

CAPÍTULO IV

El análisis de la economía en tiempo real a partir de datos masivos de transacciones en cuentas bancarias

Josep Mestres Domènech

En este trabajo se describe la experiencia de CaixaBank Research en el análisis de datos masivos de transacciones financieras. Los movimientos de cuentas bancarias, de alta frecuencia y con un nivel de granularidad elevado, permiten obtener información económica de gran calidad y precisión de manera casi instantánea y con un enorme potencial para la investigación económica. Sin embargo, su análisis es complejo, puesto que se trata de un gran volumen de datos creado para finalidades distintas del análisis económico. El trabajo describe el proceso de aprendizaje a partir de los proyectos llevados a cabo, y en particular, la creación de un portal de Economía en Tiempo Real (<https://realtimeeconomics.CaixaBankresearch.com/#/home>). Este portal monitoriza la evolución de la economía española mediante 12 indicadores contruidos con datos internos de CaixaBank, agregados mediante técnicas de *big data*.

Palabras clave: análisis de datos masivos en tiempo real, *big data*, transacciones financieras.

1. INTRODUCCIÓN

Si bien hace ya unos años que los economistas han ido incorporando los datos masivos (*big data*) a sus análisis, la pandemia del COVID-19 supuso una verdadera revolución de la economía en tiempo real. Disponer de información al momento era imprescindible para gestionar la crisis y poder tomar decisiones rápidamente. Por ejemplo, gracias a los datos de [Google Mobility Report](#) se pudo medir el impacto del COVID-19 en la movilidad de las personas tras las restricciones sanitarias y gracias a los datos de Opportunity Insights (<https://www.tracktherecovery.org/>), el impacto económico de la pandemia en EE. UU. (Chetty *et al.*, 2020). En CaixaBank Research, junto con investigadores de la Universitat Pompeu Fabra y el Institute of Political Economy and Governance (IPEG), desarrollamos un monitor pionero en medir en tiempo real la desigualdad y el papel del estado de bienestar en España (Aspachs *et al.*, 2022).

Las fuentes de información en tiempo real se generan tanto en el sector privado (compañías de telefonía, internet, financieras, etc.) como en el seno de la Administración pública, la cual atesora datos de registro que son una fuente inigualable de información económica. La Agencia Tributaria, por poner un ejemplo, posee los datos más detallados y actualizados sobre la evolución de las empresas españolas gracias a las declaraciones de impuestos.

La información que se obtiene a partir de los datos masivos es complementaria a la que ofrecen datos más tradicionales, como las encuestas o los experimentos económicos, y no la sustituye. Sin embargo, los datos masivos ofrecen una mayor granularidad e inmediatez que las fuentes más tradicionales no pueden ofrecer. Al estar creados para otras finalidades, los datos masivos ya están disponibles (no hace falta elaborar una encuesta, por ejemplo, “solo” se tienen que procesar y analizar), con lo que la información económica resultante se cosecha a un menor coste, con mayor celeridad y puede llegar a proporcionar muestras más representativas y de mayor tamaño.

La economía en tiempo real puede jugar también un importante papel en el diseño e implementación de las políticas públicas. Los datos masivos favorecen un diseño más ajustado de las políticas económicas y que se focalicen en aquellos colectivos que más lo necesitan (lo que se conoce como *smart policies*). También permiten un seguimiento *a posteriori* de las medidas, una vez implementadas, y ajustarlas a tiempo real cuando sea necesario. Y todo ello a un menor coste y una mayor disponibilidad que otro tipo de datos.

Por todo ello, CaixaBank Research lleva desde 2018 desarrollando investigación económica con los datos masivos de transacciones financieras, debidamente anonimizados. CaixaBank es un banco idóneo para llevar a cabo estos análisis puesto que es un banco generalista y cuenta con la mayor cuota de mercado en España. Por ejemplo, la cuota de CaixaBank en 2023 en depósitos corresponde al 24,9 % del total nacional, y las cuotas de mercado en ingresos de clientes son superiores, como en nóminas (36,7 % del total) o

en pensiones domiciliadas (34,3 %)¹. Esto permite obtener una muestra representativa de la población española, cuestión de suma relevancia para la elaboración de indicadores económicos.

En este capítulo analizamos la experiencia de CaixaBank Research con el análisis de datos masivos a partir de la elaboración del portal de economía en tiempo real, explicando los obstáculos y limitaciones encontradas por el camino. Al ser parte de los estudios con los datos internos experimentales, no todos han conseguido su objetivo inicial, pero en todos los casos nos han ayudado a entender mejor los datos de transacciones y su potencial para analizar la economía española. El [portal de economía en tiempo real](#) es un proyecto pionero del equipo de CaixaBank Research² cuyo objetivo es monitorizar la evolución de la economía española a través de 12 indicadores contruidos con datos internos de CaixaBank, agregados mediante técnicas de *big data*. Los indicadores se agrupan en cinco ámbitos: el consumo, la vivienda, el turismo, los salarios y la desigualdad. Cada uno de estos indicadores se puede consultar por distintas dimensiones, lo cual permite identificar diferencias entre colectivos (edad, sexo, ingresos, sector de actividad), regiones (CC. AA., municipios), etc. En total, publicamos más de 800 series de datos que actualizamos mensualmente.

2. EL ANÁLISIS DE LA ECONOMÍA EN TIEMPO REAL A PARTIR DE DATOS MASIVOS: LA EXPERIENCIA DE CAIXABANK RESEARCH

Los movimientos de cuentas bancarias, de alta frecuencia y con un nivel de granularidad elevado, permiten obtener información económica de gran calidad y precisión de manera casi instantánea y con un enorme potencial para la investigación económica. Sin embargo, su análisis es complejo, puesto que se trata de un gran volumen de datos creado para finalidades distintas del análisis económico. Por ello, la utilización de estas fuentes de datos requiere múltiples condiciones previas.

En primer lugar, los datos masivos creados por parte de empresas privadas deben estar disponibles para su análisis estadístico, idealmente en un repositorio único que sea accesible para todos. Esta cuestión no es obvia para muchas empresas que no se dedican a la ciencia de los datos, y se deben invertir recursos para que esto sea posible.

En segundo lugar, este repositorio de datos debe ser accesible para poder analizarlo. En CaixaBank Research iniciamos en 2018 el análisis de datos masivos de transacciones financieras, momento en que se pusieron a disposición del Servicio de Estudios. Sin embargo, un repositorio único de datos existía desde muchos años atrás, aunque no estaba destinado a

¹ https://www.caixabank.com/deployedfiles/caixabank_com/Estaticos/PDFs/Accionistasinversores/Informacion_economico_financiera/231027_Webcast_3T23_es.pdf

² El portal ha sido un esfuerzo colectivo en el que han participado Eduard Alcobé, Oriol Aspachs, Patricia Esteban, Alberto Graziano, Javier Ibáñez de Aldecoa, Eduard Llorens, Zoel Martín, Josep Mestres y Judit Montoriol. El presente artículo describe este trabajo colectivo, aunque cualquier error u omisión es responsabilidad mía y no del resto del equipo.

fin de investigación económica. Pero a partir de 2018 se puede acceder a esta capacidad tecnológica. En tercer lugar, se necesita un equipo humano capaz de analizar estos datos. Para ello, a partir de ese año se complementó la formación de los economistas del equipo con cursos en *data science*, a la vez que se incorporan *data scientists* al equipo dedicados plenamente a proyectos de *big data*.

El proceso de aprendizaje en análisis de datos masivos ha sido un largo camino. El conocimiento sobre los datos mismos, a menudo en tablas independientes creadas para fines distintos, y la información contenida en ellas, se fue obteniendo a medida que llevábamos a cabo distintos proyectos de *analytics* en CaixaBank Research. Los primeros análisis con datos internos de CaixaBank permitieron observar, de forma siempre anonimizada, los patrones de gasto de los turistas extranjeros (Campos y Montoriol, 2019) o, en el ámbito inmobiliario, entender las decisiones de compra o alquiler de vivienda (Montoriol, 2020), entre otros análisis. Pero fue la pandemia del COVID-19 la que supuso un revulsivo para su uso. En 2020, en plena pandemia, urgía entender cómo estaba afectando el COVID-19 a los colectivos más vulnerables y cómo estaba respondiendo a la crisis el estado de bienestar. De esa urgencia surgió el Monitor de Desigualdad, germen del portal actual de Economía en tiempo real, impulsado por investigadores de la Universitat Pompeu Fabra, el Institute of Political Economy and Governance y CaixaBank Research. El monitor nos permitió analizar la evolución de millones de nóminas cada mes para obtener una estimación precisa y representativa de la distribución de los salarios tanto para el conjunto de la sociedad, como para los colectivos más vulnerables. Y también constatamos el papel que han jugado las transferencias del sector público para amortiguar el duro golpe que supuso la pandemia (Aspachs *et al.*, 2021). Por primera vez, se pudo monitorear la desigualdad en España con una frecuencia mensual y con un extraordinario nivel de detalle.

Conscientes de las dificultades que entraña la predicción económica y del papel de los datos de alta frecuencia para ayudar a despejarlas, especialmente en un contexto tan complejo como el de la crisis del coronavirus, desarrollamos, también en 2020, el Monitor de consumo³. Mediante este indicador logramos tomar el pulso a la actividad española, semana a semana, a partir de los pagos con tarjetas. De nuevo, millones de transacciones al mes.

A partir de entonces, el análisis de la situación económica en nuestras publicaciones se nutrió cada vez más de los datos internos, siempre debidamente anonimizados. Algunos ejemplos de análisis exitosos que hemos llevado a cabo se han beneficiado de un tipo de información particular, como la domiciliación de recibos o de la pensión⁴. Tras el estallido de la guerra en Ucrania, analizamos los recibos de electricidad de clientes particulares domiciliados en CaixaBank para determinar cómo afectó el aumento de los precios en la economía de las

³ <https://www.caixabankresearch.com/es/publications/monitor-consumo>

⁴ Para más detalle y una lista completa de publicaciones, ver https://www.caixabankresearch.com/es/tendencias-fondo/economia-tiempo-real?_gl=1*_1flsxbk*_ga*NjUyOTk5NDc1LjE3MDU1ODc5MjQ*_ga_0S8KEKC-3M2*MTcwNTY4ODM5MS4zNy4xLjE3MDU2ODkxMzAuMC4wLjA

familias españolas. Hemos utilizado también los datos internos para analizar los patrones de consumo y ahorro tras la jubilación para el caso español. La utilización de datos internos permitía conocer los ingresos, el consumo y el ahorro de los individuos con precisión, así como identificar el momento de la jubilación mediante el primer cobro de la pensión. Otros proyectos no han conseguido el objetivo inicial de análisis, pero han servido como aprendizaje y muchos de los elementos se han usado posteriormente, para mejorar geolocalización de las transacciones o la identificación de tipos específicos de transferencias, para dar algunos ejemplos.

Con todo este aprendizaje se ha conseguido llevar a cabo el portal de Economía en Tiempo Real. El portal define unos indicadores económicos de calidad, publicados casi a tiempo real y con granularidad. Los indicadores publicados están contruidos a partir de datos internos anonimizados de CaixaBank, sobre una base de más de 18 millones de clientes, 11.500 cajeros automáticos y 700.000 terminales de punto de venta. Se utilizan técnicas de *big data* para procesar millones de operaciones cada mes, lo que aporta una gran cantidad de información que permite observar la marcha de la economía española de manera prácticamente instantánea y con una alta precisión. Gracias a la digitalización y a la elevada penetración de CaixaBank en el territorio, los datos obtenidos son representativos del conjunto de la sociedad, lo cual permite detectar anticipadamente las nuevas tendencias e identificar diferencias entre colectivos y comunidades autónomas. El portal, pionero en España, fue lanzado en noviembre de 2022 y su acceso está abierto al público, al igual que el resto de las publicaciones de CaixaBank Research, para dar un servicio a la sociedad al crear y divulgar análisis económico y al democratizar el acceso a la información económica de calidad.

3. EL PORTAL DE ECONOMÍA EN TIEMPO REAL: CONSTRUCCIÓN Y REPRESENTATIVIDAD

Esta sección describe la metodología aplicada en la construcción de los indicadores publicados en el portal de Economía en Tiempo Real de CaixaBank Research. Los indicadores están contruidos a partir de los datos internos de CaixaBank y se agregan mediante técnicas de *big data*. El objetivo principal fue elaborar indicadores económicos de calidad, representativos del conjunto de la economía española, y con la granularidad e inmediatez temporal que los datos masivos permiten. Su frecuencia de publicación es mensual, con un decalaje de pocos días después del cierre del mes, con la excepción del indicador de la accesibilidad a la vivienda. En este caso, el indicador se publica en bloques de tres meses tras la publicación del precio de la vivienda del MITMA.

El portal de Economía en Tiempo Real de CaixaBank Research ofrece información sobre 12 indicadores, agrupados en cinco ámbitos: consumo, vivienda, salarios, turismo y desigualdad. Cada indicador se muestra para distintas categorías (comunidad autónoma, franjas de edad, franjas de ingresos, sector de actividad, etc.) y se publican unas 850 series en total (ver [cuadro 1](#) para más detalle).

Cuadro 1.

Indicadores disponibles en el portal de Economía en tiempo real

	<i>Indicadores</i>	<i>Definición</i>	<i>Dimensiones</i>
	Consumo Consumo total, consumo presencial y comercio electrónico	Consumo registrado a partir de pagos y reintegros con tarjetas de débito y de crédito tanto españolas como extranjeras	Nacional, CC. AA., edad, nivel de ingresos, tipo de gasto
	Vivienda Accesibilidad a la vivienda	Número de años de ingresos laborales netos que el hogar mediano debe destinar a la compra de una vivienda en una zona geográfica específica	Nacional, CC. AA. y capital de provincia
	Salarios Salarios	Ingresos salariales mensuales calculados a partir de las nóminas domiciliadas en las cuentas de CaixaBank	Nacional, CC. AA., edad, género, sector de actividad, sector público/privado
	Turismo Turismo doméstico, internacional y gasto exterior	Gasto registrado a partir de pagos y reintegros con tarjetas españolas o extranjeras	Nacional, edad, país, tipo de gasto
	Desigualdad Índice de Gini, percentiles de ingresos, distribución de ingresos, curva de Lorenz	Desigualdad salarial definida a partir de las nóminas domiciliadas así como de los cobros de la prestación y del subsidio de paro	Nacional, CC. AA., edad, género, país de nacimiento

Nota: Para mayor detalle sobre la construcción de las variables y su representatividad, véase la Nota Metodológica (https://www.caixabankresearch.com/es/nota-metodologica_rte).

Fuentes: CaixaBank Research, portal de Economía en tiempo real (<https://realtimeeconomics.caixabankresearch.com>).

3.1. Consumo

Los indicadores de consumo se refieren al consumo nominal (no deflactado) y contiene tres conjuntos de indicadores: consumo total, consumo presencial y comercio electrónico. Para el consumo doméstico, se considera una muestra cerrada de clientes de CaixaBank, mientras que para los extranjeros no se utiliza ningún filtro en la muestra. Se presenta la variación interanual del indicador, que recoge la mayor parte de la estacionalidad de la serie (no se realiza ningún ajuste adicional por los efectos de calendario).

El consumo total incluye el gasto y los reintegros con tarjetas de crédito y débito emitidas por CaixaBank, y gasto con tarjetas extranjeras en TPV CaixaBank y reintegros en cajeros de CaixaBank. El consumo total engloba el consumo presencial con tarjeta, las retiradas de efectivo y el comercio electrónico (*e-commerce*).

El consumo presencial doméstico incluye el gasto presencial y reintegros con tarjetas de crédito y débito emitidas por CaixaBank y solo se consideran personas físicas. Entre las categorías disponibles, el gasto se puede visualizar para el total nacional y para la C. A. del TPV, por franja de edad (16-29 años, 30-49 años, 50-64 años, 65 años y más), franja de ingre-

tos (ingresos bajos, medios y altos) y por sector de actividad. Los sectores de actividad están clasificados a partir del sector del comercio propietario del TPV de la siguiente manera:

- Primera necesidad (alimentación, farmacias).
- Bienes duraderos (tecnología, menaje, textil, deportes, jugueterías y joyerías).
- Ocio y restauración (restaurantes, casinos, clubs deportivos, espectáculos, museos, parques temáticos).
- Transporte (transporte de viajeros, autopistas, parkings y gasolineras).
- Turismo (hoteles, agencias de viajes y alquiler de vehículos).

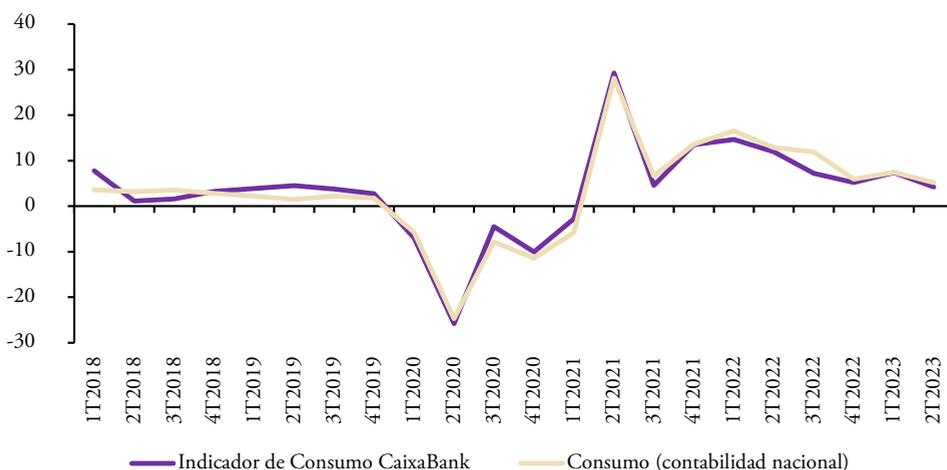
El comercio electrónico doméstico incluye el gasto con tarjetas de crédito y débito emitidas por CaixaBank en un TPV virtual. Solo se consideran personas físicas y se muestra la media móvil de los últimos dos meses.

La comparativa del indicador de consumo con una referencia externa (figura 1) muestra cómo la suma de consumo presencial doméstico y comercio electrónico doméstico puede considerarse una buena aproximación de la serie de gasto en consumo final de los hogares (nominal y no ajustada de estacionalidad y calendario) publicada por el INE en la contabilidad nacional trimestral de España. La figura 1 muestra la comparativa en términos interanuales y puede apreciarse una correlación entre ambas series muy elevada, del 0,98.

Figura 1.

Comparativa del indicador de consumo con una referencia externa

Variación interanual (%)



Fuentes: CaixaBank Research, a partir de datos internos de CaixaBank y del INE.

3.2. Acceso a la vivienda

En la sección de vivienda, el indicador clave es la *ratio de accesibilidad a la vivienda* en propiedad. Esta ratio mide el número de años de ingresos laborales netos que el hogar mediano debe destinar a la compra de una vivienda en la zona geográfica i . Un valor más alto indica mayor dificultad para acceder a una vivienda en propiedad. Siguiendo la metodología del Banco de España, se calcula con la siguiente expresión:

$$\text{Ratio Accesibilidad}_i = \frac{PMC_i * 93,75}{\text{mediana ingresos}_i * 1,6} \quad [1]$$

Donde

- PMC_i = precio de la vivienda por metro cuadrado en la zona geográfica i (media móvil de 12 meses).
- 93,75 m² es la superficie de referencia que usa el Banco de España para calcular la ratio de accesibilidad a nivel nacional.
- La mediana de los ingresos laborales (nóminas y subsidio de paro) por persona en la zona geográfica i (media móvil de 12 meses para reducir la estacionalidad por las pagas extras).
- 1,6 es el número promedio de perceptores de ingresos en un hogar (EPA).

Los datos del precio de la vivienda (PMC) provienen del Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA) y tienen frecuencia trimestral. Para pasarlos a frecuencia mensual se replica el mismo dato de PMC para todos los meses del trimestre, y se calcula una media móvil de 12 meses del PMC para suavizar la serie. En consecuencia, los datos de la ratio de accesibilidad se publican en el portal de Economía en tiempo real en bloques de tres meses tras la publicación del dato del MITMA. Las categorías disponibles incluyen el total nacional, CC. AA. y capitales de provincia (68 series).

La comparativa del indicador de accesibilidad a la vivienda con una referencia externa se realiza con la ratio de accesibilidad a la vivienda que publica del Banco de España (BdE) a nivel nacional (figura 2)⁵. El BdE usa el precio de la vivienda del MITMA hasta 2005; luego se enlaza con la variación interanual de Tinsa para los años 2005 y 2006, y a partir de 2007 se enlaza con la variación interanual del IPV del INE. El indicador CaixaBank usa el precio del MITMA por la disponibilidad de datos de PMC por capitales de provincia.

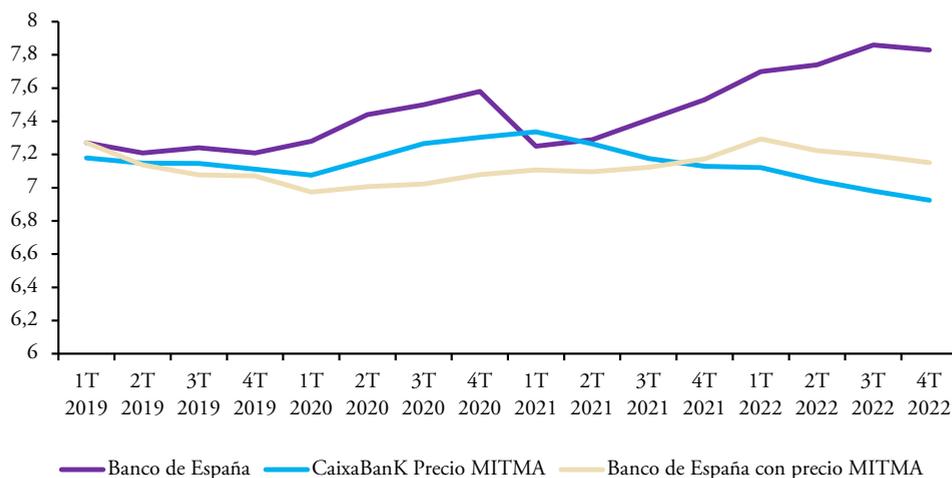
Para los ingresos, el BdE usa la renta bruta disponible del hogar mediano procedente de la *Encuesta Financiera de las Familias* (trianual) y extrapolando el resto de los años a partir de otras fuentes estadísticas. El indicador CaixaBank usa la mediana de los ingresos (nóminas y subsidios de paro) por persona y en términos netos (ingreso a cuenta del trabajador).

⁵ El Banco de España publica trimestralmente la ratio de accesibilidad en la Síntesis de Indicadores, cuadro 1.5: <https://www.bde.es/webbde/es/estadis/infoest/sindi.html>

Los datos del BdE tienen frecuencia trimestral. Para la comparación con el indicador creado a partir de datos internos de CaixaBank, escogemos el último mes del trimestre (serie suavizada). La siguiente figura muestra tres series de datos: 1) la ratio de accesibilidad CaixaBank calculada a partir de la mediana de los ingresos (datos internos) y el precio de la vivienda del INE; 2) el indicador CaixaBank de la ratio de accesibilidad, calculado a partir de la mediana de los ingresos (datos internos) y el precio de la vivienda del MITMA; y (3) la ratio de accesibilidad del BdE. El primer indicador CaixaBank es el directamente comparable con el BdE por usar el mismo precio de la vivienda en los años recientes (INE). Su correlación es de 0,97. Sin embargo, el indicador CaixaBank publicado en el portal usa el precio de la vivienda del MITMA por tener mayor detalle geográfico (municipios). En este caso, la correlación con la ratio de accesibilidad del BdE es positiva, pero más débil (0,13) debido al menor crecimiento del precio del MITMA en comparación con el precio del INE en el periodo considerado.

Figura 2.

Comparativa de la ratio de accesibilidad con una referencia externa (Banco de España)



Fuentes: CaixaBank Research, a partir de datos internos de CaixaBank, del Banco de España y del MITMA.

3.3. Salarios

Los salarios se miden en términos nominales (no deflactados) a partir de los ingresos salariales mensuales de personas físicas desde los movimientos en la cuenta corriente identificados específicamente como nómina. Se incluyen todos los ingresos de nómina acumulados en el mes para el cálculo de un salario mensual. Se consideran todos los clientes con nómina presente en el mes anterior como criterio de estabilización. El indicador de salarios corresponde a la mediana de la variación interanual de la nómina mensual, que se calcula cliente a cliente. El indicador corresponde a la media móvil de dos meses.

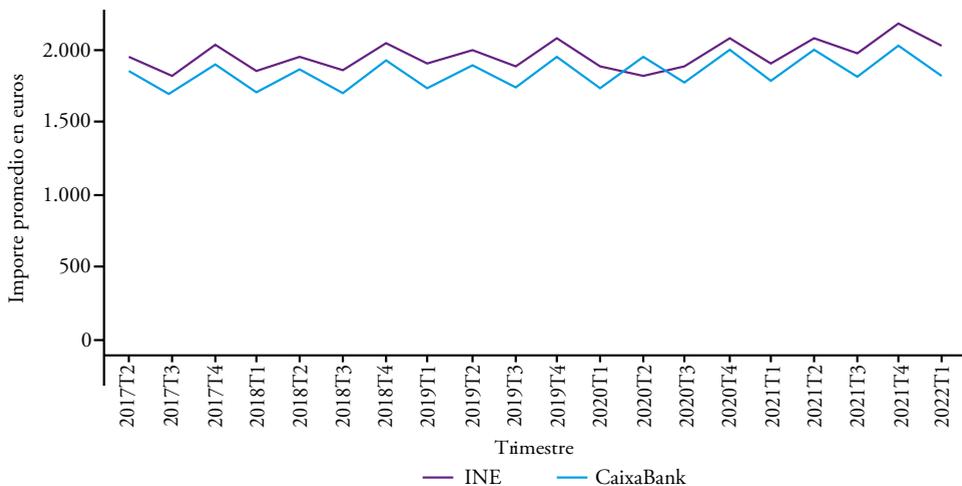
Las categorías disponibles corresponden a ámbitos geográficos (total nacional y CC. AA.), franja de edad, género, sector público/privado y cinco sectores de actividad:

- a. Agricultura (CNAE A)
- b. Industria (CNAE B-E)
- c. Construcción (CNAE F)
- d. Servicios turísticos (CNAE a dos dígitos 49-51, 55, 56, 77, 79, 90, 91, 93)
- e. Servicios no turísticos (CNAE G a J y L a U, excluyendo a aquellos incluidos en los servicios turísticos).

La primera referencia externa para comparar el indicador interno de salarios corresponde al coste salarial total por trabajador de la *Encuesta Trimestral de Coste Laboral (ETCL)* del INE (figura 3). Este coste salarial es el que soportan los empleadores para emplear a su plantilla (excluye otros costes laborales como las cotizaciones sociales, etc.) y está expresado en términos brutos (al contrario que el indicador interno de salarios, que está expresado en

Figura 3.

Comparativa del indicador de salario con una referencia externa (ETCL)



Nota: El coste salarial total por trabajador de la *Encuesta Trimestral de Coste Laboral (ETCL)* del INE corresponde al coste salarial promedio que los empleadores soportan para emplear a su plantilla (excluye otros costes laborales como las cotizaciones sociales, etc.) y está expresado en términos brutos. El indicador interno de salarios está expresado en términos netos y la comparativa con el INE realizada a partir del promedio.

Fuentes: CaixaBank Research, a partir de datos internos de CaixaBank y de la *Encuesta Trimestral de Coste Laboral (ETCL)* del INE.

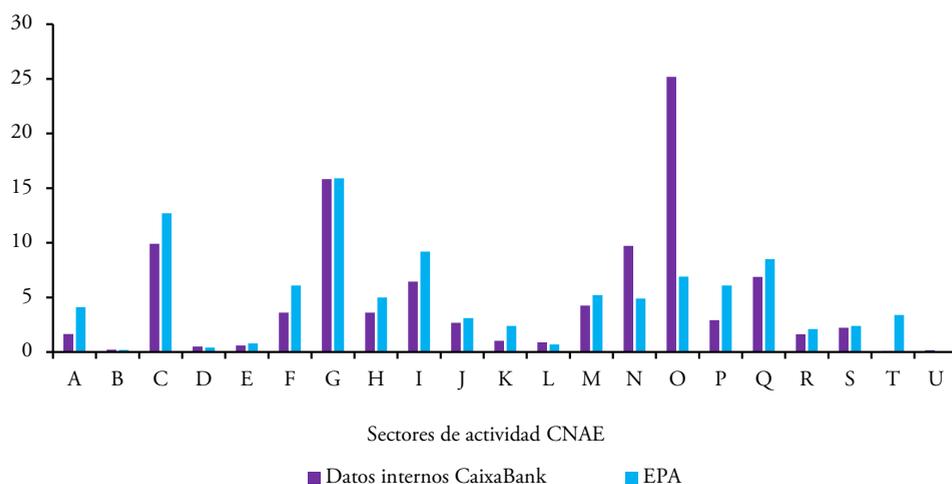
términos netos). En la figura se puede observar la similitud entre las series en el agregado para España, tanto en el nivel como en la estacionalidad (con la excepción del segundo trimestre de 2020).

Asimismo, se puede comprobar la representatividad de los datos internos de nóminas de CaixaBank con la *Encuesta de Estructura Salarial (EES)* en la sección de desigualdad (sección 3.4.). Otra comparativa realizada corresponde a la representatividad sectorial a nivel de sección CNAE entre los datos internos y la *Encuesta de Población Activa*. La **figura 4** muestra que la distribución es muy similar en la mayoría de los sectores de actividad. Las dos únicas excepciones corresponden al sector público (grupo CNAE O), en el que la muestra de CaixaBank está sobrerrepresentada por su mayor cuota de mercado, y el sector de empleadas del hogar (grupo CNAE T), en el que la muestra interna está infrarrepresentada puesto que a menudo la retribución se paga en efectivo.

Figura 4.

Comparativa de la distribución sectorial de los asalariados entre la muestra interna y la EPA

Peso del número de asalariados en cada sector sobre el total (%)



Fuentes: CaixaBank Research, a partir de datos internos de CaixaBank y de la *Encuesta de Población Activa (EPA)* del INE correspondientes al 3T 2022.

3.4. Turismo

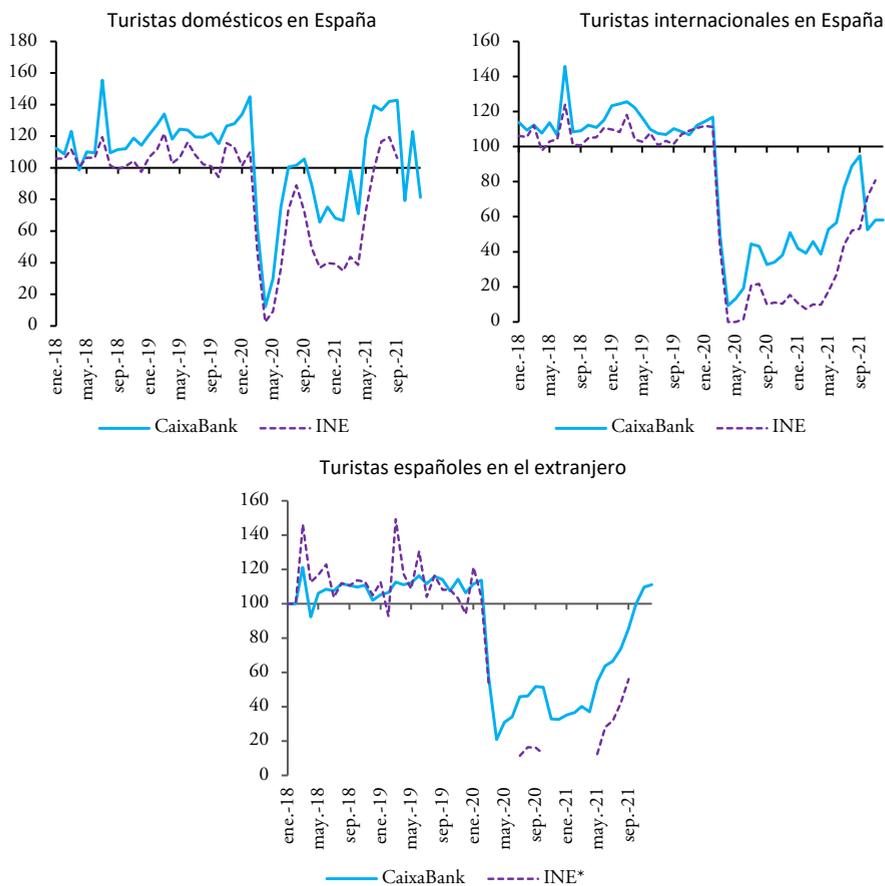
El bloque de turismo contiene tres indicadores de gasto nominal (no deflactado): turismo doméstico, turismo internacional y gasto exterior (gasto turístico de españoles en el extranjero). Se aplica una media móvil de tres meses sobre la serie de gasto y, a continuación, se calcula la tasa de variación interanual de la serie suavizada.

El turismo doméstico (gasto turístico doméstico) se identifica mediante los pagos presenciales y retiradas de efectivo realizados por los turistas españoles en España. Consideramos como gasto de los turistas domésticos los pagos en TPV de CaixaBank con tarjetas emitidas por entidades españolas efectuados fuera de sus áreas habituales de consumo (área de residencia, trabajo, consumo rutinario, etc.). Las categorías disponibles corresponden a distintos ámbitos geográficos (total nacional y CC. AA.) y distintos sectores de actividad (alojamiento, comercio y ocio y restauración).

Figura 5.

Comparativa de los niveles de gasto según destino y origen

Índice (100 = mismo mes de 2017)



Nota: El INE registra gasto nulo para turistas internacionales en España en los meses de abril y mayo de 2020, y no publica resultados de gasto de españoles en el extranjero para 8 meses de 2020 y 2021, debido al reducido número de encuestas realizadas.

Fuentes: CaixaBank Research, a partir de datos internos de CaixaBank y del INE.

El turismo internacional (gasto turístico internacional en España) corresponde a los pagos presenciales y retiradas de efectivo realizados por los turistas internacionales en España. Consideramos como gasto de turistas internacionales los pagos efectuados en España en TPV de CaixaBank con las tarjetas emitidas por entidades extranjeras. Las categorías disponibles corresponden a ámbitos geográficos (total nacional y CC. AA.), país de origen (Alemania, Estados Unidos, Francia, Reino Unido, otros de Europa y otros países) y sectores de actividad (alojamiento, comercio y ocio y restauración).

El gasto exterior (Gasto turístico español en el extranjero) se identifica mediante los pagos presenciales y retiradas de efectivo con tarjetas de crédito y débito emitidas por CaixaBank realizados en el extranjero. Las categorías disponibles corresponden a ámbitos geográficos (total nacional y CC. AA. de residencia del turista), franja de edad, franja de ingresos y país/región de destino.

Las referencias externas para comparar los indicadores de CaixaBank son los indicadores de gasto turístico publicados por el INE en las encuestas *EGATUR* (encuesta de gasto de turistas internacionales) y *FAMILITUR* (encuesta de turismo doméstico). En la **figura 5** puede apreciarse la similar dinámica de las series del INE y los indicadores calculados con datos internos de CaixaBank, si bien se observa un sesgo positivo en el nivel capturado por estos últimos. Este sesgo puede deberse a multitud de factores, si bien estimamos que los principales son el uso cada vez más predominante de los pagos con tarjeta, la inclusión de tarjetas de extranjeros residentes en España y la mejor representación del turismo doméstico de cercanía (excursiones o estancias muy cortas) y el turismo de frontera. Este turismo se mostró considerablemente más resiliente a partir de marzo de 2020, con el inicio de la pandemia.

3.5. Desigualdad

El bloque de desigualdad contiene cuatro indicadores que miden la desigualdad de ingresos salariales: el índice de Gini, los percentiles de ingresos, las distribuciones de ingresos y las curvas de Lorenz.

El indicador del índice de Gini corresponde a la variación del índice de Gini⁶, en puntos porcentuales, respecto a febrero de 2020, corregida por la variación estacional promedio experimentada durante el mismo periodo en 2018 y 2019. Se presentan dos series del índice de Gini, que se calculan a partir de la distribución mensual de los ingresos salariales antes y después de las transferencias públicas. La diferencia entre los índices antes y después de las transferencias públicas refleja en qué medida los esquemas de soportes de rentas (como pueden ser las prestaciones por desempleo o las que reciben los trabajadores que se encuentran en un ERTE) reducen la desigualdad salarial. Las categorías disponibles corresponden a ámbitos geográficos (total nacional y CC. AA.), franja de edad, género y país de nacimiento (España, extranjero).

⁶ El índice de Gini es el índice de referencia para analizar el nivel de desigualdad salarial. Puede tomar valores entre 0 y 100. Cifras más elevadas reflejan niveles de desigualdad más altos, y viceversa.

Los percentiles de ingresos se definen como el ratio entre dos percentiles de la distribución de ingresos después de las transferencias del sector público: 1) ratio entre los percentiles 80 y 20 (p80/p20); 2) ratio entre los percentiles 80 y 50 (p80/p50), y 3) ratio entre los percentiles 50 y 20 (p50/p20). En este caso, no se presentan las ratios de percentiles antes de las transferencias públicas. Las categorías disponibles corresponden a ámbitos geográficos (total nacional y CC. AA.) y género.

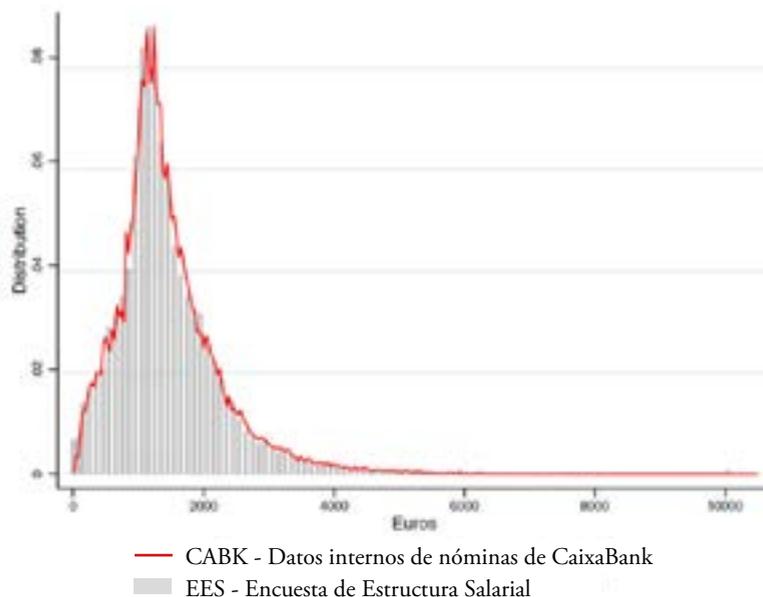
Las distribuciones de ingresos corresponden a la distribución mensual de los ingresos salariales antes y después de las transferencias del sector público. Las franjas representan el porcentaje de población con “ingresos bajos”, que se refiere a salarios o transferencias inferiores a 1.000 euros/mes; “ingresos medios”, entre 1.000 y 2.000 euros/mes; “ingresos altos”, más de 2.000 euros/mes; y “sin ingresos”. Las categorías disponibles corresponden a ámbitos geográficos (total nacional y CC. AA.), franja de edad, género y país de nacimiento (España, extranjero).

Las curvas de Lorenz corresponden a la distribución cumulativa de los ingresos salariales a lo largo de la población. Cada punto de la curva de Lorenz muestra qué porcentaje

Figura 6.

Distribución de salarios mensuales netos en España

Frecuencia (%)



Nota: Para facilitar la comparación entre muestras, ajustamos la distribución salarial de la ESS de 2018 por el aumento salarial promedio entre 2018 y 2019.

Fuentes: CaixaBank Research, a partir de datos internos de CaixaBank y de la *Encuesta de Estructura Salarial* (INE).

acumulado de los ingresos tiene un determinado porcentaje acumulado de la población. La curva empieza en el origen (0,0) y termina en el punto (100,100). Si los ingresos salariales estuvieran distribuidos de manera igual en toda la población, la curva correspondería con la línea de 45 grados. Si una persona dispusiera de todos los ingresos, la curva correspondería al eje horizontal hasta el punto (100,0), donde pasaría al punto (100,100). Cuanto más cerca está la curva de la perfecta igualdad, más equitativa es la distribución de ingresos (es decir, menor desigualdad existe), y viceversa.

Para confirmar la representatividad de los datos internos de CaixaBank con datos externos, utilizamos la *Encuesta de Estructura Salarial (EES)*, y restringimos la muestra a los asalariados para poder comparar la distribución de las nóminas. Como se puede observar en la *figura 6*, las dos distribuciones son extremadamente parecidas.

Para corroborar este hecho, comparamos también los ratios de los percentiles de cada distribución y las características de las personas con una nómina en CaixaBank con los de la *ESS* en el *cuadro 2*. Así, podemos comprobar que el peso de distintos colectivos en los datos de CaixaBank es muy parecido al de la *EES*, y, por tanto, igual de representativo del conjunto de la población asalariada (aunque con un mayor tamaño muestral).

Cuadro 2.

Comparativa de los ratios de percentiles de la distribución de salarios netos y de la distribución por edad y género

	<i>CaixaBank (2018)</i>	<i>Encuesta de Estructura Salarial (2018)</i>
Ratio de percentiles		
P90/P10	4,23	4,04
P90/P50	1,95	1,86
P10/P50	0,46	0,46
P75/P25	1,87	1,82
Género (%)		
Hombre	53	52
Mujer	47	48
Edad (%)		
15-19	1	0
20-29	19	12
30-39	25	25
40-49	28	32
50-59	21	24
60+	6	7
Tamaño muestral	2.325.908	216.726

Nota: La tabla muestra los ratios de los percentiles de la distribución de salarios netos y la distribución por edad y género separadamente para la muestra interna de CaixaBank y para la *Encuesta de Estructura Salarial (ESS)* en el año 2018.

Fuentes: CaixaBank Research, a partir de datos internos de CaixaBank y de la *Encuesta de Estructura Salarial* (INE).

Tener unos datos que sean representativos del conjunto de la población nos permite construir indicadores de desigualdad, como el índice de Gini o las curvas de Lorenz, para el conjunto de la población y por los distintos subgrupos de población, y analizar su evolución a lo largo del tiempo.

4. TOMANDO EL PULSO DE LA ECONOMÍA ESPAÑOLA A PARTIR DE LOS INDICADORES DEL PORTAL DE ECONOMÍA EN TIEMPO REAL

El portal de Economía en Tiempo Real permite analizar la coyuntura económica con indicadores representativos del conjunto de la economía española, como vimos en el apartado anterior. En esta sección se describen algunos de los resultados que se pueden obtener con los indicadores. Con todo, al actualizar el portal mensualmente y con poca latencia, le recomendamos al lector ir directamente a la página web <https://realtimeeconomics.CaixaBankresearch.com/#/home> para analizar los últimos indicadores. Parte del esfuerzo de la creación del portal consistió en crear distintos niveles de entrada, con un apartado “Descubrir” para todos los públicos en el que la información ya está analizada y se proporcionan directamente los resultados principales; otro más interactivo para un público, como periodistas o *policy makers*, más interesado en cuestiones concretas y donde se pueden seleccionar los indicadores y el nivel de detalle deseado; hasta el nivel en el que se pueden descargar datos y gráficos con mayor desagregación para realizar análisis e investigaciones académicas propios.

El indicador de consumo muestra el dinamismo del consumo en la segunda mitad del año 2023, con un ritmo de crecimiento alrededor del 7,0 % en promedio, aunque algo inferior a la primera mitad del año (9,4 % interanual). El consumo presencial doméstico en España, medido a partir del gasto presencial y reintegros con tarjetas de crédito y débito emitidas por CaixaBank, cerró el año 2023 creciendo un 3,3 % interanual en diciembre (3,7 % en 2S 2023), con una desaceleración respecto al primer semestre (5,8 %), pero que no fue a más en el segundo semestre del año. Por su parte, el comercio electrónico, medido a partir del gasto con tarjetas de crédito y débito emitidas por CaixaBank en un TPV virtual, mantiene su ritmo de crecimiento de doble dígito, al avanzar un 13,3 % interanual en diciembre de 2023 (11,4 % en 2S 2023).

El indicador de salarios muestra que en 2023 aumentaron las retribuciones de los trabajadores, pero sin producirse efectos de segunda ronda significativos. Los ingresos salariales del sector privado español aumentaron un 4,3 % interanual en diciembre de 2023, con lo que se mantiene el avance alrededor del 4 % a lo largo de 2023 en las nóminas del sector privado (4,1 % en promedio). Con todo, este ritmo es 1,7 p.p. superior al que experimentaron un año antes (+2,6 % interanual en diciembre de 2022). Por sectores, los ingresos salariales avanzaron un 4,3 % en 2023 en el sector de servicios turísticos, un ritmo algo superior al del sector de servicios no turísticos (3,9 %) y de la industria (3,8 %). Por su parte, los ingresos salariales en 2023 avanzaron algo menos en la construcción (3,1 %) y en la agricultura (2,5 %).

El indicador de accesibilidad de la vivienda permite observar cómo un hogar típico destina 6,8 años de renta íntegra para comprar una vivienda en España, aunque existen marcadas

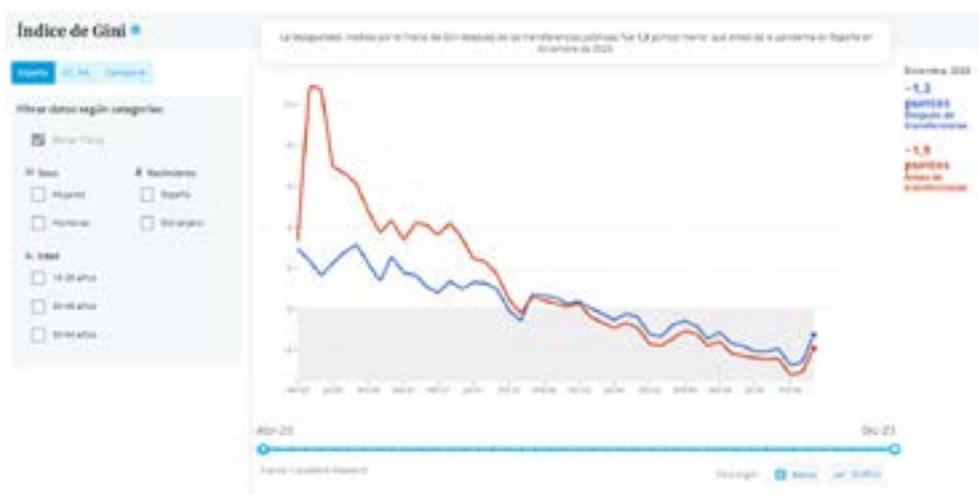
diferencias entre regiones. Madrid y Baleares son las CC. AA. donde más cuesta acceder a la vivienda (11 y 10 años de renta); mientras que Extremadura y Castilla-La Mancha son las CC. AA. donde menos (3,6 años).

Los indicadores de turismo muestran un fuerte crecimiento del turismo internacional en España en 2023. En la segunda mitad del año, el gasto turístico de los extranjeros en España pasó de crecer un 8,3 % en agosto a un 13,2 % en diciembre. El turismo interior de los españoles se mantuvo en niveles elevados al inicio del año (9,2 % interanual en enero), aunque ha ido perdiendo fuelle a lo largo del año (1,2 % en diciembre). Por otro lado, los indicadores de gasto en el extranjero muestran que los españoles han salido al extranjero mucho más en 2023 que en el año anterior, con crecimientos del gasto por encima del 6 % interanual en todo el año.

Los indicadores disponibles en el apartado de desigualdad nos muestran que se ha deshecho el impacto de la pandemia en la desigualdad y que esta ya alcanza un nivel inferior al que existía antes de la pandemia (figura 7). Concretamente, la desigualdad medida por el índice de Gini antes de las transferencias públicas fue 1,9 puntos menor que antes de la pandemia en España en diciembre de 2023 (-1,4 puntos en diciembre de 2022). Alcanzar una situación de desigualdad salarial menor a la de febrero de 2020 es un hito remarcable, porque la pandemia afectó fuertemente a las nóminas de los trabajadores. La desigualdad en España, medida por el índice de Gini, aumentó en 10,8 puntos en tan solo dos meses, de febrero a abril de 2020 antes de las transferencias públicas. A lo largo de los siguientes meses, y en particular a partir de la primavera de 2021, con una mayor reactivación económica, se aceleró la

Figura 7.

Índice de Gini



Fuentes:

reducción de la desigualdad. Las transferencias públicas jugaron un papel clave para contener ese aumento de la desigualdad, dado que las ayudas permitieron, en el momento álgido de la crisis (abril de 2020), reducir la desigualdad en un 80 % aproximadamente. Dos años y medio después, en diciembre de 2023, el índice de Gini, una vez incluidas las transferencias públicas era ya 1,3 puntos menor que antes de la pandemia en España.

5. CONCLUSIÓN

Los movimientos de cuentas bancarias, de alta frecuencia y con un nivel de granularidad elevado, permiten obtener información económica de gran calidad y precisión de manera casi instantánea y con un enorme potencial para la investigación económica. En este trabajo hemos descrito la experiencia de CaixaBank Research en el análisis de datos masivos de transacciones financieras a partir de los proyectos llevados a cabo, y en particular, mediante la creación de un portal de Economía en Tiempo Real. Este proyecto monitoriza la evolución de la economía española a través de 12 indicadores construidos con datos internos de CaixaBank, agregados mediante técnicas de *big data*. Los indicadores se agrupan en cinco ámbitos: el consumo, la vivienda, el turismo, los salarios y la desigualdad. Cada uno de estos indicadores se puede consultar por distintas dimensiones, lo cual permite identificar diferencias entre colectivos (edad, sexo, ingresos, sector de actividad), regiones (CC. AA., municipios), etc. En total, actualmente publicamos más de 800 series de datos que actualizamos mensualmente. El proyecto está en curso de mejora constante, y se añadirán al portal nuevos indicadores y dimensiones a medida que estén disponibles.

Referencias

- ASPACHS, O., DURANTE, R., GRAZIANO, A., MESTRES, J., G. MONTALVO, J. y REYNAL-QUEROL, M. (2021). Tracking the impact of COVID-19 on economic inequality at high frequency. *PLoS One*, 16(3), e0249121. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249121>
- ASPACHS, O., DURANTE, R., GRAZIANO, A., MESTRES, J., G. MONTALVO, J. y REYNAL-QUEROL, M. (2022). Real-time inequality and the welfare state in motion: evidence from COVID-19 in Spain.
- BISQUERT PERLES, M., MESTRES DOMÈNECH, J., SÁNCHEZ PALOMINO, J. A. y VIDAL MARTÍNEZ, A. (2023). "Patrones de consumo y ahorro tras la jubilación. *Informe Mensual Junio 2023*. CaixaBank Research. <https://www.caixabankresearch.com/es/economia-y-mercados/mercado-laboral-y-demografia/patrones-consumo-y-ahorro-tras-jubilacion>
- CAMPOS, A. y MONTORIOL, J. (2019). ¿Cómo puede el big data potenciar la sostenibilidad del sector turístico? *Informe Sectorial Turismo*, enero de 2019. CaixaBank Research. https://www.caixabankresearch.com/sites/default/files/content/file/2022/04/05/91184/is_turismo1_2022_esp.pdf
- CHETTY, R., FRIEDMAN, J. N. y STEPNER, M. (2020). The economic impacts of COVID-19: Evidence from a new public database built using private sector data (No. w27431). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w27431/w27431.pdf

- GRAZIANO, A. y MESTRES, J. (2022). Los precios de la luz están por las nubes, ¿y el importe de su recibo? *Informe Mensual Enero 2022*. CaixaBank Research. <https://www.caixabankresearch.com/es/economia-y-mercados/actividad-y-crecimiento/precios-luz-estan-nubes-y-importe-su-recibo?1150>
- MONTORIOL, J. (2020). ¿Comprar o alquilar? Una cuestión de ingresos, pero sobre todo de capacidad de ahorro. *Informe Sectorial Inmobiliario Julio 2020*. CaixaBank Research. <https://www.caixabankresearch.com/es/analisis-sectorial/inmobiliario/comprar-o-alquilar-cuestion-ingresos-sobre-todo-capacidad-ahorro?1150>