

# Big data y market intelligence en el Banco de España

## RICARDO GIMENO

Jefe de División de Análisis e Inteligencia de Mercados  
Banco de España

## JOSÉ MANUEL MARQUÉS

Jefe de División de Innovación Financiera  
Banco de España

## RESUMEN

Los bancos centrales cuentan con enormes bases de datos como consecuencia de su papel central en el sistema financiero. Aunque su uso primario está relacionado con sus funciones primarias (política monetaria, supervisión, estabilidad financiera), su uso puede extenderse a otros múltiples usos. Dentro de esos usos secundarios se encuentra el desarrollo de indicadores que pueden ponerse al servicio de la sociedad, como son aquellos que permiten hacer previsiones en tiempo real del PIB o la Inflación, o, usando técnicas novedosas de tratamiento del lenguaje natural, para seguimiento de noticias, tendencias o informes de analistas.

Como parte de su vocación como servicio público, el Banco de España ha desarrollado el laboratorio de datos BELab, para permitir el acceso a investigadores externos a dichas fuentes de información, pero manteniendo un escrupuloso cumplimiento de la Ley de Protección de datos.



Los bancos centrales, dadas las competencias tradicionales que suelen tener asignadas en cuanto a la definición y ejecución de la política monetaria, promover y vigilar el buen funcionamiento de los sistemas de pago y emisión y distribución del efectivo, se sitúan en el centro mismo del sistema financiero, por lo que suelen acumular bases de datos de tamaño considerable con información muy valiosa, pero que, al mismo tiempo, puede ser compleja de analizar. Baste mencionar, y sin ánimo de ser exhaustivo, la necesidad de recopilar información sobre los activos que son usados como colateral en las operaciones de financiación, las adquisiciones de activos en los programas de compras, el uso de las transacciones de pagos minoristas o la circulación de los billetes. Esto lleva a la **existencia de bases de datos, que muchas veces alcanzan el conjunto del Eurosistema** como la SHS (*Securities Holdings Statistics*), que recoge información de los tenedores de los activos financieros, la CEPH (*Common Eurosystem Pricing Hub*) y la CSDB (*Centralised Securities Database*), con datos de precios, o RIAD (*Register of Institutions and Affiliates Data*) con información de los lazos entre entidades.

En algunos casos, como es el caso del Banco de España, además tienen asignadas competencias de **supervisión prudencial** —lo que proporciona acceso a los estados que tienen que reportar las entidades financieras sobre sus actividades de intermediación—, desarrollan herramientas para gestionar los riesgos asumidos por las entidades como las centrales de información de riesgo (CIRBE) o recopilan reclamaciones en relación con las prácticas de las entidades supervisadas en virtud de las competencias relacionadas con la supervisión de conducta.

Por último, cabe mencionar que las atribuciones que suelen tener estas autoridades en cuanto al seguimiento de la actividad económica han propiciado el desarrollo de **bases de datos detalladas sobre aspectos como los estados económicos y financieros de las empresas** (Central de Balances) o la situación patrimonial de los hogares (Encuesta Financiera de los Hogares).

En la mayoría de los casos —probablemente siendo los mencionados en el último párrafo la principal excepción— la recopilación de esta información se ha realizado para fines muy concretos, vinculados con las competencias que daban origen al requerimiento de esta información y para lo cual era necesario un análisis y procesamiento laborioso. Sin embargo, la irrupción de las nuevas tecnologías, tanto en el almacenamiento como en el procesamiento o en el análisis de los datos, no solo están permitiendo que esas labores puedan realizarse de un modo mucho más eficiente, sino que, además, posibilitan nuevos usos de estas bases de datos.

En relación con el primer aspecto, existen numerosos ejemplos que abarcan elementos como el uso de **reconocimiento óptico de caracteres** (OCR) para capturar y procesar la información de cláusulas contractuales o las características de los billetes, el uso de **algoritmos de inteligencia artificial**, o *machine learning*, para la detección de errores en la información reportada por las empresas o entidades financieras o la aplicación de **técnicas de procesamiento de lenguaje natural** para categorizar información como la relacionada con la exposición al riesgo de cambio climático. Estas mejoras no solo implican un mejor tratamiento de la información,



sino que, en algunos casos, está llevando a los supervisores a plantearse un cambio en la forma en la que se solicita la información, facilitando, de este modo, no solo mayores posibilidades para analizarla sino también una considerable reducción del coste que para los supervisados tiene proporcionar la información (véase, como ejemplo, Bank of England, 2020).

Sin embargo, como se ha mencionado anteriormente, el uso de las nuevas tecnologías también amplía los posibles usos de esta información, incluyendo usos como los relacionados con el análisis o la investigación. En este sentido, son numerosos los ejemplos de trabajos analíticos en los que, gracias a las posibilidades analíticas y de procesamiento de la información, se está usando **información recopilada por los bancos centrales**, en muchos casos combinándola con nuevas fuentes de información como la proveniente de las redes sociales o de las búsquedas en internet.

Sin ánimo de ser exhaustivos, a continuación se describen algunos ejemplos en el ámbito de las **estadísticas recopiladas por el Banco de España** que ilustran esta tendencia que, por otro lado, es común en otros bancos centrales<sup>1</sup>. Así, por ejemplo, los trabajos de Conesa *et al.* (2015) con datos de pagos, que permiten obtener predictores en tiempo real del PIB; los múltiples trabajos con la Central de Información de Riesgos del Banco de España, que se han usado por Jiménez *et al.* (2012; 2014; 2017 y 2020), Arce *et al.* (2021), de Andrés *et al.* (2021) para explorar los mecanismos de la concesión de créditos en España, así como sus canales de interconexión con la política monetaria y la estabilidad financiera.

Dentro de estos trabajos, destacan también aquellos que están permitiendo transformar datos que antes no eran considerados cuantificables, como son los textos (*text mining*) para identificar el alineamiento entre los mensajes del banco y la prensa económica (Moreno y González, 2020), la predicción del PIB (Díaz Sobrino *et al.*, 2020) o los resultados bancarios (Aguilar y Torres, 2021).

Sin embargo, esta tendencia no está exenta de retos que es necesario abordar. En primer término, este tipo de análisis requieren personal especializado en el tratamiento de la información, como los científicos de datos, así como el *software* y equipo necesarios para almacenar y procesar la información con las técnicas disponibles. Además de las implicaciones presupuestarias de estas necesidades en el contexto de un banco central, pueden existir otro tipo de condicionantes, como la necesidad de garantizar unos altos estándares de seguridad en la gestión de la información, que puede limitar el rápido desarrollo de infraestructuras como el acceso a la nube o a la información de las redes sociales.

Por otro lado, es necesario cumplir de modo escrupuloso con la normativa que regula el uso de la información. Así, aspectos como el **reglamento general de protección de datos** en el caso de los datos de carácter personal, o los **compromisos de confidencialidad** en mucha de

<sup>1</sup> Véase, por ejemplo, HIRSCHBÜHL *et al.* (2021) para una revisión de estas aplicaciones en el caso del BCE con respecto al análisis de los ciclos económicos.



la información reportada en el ámbito de la supervisión deben ser tenidos en cuenta a la hora de realizar análisis y, especialmente, en el momento de difundirlos o publicarlos.

Por último, es preciso abordar la organización en torno a la **gobernanza y la gestión de los datos**. El entorno al que se enfrenta el analista implica el acceso potencial a numerosas bases de datos, con distintas características, restricciones y tradicionalmente gestionadas de un modo muy heterogéneo y poco habituado a lidiar con el enfoque de un investigador. Por ello, es necesario abordar planteamientos integrales que contemplen la información estandarizada que debe incorporarse en todos los datos que gestiona y almacena el banco central en lo que se conoce como el **metadato**, asignen la responsabilidad en la gobernanza de la información, expliciten de forma homogénea y controlada las restricciones en el acceso y uso de las distintas bases de datos o garanticen el cumplimiento actualizado con las normas de privacidad o confidencialidad.

El **plan estratégico vigente del Banco de España** (Banco de España, 2020), reconoce la importancia de las bases de datos y las herramientas elaboradas por la institución no solo para el desarrollo de sus funciones, o el trabajo de los propios analistas del banco, sino como forma de generar valor para el conjunto de la sociedad. Para ello, se ha elaborado un plan para la creación de un **marco de actuación reglado para el acceso de terceros a las bases de datos, estadísticas y herramientas analíticas del Banco de España**. Esto ya se ha concretado en el lanzamiento del **Laboratorio de Datos BELab**<sup>2</sup> que, aunque se encuentra todavía en su fase piloto, ya permite poner a disposición de investigadores externos los recursos del Banco de España.

BELab nace de la necesidad de aunar el acceso a los microdatos disponibles con la protección escrupulosa de los requisitos legales, y especialmente los de protección de datos. Para ello, los proyectos de los investigadores deben ser evaluados previamente, los datos *curados* adecuadamente y, finalmente, asegurarse de que los trabajos realizados por el investigador visitante no permitan la identificación de agentes individuales. Para ello, el acceso se hace mediante el establecimiento de *Data rooms* (por ahora en las sedes de Madrid y Barcelona) a los que los investigadores pueden trasladarse de forma física y trabajar con los microdatos de las distintas bases y usando las herramientas de tratamiento de los datos que el banco pone a su disposición (como Python, R, o Stata).

En la actualidad se encuentran ya disponibles en este laboratorio los datos de empresas individuales de la Central de Balances, las Emisiones de Valores representativos de Deuda, los datos de los préstamos a personas jurídicas de la Central de Información de Riesgos, la Encuesta Financiera de las Familias o la Encuesta de Competencias Financieras. La existencia de un mecanismo para el uso de estas bases de datos por parte de todos los investigadores es muy relevante para lograr que sean base para el desarrollo científico, una de cuyas piedras angulares es la posibilidad de replicar los estudios desarrollados por otros colegas, y que

<sup>2</sup> [www.bde.es/bde/es/areas/analisis-economi/otros/laboratorio-de-d/](http://www.bde.es/bde/es/areas/analisis-economi/otros/laboratorio-de-d/)



resultaría imposible si estos microdatos fueran privados y no estuvieran al alcance de cualquier investigador.

En cualquier caso, siguen existiendo restricciones normativas u operativas que limitan la accesibilidad a cierta información. No obstante, estas limitaciones no deben servir como coartada por la complacencia, y con ello socavar la vocación de servicio público del Banco de España para facilitar el uso de la información que atesora. En este sentido, siguen abiertas vías alternativas para la disposición de información por parte del banco. Así, por ejemplo, las distintas áreas involucradas en la elaboración de estadísticas e información financiera hacen una labor fundamental para la generación interna de distintos indicadores económicos y financieros agregados de gran utilidad en múltiples modelos y que se han puesto a disposición de los investigadores desde hace décadas.

En el futuro existen vías que se pueden explorar, y que no requeriría el uso de los microdatos individuales, como son el uso de parámetros (como los momentos de las distribuciones conjuntas) que caractericen los datos que no pueden hacerse públicos de forma individual, pero que permiten la creación de datos sintéticos. Así, por ejemplo, el uso de medias y matrices de varianzas y covarianzas se usan ya de forma habitual en metaanálisis que no requieren el acceso a los datos de partida.

## Referencias bibliográficas

AGUILAR, A. y TORRES, D. J. (2021): «The impact of COVID-19 on analysts' sentiment about the banking sector», *Documentos de Trabajo*, N.º 2124. Banco de España.

ANDRÉS, P. de, GIMENO, R. y MATEOS DE CABO, R. (2021): «The gender gap in bank credit access», *Journal of Corporate Finance*, 71, 101782.

ARCE, O., MAYORDOMO, S. y GIMENO, R. (2021): «Making room for the needy: the credit-reallocation effects of the ECB's corporate QE», *Review of Finance*, 25(1), pp. 43-84.

BANCO DE ESPAÑA (2020): *Plan Estratégico 2024*.

 [www.bde.es/f/webbde/SSICOM/20200115/planestrategico.pdf](http://www.bde.es/f/webbde/SSICOM/20200115/planestrategico.pdf)

BANK OF ENGLAND (2020): «Transforming data collection from the UK financial sector», *Discussion paper*.

CONESA, C., GAMBACORTA, L., GORJÓN, S. y LOMBARDI, M. J. (2015): «The use of payment systems data as early indicators of economic activity», *Applied Economics Letters*, 22(8), pp. 646-650.



- DÍAZ SOBRINO, N., GHIRELLI, C., HURTADO, S., PÉREZ, J. J. y URTASUN, A. (2020): «The narrative about the economy as a shadow forecast: an analysis using Banco de España quarterly reports», *Documentos de Trabajo*, N.º 2042. Banco de España.
- HIRSCHBÜHL, D., ONORANTE, L. y SAIZ, L. (2021): «Using machine learning and big data to analyse the business cycle», *ECB Economic Bulletin*, Issue 5/2021.
- JIMÉNEZ, G., MIAN, A., PEYDRÓ, J. L. y SAURINA, J. (2020): «The real effects of the bank lending channel», *Journal of Monetary Economics*, 115, pp. 162-179.
- JIMÉNEZ, G., ONGENA, S., PEYDRÓ, J. L. y SAURINA, J. (2012): «Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications», *American Economic Review*, 102(5), pp. 2301-26.
- (2014): «Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking?» *Econometrica*, 82(2), pp. 463-505.
- (2017): «Macroprudential policy, countercyclical bank capital buffers, and credit supply: Evidence from the Spanish dynamic provisioning experiments», *Journal of Political Economy*, 125(6), pp. 2126-2177.
- MORENO, A. y GONZÁLEZ, C. (2020): «Análisis de sentimiento del *Informe de Estabilidad Financiera*», *Documentos de Trabajo*, N.º 2011. Banco de España.

