta los desequilibrios demográficos entre la muestra de clientes activos y los datos censales españoles de 2018. Se define  $x_{g,a,q}^{INE}$  como el recuento total de adultos españoles según los datos censales en la celda ( $g, a, q$ ) en 2018. Asimismo, se define  $x_{\tau(t);g,a,q}^{BBVA}$  como el recuento total de clientes activos en la celda ( $g, a, q$ ) en el año  $\tau$  que depende del periodo de interés  $t$ . El consumo total en cada celda en el momento  $t$  es:

$$c_{t;g,a,q}^W = c_{t;g,a,q} \times \left( \frac{x_{g,a,q}^{INE}}{x_{\tau(t);g,a,q}^{BBVA}} \right)$$

El consumo agregado es la suma de todas las celdas, mientras el consumo del cuartil regional del distrito ( $q$ ) es la suma de todas las categorías de género y edad manteniendo ( $q$ ) fija, y así sucesivamente. El consumo específico de cada categoría se obtiene considerando únicamente el subconjunto de  $c_{t;g,a,q}$  que pertenece a la categoría COICOP de interés. Con esta ponderación corregimos los desequilibrios demográficos en la elaboración de medidas de consumo agregadas. Para el análisis de las cuentas nacionales distributivas y la dinámica del consumo necesitamos una muestra representativa a nivel nacional del consumo individual. Para ello, para cada celda demográfica ( $g, a, q$ ) en el año  $\tau$  extraemos con reemplazo de la población de clientes activos  $x_{g,a,q}^{INE}$  veces. Esto produce una muestra nacional de tamaño igual a la población adulta española para la que se puede realizar un análisis distributivo.

#### 4. EL CONSUMO EN TIEMPO REAL TR

Con la metodología descrita en la sección anterior, agregamos todas las observaciones de consumo con frecuencia trimestral para formar una medida de consumo nacional agregada y compararla con el consumo final de los hogares oficial publicado por el INE. De igual manera, comparamos el desglose del consumo de cada categoría COICOP entre nues-



tros datos y los oficiales. En ambos casos la correlación entre nuestros indicadores y las cifras oficiales es significativamente alta, a pesar de los enfoques y metodologías distintas para llegar a las cifras agregadas, con la ventaja de que nuestros indicadores tienen una frecuencia diaria, pudiendo dar respuestas rápidas a la evolución del consumo, así como gran detalle regional, sectorial y por características del consumidor. De esta forma, describimos a continuación la evolución del consumo diario; el consumo agregado de los hogares a nivel provincial; una medida de alta frecuencia de la cesta de consumo de los hogares nacionales; y un desglose del consumo por medio de pago. Estos ejemplos ponen de manifiesto como nuestros datos pueden ayudar a generar nuevas medidas relevantes para el desarrollo de políticas, no presentes en los datos oficiales.

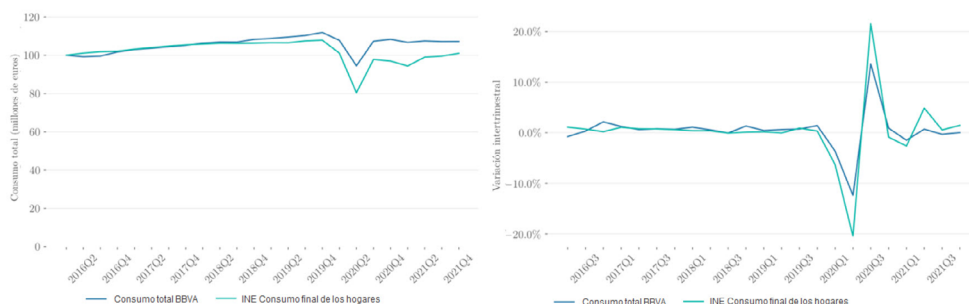
#### 4.1. Relación con la contabilidad nacional

La estimación del consumo de los hogares en las cuentas nacionales se basa en diversas fuentes de información, como las ventas de las empresas registradas a través de encuesta y los requisitos de información obligatoria, los datos administrativos sobre, por ejemplo, las compras de automóviles, y las encuestas a los hogares para compras específicas, por ejemplo, el Panel de Consumo Alimentario. Estas diversas fuentes de datos se combinan después en un modelo estadístico y el INE publica el consumo final de los hogares con una periodicidad trimestral y presenta los desgloses por COICOP con una periodicidad anual.

La **figura 4** compara las series agregadas trimestrales del INE en términos reales, tanto en niveles como en tasas de crecimiento intertrimestral, con nuestros datos, excluyendo los efectos estacionales en ambas series. Las series se deflactan utilizando el índice de precios al consumo (IPC) oficial definido a nivel de mes-región-COICOP. Ambas series tienen una alta correlación tanto en niveles como en tasas de crecimiento con frecuencia trimestral pese a que se construyen de forma distinta. Aunque seguimos cuidadosamente las definiciones de la contabilidad nacional al diseñar qué transacciones filtrar, obtenemos el consumo agre-

Figura 4.

#### Consumo agregado de BBVA frente a la contabilidad nacional (real)



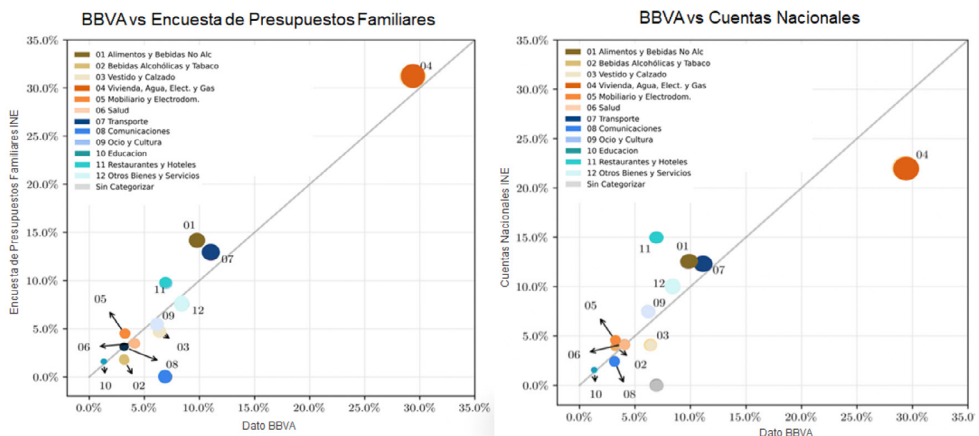
gado mediante una simple suma ponderada sobre las medidas de consumo individuales, a diferencia del SEC (Sistema Europeo de Cuentas, que el INE utiliza para elaborar las cuentas nacionales), que define un complejo conjunto de procesos para combinar múltiples fuentes de datos en los agregados de las cuentas nacionales.

Una discrepancia notable en las series se produce durante la crisis del COVID-19, cuando observamos que tanto la caída como la recuperación del consumo son más estables según nuestra medida que según los datos oficiales. Es difícil determinar cuál de estas series es más precisa en captar la dinámica del consumo durante la crisis. Una hipótesis es que los datos oficiales utilizan como insumo las ventas de las empresas, pero que durante y después de la pandemia los consumidores comenzaron a adquirir insumos de empresas diferentes a las incluidas en el modelo del INE. Esto implicaría que captaríamos un mayor consumo y por tanto generaríamos un nivel más alto.

A nivel sectorial, comparamos la distribución del consumo agregado entre las categorías *COICOP* según las cuentas nacionales y la *EPF* con nuestros datos en 2019. En estos últimos, distribuimos el efectivo entre las categorías *COICOP* utilizando los mismos porcentajes que observamos para el gasto con tarjeta presencial. El supuesto es que el gasto en efectivo y en tarjetas presencial son sustitutos y, por tanto, deberían gastarse en artículos relacionados. La figura 5 compara los niveles (logarítmicos) de consumo de las diferentes medidas con nuestros datos. En general, existe una estrecha relación entre los niveles de consumo específicos *COICOP* de las cuentas nacionales y los de la *EPF* y nuestros datos. Sin embargo, nuestros datos tienen una mejor representatividad de las cuentas nacionales: el error absoluto medio con respecto a las cuentas nacionales (*EPF*) es de 0,266 (0,333) puntos logarítmicos en todas las categorías *COICOP*. La mayor diferencia aparece en *COICOP* 2 “Bebidas alcohólicas y

Figura 5.

### Distribución del gasto en las categorías de la *COICOP*



tabaco”. Los hogares parecen especialmente propensos a infradeclarar el gasto en esta categoría debido a la estigmatización social, mientras que nuestros datos coinciden con las cuentas nacionales. La mayor divergencia entre los datos naturales y las cuentas nacionales se da en COICOP 3 “Vestido y calzado”. Esto se debe probablemente a la atribución de dinero en efectivo en proporción al gasto con tarjeta presencial, donde esta COICOP está sobrerrepresentada (como se ve en la [figura 3](#)).

Por último, para mostrar el valor añadido de la medida de consumo total construida, mostramos la comparativa de nuestros datos con los datos oficiales si tenemos en cuenta una medida de consumo agregada derivada únicamente del gasto con tarjeta. El gasto con tarjeta se ha utilizado recientemente para rastrear los efectos del COVID-19 en varios trabajos (Andersen *et al.*, 2020; Carvalho *et al.*, 2021). ¿Hasta qué punto el gasto con tarjeta nos permite monitorizar el consumo agregado? En la [figura 6](#) se representan las mismas curvas que en la [figura 4](#), pero utilizando únicamente la parte del consumo derivada del gasto con tarjeta, incluida la retirada de efectivo con tarjeta. La serie de datos de tarjetas tiene una escasa cobertura agregada del consumo de las cuentas nacionales. Además, su tasa de crecimiento tiene un notable sesgo al alza que no está presente en la medida completa del consumo construida. La tasa media de crecimiento intertrimestral del consumo oficial de los hogares entre 2016 y 2021 es del 0,55 %, mientras la tasa de crecimiento de la medida completa de consumo es del 0,83 % y la tasa de crecimiento de la medida de consumo basada en tarjeta es del 2,01 %. Es por tanto que el gasto con tarjeta tiene notables limitaciones para monitorizar la evolución del consumo.

Figura 6.

### Consumo agregado de BBVA frente a la contabilidad nacional (solo tarjetas)



#### 4.2. Nuevos indicadores del dato en tiempo real y alta definición

Una de las principales ventajas de nuestros datos es su alta granularidad, pudiendo producir nuevas medidas de contabilidad nacional que van más allá de lo que ya está disponible en los institutos de estadística. A continuación mostramos varios ejemplos que lo ilustran.

#### 4.2.1. Cuentas nacionales en alta frecuencia

En tiempos de alta incertidumbre económica como los actuales es importante comprender cómo evoluciona la economía en tiempo real. La definición de  $t$  en  $c_{t:g,a,q}^W$  depende del usuario y puede ajustarse a la frecuencia que se necesite. Para ilustrar una versión de alta frecuencia del consumo agregado, la **figura 7** toma  $t$  como un día. La vivienda se imputa a nivel mensual, por lo que la dividimos por igual entre todos los días de un mes determinado. Además, para evitar que los grandes pagos en un solo día influyan en la medida (por ejemplo, las fechas periódicas en las que se pagan las facturas de suministros), calculamos una media móvil no centrada de 28 días. Por último, tenemos en cuenta la estacionalidad diaria trazando las tasas de crecimiento interanual con respecto a los días comparables (por ejemplo, el crecimiento del consumo desde el primer domingo del año  $t - 1$  hasta el primer domingo del año  $t$ ). Como cabría esperar, la serie diaria es más volátil –en parte debido a los problemas de calidad de los datos en 2015 T2-2016 T4– que la serie trimestral, pero también es capaz de captar el impacto económico de las grandes perturbaciones. La caída del consumo agregado debida al confinamiento del COVID-19 es severa e inmediata.

Otra medida de alta frecuencia que resulta de interés es la cesta de consumo de los consumidores. La **figura 7** muestra la evolución de las cuotas de consumo de las categorías COICOP con una frecuencia mensual, donde de nuevo se observa un cambio drástico debido al COVID-19. El gasto en restaurantes y hoteles se desploma en 2020, mientras que otras categorías, como la comunicación, permanecen relativamente estables a lo largo de la muestra.

Figura 7.

#### Crecimiento del consumo diario (media móvil interanual de 28 días)

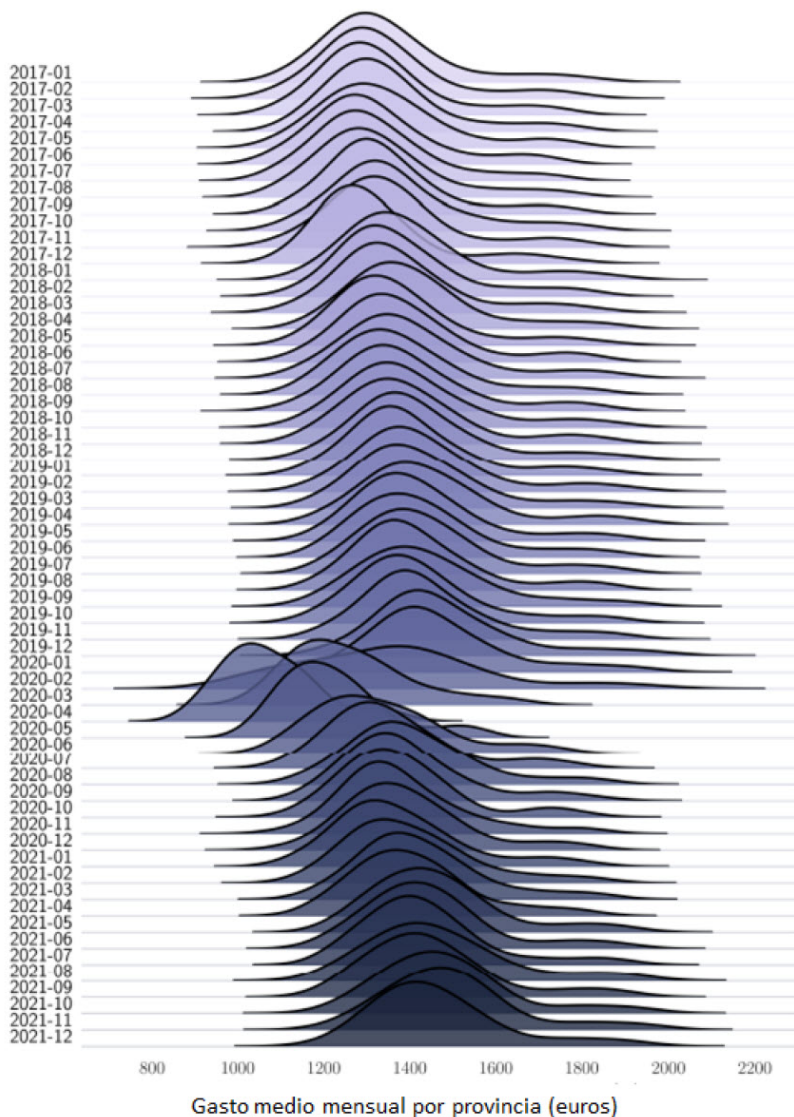


**4.2.2. Cuentas nacionales geográficamente detalladas**

Otra de las ventajas de nuestros datos es el detalle geográfico, que nos deja una huella precisa la evolución y desigualdades del consumo a nivel regional. La despoblación rural ha

Figura 8.

**Distribución del consumo por persona por comunidades autónomas**



sido una de las principales preocupaciones de la última década, pero no existen estadísticas regionales nacionales que permitan estudiar esta desigualdad regional con detalle. En España, el INE elabora estimaciones del PIB a nivel regional, pero sin desagregar el consumo y en ningún caso elabora estadísticas oficiales a un nivel provincial más detallado. Es aquí donde nuestros datos, de nuevo, aportan un valor diferencial.

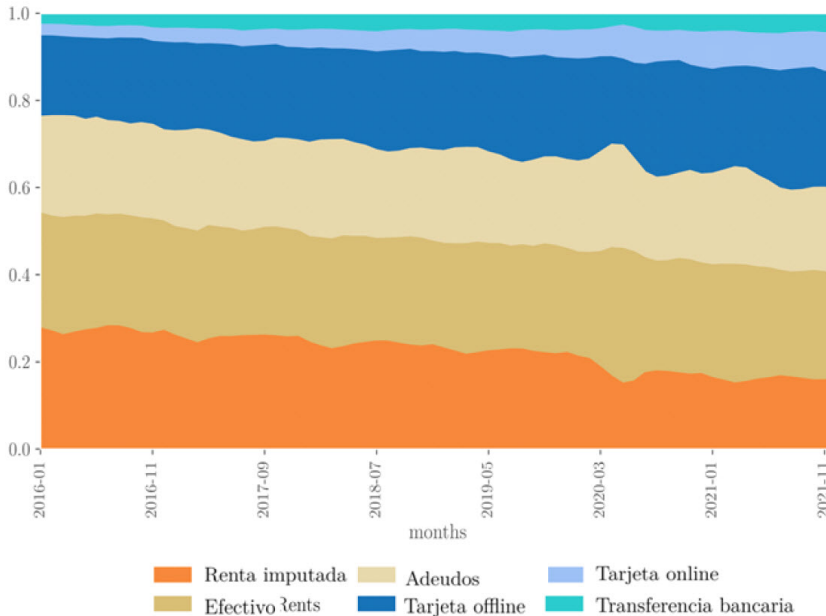
La *figura 8* ilustra la desigualdad provincial medida con nuestros datos. Para cada una de las 52 provincias españolas, calculamos el consumo agregado mensual desestacionalizado y lo dividimos por el número de adultos de la provincia. En general, se observa una masa de provincias en el centro de la distribución y una larga cola derecha de provincias de mayor consumo. De esta forma, estas estadísticas pueden constituir una importante base de datos para el diseño de políticas “inteligentes” que permitan atacar la desigualdad geográfica regional en España.

#### 4.2.3. Cuentas nacionales por medio de pago

Como hemos destacado a lo largo del capítulo, nuestro estudio monitoriza las transacciones financieras de todos los medios de pago a través de los cuales se efectúa el consumo. Aunque existen algunas encuestas que recogen la frecuencia con la que los individuos utilizan

Figura 9.

#### Distribución por medio de pago



los diferentes medios de pago, éstas no se restringen a los pagos relacionados específicamente con el consumo, por lo que nuestros indicadores aportan un valor diferencial. La **figura 9** muestra la evolución del medio de pago utilizado a lo largo de nuestro ejercicio. Se observa un aumento constante del uso de las transacciones online y un descenso marcado del uso del efectivo.

## 5. LA ECONOMÍA EN ALTA DEFINICIÓN: LA DISTRIBUCIÓN DEL CONSUMO EN LAS CUENTAS NACIONALES

Una vez bien procesados y clasificados según los principios de contabilidad nacional, los datos de transacciones financieras logran una alta correspondencia con las cuentas nacionales publicadas por el INE (Buda *et al.*, 2022). Esto supone que, además de utilizar estos datos para el seguimiento del consumo en tiempo real, los microdatos subyacentes pueden ser utilizados para construir cuentas nacionales distributivas del consumo. Es decir, pueden ser utilizados para describir la distribución de los niveles de consumo y sus categorías por diversas variables como nivel de consumo, renta, edad... incluyendo cuestiones relacionadas como la desigualdad y su distribución en el tiempo en línea con los trabajos desarrollados por Piketty *et al.* (2018).

Desde el trabajo seminal de Piketty, Saez, Zucman (2018), ya existen cuentas nacionales distributivas de la renta para un gran número de países incluido España. Esta metodología es macroconsistente en el sentido que combina los agregados de las cuentas nacionales existentes, los censos, las encuestas de hogares y los microdatos de los impuestos sobre la renta. Sus trabajos han tenido una gran repercusión en los debates académicos y públicos sobre la desigualdad de ingresos y su evolución en el tiempo.

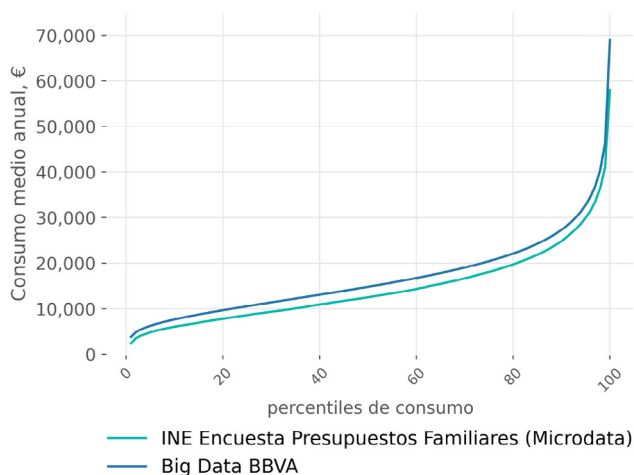
Sin embargo, y hasta donde sabemos, las cuentas nacionales de distribución del consumo son prácticamente inexistentes. Nuestras cuentas nacionales distributivas del consumo captan el 100 % del consumo agregado, lo que nos permite calcular el consumo para cada percentil de la distribución del consumo de forma coherente con los agregados macroeconómicos.

La **figura 10** representa la distribución transversal del consumo a través de los percentiles de consumo en 2017, con el eje Y representando el consumo medio anual de un adulto español en un percentil determinado de la distribución de consumo. El consumo medio de los adultos derivado de esta distribución coincide con la cifra de consumo agregado español por adulto, que en 2017 se situaba en 16.907 euros (56 % del PIB por adulto en España). Esta distribución es además muy similar cuando se compara con la distribución equivalente de la *Encuesta de Presupuestos Familiares Española (EPF)*.

Un primer indicio de la magnitud de la desigualdad en el consumo viene dado entonces por el hecho de que el adulto medio español en 2017 consumió 14.971 euros, un 12 % menos

Figura 10.

### Distribución del consumo en España: *big data* vs. Encuesta de Presupuestos Familiares



que el consumidor medio en España. La desigualdad del consumo se ve reforzada al observar directamente las colas de la distribución del consumo. Así, el consumidor adulto típico en el percentil 90 de la distribución del consumo de 2017 consumió aproximadamente dos veces más que el consumidor medio (28.115 euros). En otras palabras, la conocida relación entre los percentiles 90 y 50 ( $p_{90}/p_{50}$ ) en 2017 es de 1,87. Si nos acercamos más a la cola de la distribución del consumo, el consumidor medio del 1 % superior consumió unos 68.893 euros, lo que implica una relación  $p_{99}/p_{50}$  de 4,6. En el extremo opuesto, un adulto típico en el percentil décimo inferior de la distribución de consumo de 2017 consumió sólo 7.869 euros, aproximadamente la mitad del consumidor adulto mediano y 3,6 veces menos que el 10 % superior entre adultos.

Hasta ahora, hemos centrado nuestra atención en la distribución del consumo total por adulto. Sin embargo, la elevada granularidad de nuestro trabajo nos permite desglosar los resultados del consumo por las principales categorías de consumo oficiales (COICOP) como hemos mostrado en secciones anteriores.

Podemos obtener un resumen de la información de la desigualdad por tipo de consumo agrupando las categorías en dos grandes subgrupos: necesidades básicas y bienes de lujo<sup>2</sup>.

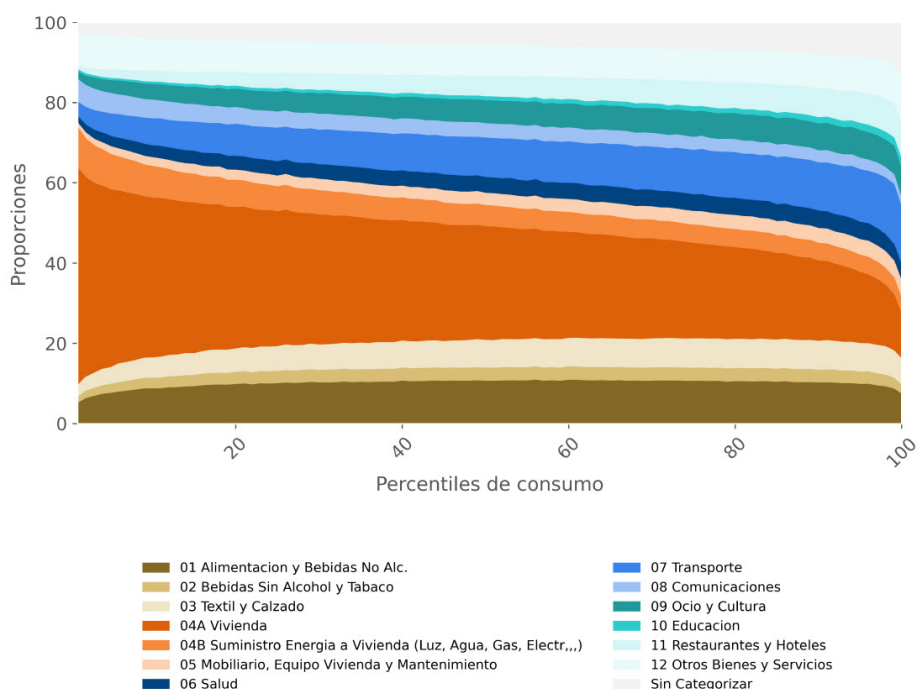
<sup>2</sup> Para ello Incluímos en las necesidades de consumo los alimentos y bebidas no alcohólicas (categoría 01 de la COICOP), el alcohol y el tabaco (02), el vestido y el calzado (03), el gasto en vivienda y suministros (4A y 4B) y la salud (06). Por el contrario, incluimos en el gasto de bienes de lujo, el menaje de hogar (05), el transporte (07), las comunicaciones (08), el ocio y la cultura (09), la educación (10), los restaurantes y hoteles (11), los bienes y servicios diversos (12) y la categoría residual no clasificada, los gastos no clasificados. Procedemos a nuestro análisis basándonos en estas agrupaciones.



La **figura 11** muestra la distribución del consumo por percentiles y categorías durante el año 2017. Si tenemos en cuenta en primer lugar la cuota de gasto en necesidades (representado por tonalidades marrones y naranjas), éstas constituyen el 57,4 % del consumo total del adulto medio en la distribución del consumo. En consonancia con el concepto de necesidades, este porcentaje disminuye fuertemente a lo largo de la distribución del consumo, representando el 67 % del consumo total de los adultos del 10 % inferior, el 49 % del consumo total del 10 % superior y sólo el 29 % del 0,1 % superior. Un resultado importante es que, aunque el consumo total de artículos de primera necesidad aumenta con el consumo total (como se desprende del panel (a) para los niveles de gasto), la desigualdad de consumo implícita derivada del consumo de artículos de primera necesidad es algo menor que la del consumo total. Por ejemplo, la relación p90-p50 es de 1,60 (en relación con el 1,87 del consumo total) y la proporción del 10 % superior del consumo agregado de artículos de primera necesidad es de 0,19 (en relación con el 0,22 del consumo agregado total).

Figura 11.

### Distribución del consumo por percentiles y categorías en 2017



Por el contrario, la distribución del consumo de bienes de lujo es muy desigual. En efecto, el 50 % inferior de la distribución del consumo sólo representa el 24 % del gasto total en lujo, mientras que el 10 % superior representa un desproporcionado 30 %. Como era de

esperar, el consumo de bienes de lujo se concentra en la parte superior y, por ejemplo, representa el 71 % del consumo del adulto medio en el 0,1 % superior de la distribución del consumo. En resumen, y complementando toda esta información con el índice de Gini como medida para resumir la desigualdad en la distribución, los resultados presentados implican que el consumo de bienes de lujo está distribuido un 38 % de manera más desigual que el consumo total.

Otra de las ventajas de los datos de las transacciones financieras es que incluyen información adicional sobre cada consumidor (edad, género, ingreso, provincia...), por lo que podemos presentar estas cuentas distributivas del consumo desagregadas por características demográficas.

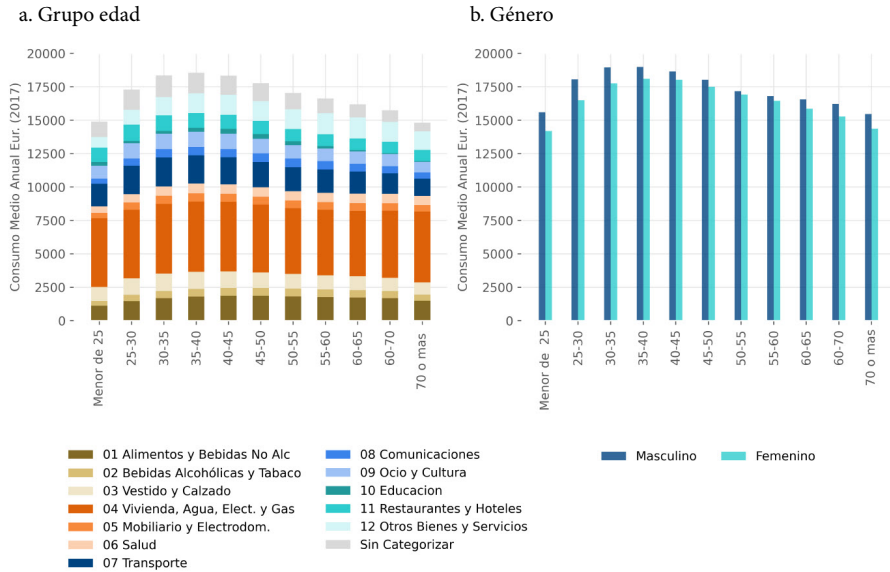
Como ejemplo las [figuras 12a y 12b](#), que muestran la distribución del consumo de 2017 por grupos de edad y género. En la primera, el eje Y indica tanto el consumo total medio de los adultos en una categoría de edad determinada como su desglose entre las categorías de consumo *COICOP*. La pauta de consumo en forma de U invertida reflejada en la literatura<sup>3</sup> es también evidente en nuestra muestra, en la que observamos que el consumo de los adultos crece a lo largo de los 20 y los 30 años, alcanza su máximo en la mediana edad y disminuye suavemente a partir de entonces. Es decir, a lo largo del ciclo vital se observa un aumento del 20 % en el consumo desde la juventud hasta la mediana edad, seguido de un descenso del mismo tamaño en el consumo hasta la vejez. Dada la elevada granularidad de los datos es posible explorar los ricos metadatos asociados a cada transacción para obtener una contabilidad distributiva de esta U invertida por edad y categorías de consumo. En particular, se confirma que el descenso del consumo después de la edad media es en parte el resultado de un descenso en el consumo de restaurantes y hoteles, transporte y, en menor medida, ropa y calzado. A diferencia de los resultados de Aguiar y Hurst (2013) para EE. UU., también encontramos un papel importante para la disminución de los gastos de educación, de ocio y cultura.

La información nos permite extender el análisis a la diferencias de género ([figura 12b](#)). Aquí, el consumo agregado muestra una brecha de género del 6 % en el consumo en 2017. El hombre adulto medio en España consumió 17.390 euros, mientras que la mujer adulta media consumió aproximadamente 1.000 euros menos, con 16.399 euros. Curiosamente, esta diferencia no es constante a lo largo del ciclo vital de hombres y mujeres. Tanto los hombres como las mujeres muestran un claro perfil de ciclo vital que alcanza su punto máximo en la mediana edad, pero la brecha de consumo entre ambos sexos es mayor para los que tienen entre 20 y 30 años, y luego disminuye lentamente hasta alcanzar un mínimo casi paritario para los que tienen entre 50 y 55 años, mientras que vuelve a abrirse a partir de los 60<sup>4</sup>.

<sup>3</sup> Véase Aguiar y Hurst (2013); Fernández-Villaverde y Krueger (2007), entre otros.

<sup>4</sup> Este hecho es coherente con una penalización de la brecha de ingresos de género ampliamente documentada debido a las interrupciones de la trayectoria profesional durante las edades típicas de maternidad: véase, por ejemplo, Guvenen, Kaplan y Song (2020) como estudio reciente.

Figura 12.

**Distribución del consumo en España por edad y género (2017)****6. UTILIZACIÓN DEL *BIG DATA* PARA POLÍTICA ECONÓMICA**

Las posibilidades de obtener la información *big data* de las transacciones financieras en tiempo real y alta definición convierten a este tipo de datos en idóneos para ser utilizados en materia de política económica. En particular, hay una tendencia creciente durante los últimos años a utilizar estos datos como indicadores para el seguimiento en tiempo real de la actividad económica, el análisis de la heterogeneidad y desigualdad y el análisis de inferencia causal para la evaluación de política económica.

La inmediata disponibilidad de los datos favorece una rápida respuesta en términos de política económica, tal como se pudo apreciar durante el COVID-19, sobre todo en el primer y segundo trimestre de 2020. El uso de esta información ofrece la oportunidad de extraer señales económicas casi en tiempo real y anticipar series económicas antes de que se publiquen las cifras oficiales.

La información obtenida de las transacciones financieras se ha utilizado bien para evaluar directamente la actividad económica en tiempo real. En el caso de España tanto Carvalho *et al.* (2021) como Mestre *et al.* (2022) mostraron ya en 2020 como utilizar los datos de transacciones con tarjetas para el seguimiento de la actividad económica. Más recientemente, Buda *et al.* (2022) muestra cómo se puede hacer un seguimiento en tiempo real no sólo del consumo de tarjetas (cerca de un 35 % del total) sino del resto de las transacciones comple-

tando así el 100 % del consumo y sus categorías oficiales (que se publican con más de un año de retraso oficialmente). La información de las transacciones puede ser introducida también en modelos de *nowcasting* junto a otras variables.

El análisis de la heterogeneidad y las cuentas distribucionales de la renta o el consumo también se ha visto reforzado con los datos de las transacciones financieras. Este puede constituir un elemento importante dentro del campo del análisis de la desigualdad de la renta (Mestre *et al.*, 2022 y Buda *et al.*, 2022) y del consumo (Buda *et al.*, 2022). Este último trabajo (Buda *et al.*, 2022) muestra un elevado grado de consistencia de la distribución de la renta con las nuevas cuentas distribucionales de la renta GRID para la economía española y un elevado grado de ajuste en la desigualdad del consumo con la obtenida en la encuesta oficial de presupuestos familiares. La alta granularidad de los datos también ha sido utilizada para analizar el impacto de la inflación en la renta de las familias a través de diferentes canales (Cardoso *et al.*, 2022).

## 7. CONCLUSIONES

Los abundantes datos de transacciones financieras que se producen de forma natural pueden utilizarse a un coste relativamente bajo para generar información compleja, cuidada, precisa y exhaustiva sobre la actividad económica. Nuestro trabajo aboga por el uso de estos datos tanto para la construcción de cuentas nacionales agregadas y su utilización en tiempo real, así como la información que se genera en alta definición en forma de cuentas distributivas para el estudio de la microestructura de la actividad económica.

La disponibilidad en tiempo real de la información del consumo total (55 % de la economía del PIB español) constituyen un avance notable para el campo de la coyuntura económica. No menos importante, es el hecho de que las doce categorías oficiales también están disponibles en tiempo real, cuando son oficialmente publicadas una vez al año.

Además, las buenas propiedades de agregación de los datos permiten un análisis distributivo del consumo agregado, proporcionando una descripción rica y macroconsistente de la desigualdad del consumo y su evolución temporal. Todo esto es posible porque los datos de las transacciones, una vez organizados adecuadamente, pueden desplegarse como una encuesta de consumo de alta calidad, a gran escala y en tiempo real, que contiene tanto información sobre las decisiones de consumo de millones de individuos como ricos metadatos que etiquetan miles de millones de transacciones.

Evidentemente, hay una plétora de covariables adicionales dentro de los datos de transacciones financieras que pueden utilizarse para ir más allá del consumo y hacer un mejor seguimiento de otras áreas de la actividad económica. Por ejemplo, los datos de las transacciones incluyen una gran cantidad de información a nivel de empresa que puede utilizarse para comprender la actividad de producción.

Creemos que también es posible analizar más a fondo el funcionamiento del sector exterior y la actividad gubernamental, o generar tablas de insumo-producto regionales y sectoriales, un objeto clave para las cuentas nacionales. Por último, volviendo al consumo, a nivel microeconómico también es posible aumentar nuestro panel de consumo con otras covariables (principalmente la renta y la riqueza), lo que permite comprender mejor las decisiones de consumo-ahorro y ayudar al desarrollo de la teoría económica.

## Referencias

- ANDERSEN, A. L., HANSEN, E. T., JOHANNESSEN, N. y SHERIDAN, A. (2020). Consumer responses to the covid-19 crisis: Evidence from bank account transaction data. *CEBI Working Paper Series*, No. 1820.
- ASPACHS, O., DURANTE, R., GRAZIANO, A., MESTRES, J., MONTALVO, J. G. y REYNAL-QUEROL, M. (2002). Real-time inequality and the welfare state in motion: evidence from COVID-19 in Spain. *Economic Policy*, Volume 37, Issue 109, January 2022, pp. 165–199.
- BAKER, S. y KUENG, L. (2022). Household Financial Transaction Data. *Review of Economics* 14(1), pp.47-67.
- BARLAS, A. B., GULER, S., ORTIZ, A. y RODRIGO, T. (2020). Investment in real time and High definition: A Big data Approach. *BBVA Research Working Papers*, No. 2013.
- BEAN, C. (2016). Independent review of UK economic statistics: Final report. Technical report.
- BOUNIE, D. CAMARA, Y. y GALBRAITH, J. (2020). Consumers' mobility, expenditure and onli- neoffline substitution response to covid-19: Evidence from french transaction data. *Hal working papers*, no. 02566443.
- BUDA, G., CARVALHO, V., HANSEN, S., ORTIZ, A., RODRIGO, T. y MORA, J. (eds) (2022). DP17519 National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. *CEPR Press Discussion Paper* No. 17519.
- CARDOSO, M., FERREIRA, C., LEIVA, J. M., ORTIZ, A., RODRIGO, T. y VAZQUEZ, S. (2022). The Heterogeneous Impact of Inflation on Households' Balance Sheets?. Mimeo.
- CARVALHO, V., GARCIA, J., HANSEN, S., ORTIZ, A., RODRIGO, T., RODRIGUEZ MORA, S. y RUIZ, P. (2021). Tracking the COVID-19 crisis with high-resolution transaction data. *Royal Society Open Science*, 8.
- CHETTY, R., FRIEDMAN, J. N., HENDREN, N. y STEPNER, M. (2020). How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data. *Working Paper*.
- COX, N., GANONG, P. NOEL, J. VAVRA, A., WONG, FARRELL, D., GREIG, F. y DEADMAN, E. (2020). Initial impacts of the pandemic on consumer behavior: Evidence from linked income, spending, and savings data. University of Chicago, Becker friedman institute for economics working paper.
- CHRONOPOULOS, D. K., LUKAS, M. y WILSON, J. O. (2020). Consumer spending responses to the covid-19 pandemic: An assessment of great britain. *Working papers in responsible banking and finance* no.20-012.
- EHRlich, G., HALTIWANGER, J. C., JARMIN, R. S., JOHNSON, D. y SHAPIRO, M. D. (2022). Re-engineering Key National Economic Indicators. En *Big Data for Twenty-First-Century Economic Statistics* (pp. 25–68). University of Chicago Press.
- EINAV, L. y KLENOW, P. (2021). Customers and retail growth. *Working Paper*, Stanford University, Stanford, CA.
- FERNANDEZ-VILLAVARDE, J. y KRUEGER, D. (2007). Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data. *The Review of Economics and Statistics*, 89(3), pp. 552–565.
- GATHERGOOD, J., LOEWENSTEIN, G., QUISPE-TORREBLANCA, E. G. y STEWART, N. (2019). The red, the black and the plastic: paying down credit card debt for hotels, not sofas. *Management Science*, 65(11), pp. 4951–5448.

- GUVENEN, F., KAPLAN, G. y SONG, J. (2020). The Glass Ceiling and the Paper Floor: Changing Gender Composition of Top Earners since the 1980s. En *NBER Macroeconomics Annual 2020*, Volume 35, pp. 309–373. University of Chicago Press.
- GUVENEN, F., PISTAFERRI, L. y VIOLANTE, G. (2022). Global Trends in Income Inequality and Income Dynamics: New Insights from GRID. *NBER Working Papers* 30524. National Bureau of Economic Research, Inc.
- HACIOGLU, S., KANZIG, D. R. y SURICO, P. (2020). Consumption in the time of covid-19: Evidence from uk transaction data. *CEPR discussion papers* no. 14733.
- PIKETTY, T., SAEZ, E. y ZUCMAN, G. (2018). Distributional National Accounts: Methods and Estimates for the United States. *The Quarterly Journal of Economics*, 133(2), pp. 553–609.