BBVA Research

Big Data in Economics:
our journey in BBVA Research

Facultad de Ciencias Económicas y Empresariales
Universidad Complutense de Madrid

Jorge Sicilia
Chief economist BBVA

February 2022

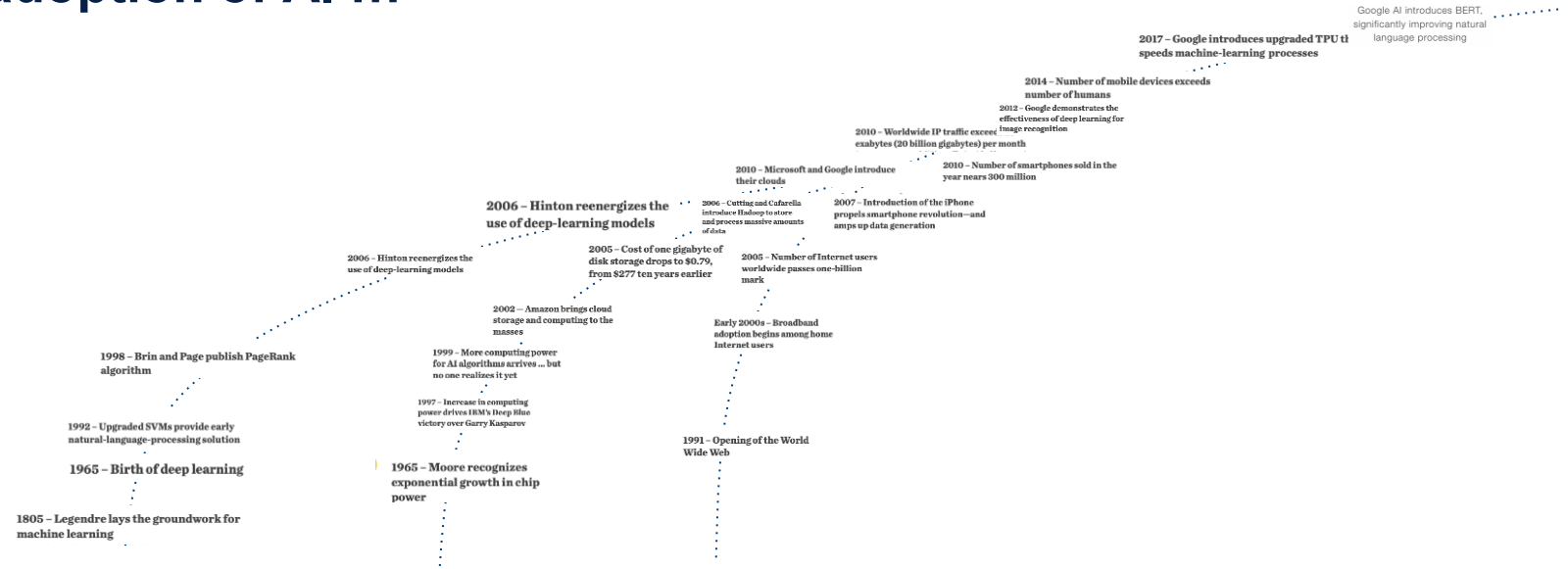
Creating Opportunities

BBVA Research

01

What's Big Data and what can we do with it? A quick global overview

Why Now? The aligment of several trends has triggered a rapid adoption of AI …



Source: An executive's guide to AI (Mckinsey)

At present, AI Industrial and Daily Life applications are found everywhere

Image Recognition

Voice Recognition

Call Centers

Chatbots

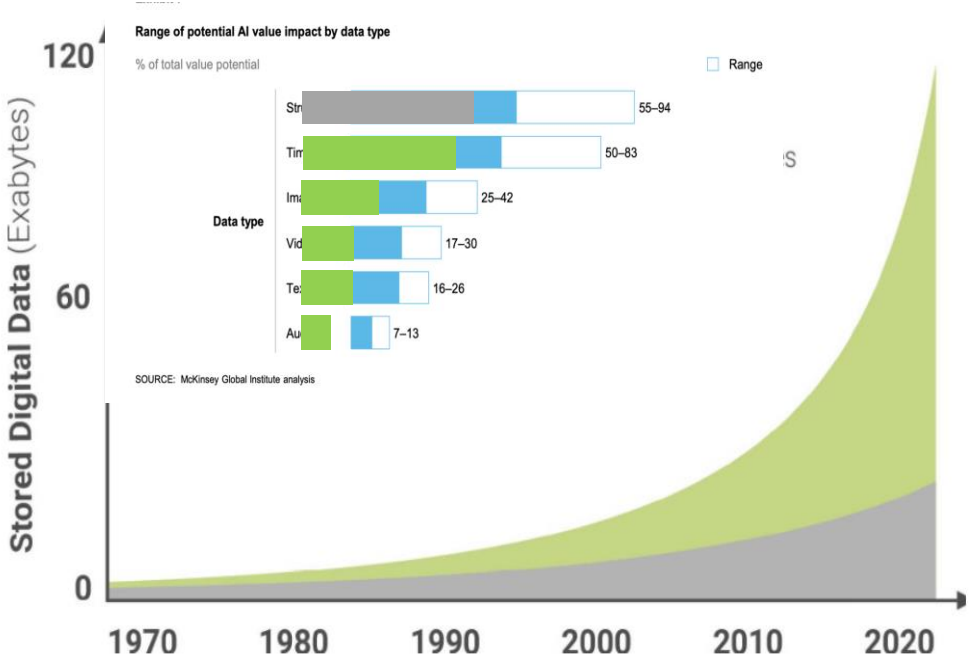
Defence

Medical Diagnosis

Gaming

Agriculture

Driverless Car



The graph shows the huge increase in unstructured data since the 2000s

The increasing availability of Big Data is changing the nature of Economic Research towards becoming more multi-disciplinary ones

Six Decades of Top Economics Publishing: Who and How?

DANIEL S. HAMERMESH*

Presenting data on all full-length articles in the three top general economics journals for one year in each decade 1960s–2010s, I analyze changes in patterns of coauthorship, age structure and methodology, and their possible causes. The distribution of number of authors has shifted steadily rightward. In the last two decades, the fraction of older authors has almost quadrupled. Top journals are publishing many fewer papers that represent pure theory, regardless of subfield, somewhat less empirical work based on publicly available data sets, and many more empirical studies based on data collected by the author(s) or on laboratory or field experiments. (*JEL* A14)

https://pubs.aeaweb.org/doi/pdfplus/10.1257/jel.51.1.162

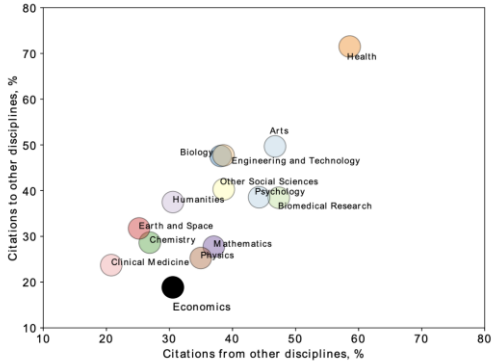


Figure 1. Citations in and out of disciplines by discipline. Source: Van Noorden (2015).

How Did COVID-19 and Stabilization Policies Affect Spending and Employment?

A New Real-Time Economic Tracker Based on Private Sector Data

Raj Chetty, Harvard
John N. Friedman, Brown
Nathaniel Hendren, Harvard
Michael Stepner, Univ. of Toronto
and the Opportunity Insights Team

June 17, 2020

OPPORTUNITY INSIGHTS

9 data providers and
Near 45 authors !!!

intuit earnin womply affinity solutions homebase ZEARN burningglass DDD BILL & MELINDA GATES foundation

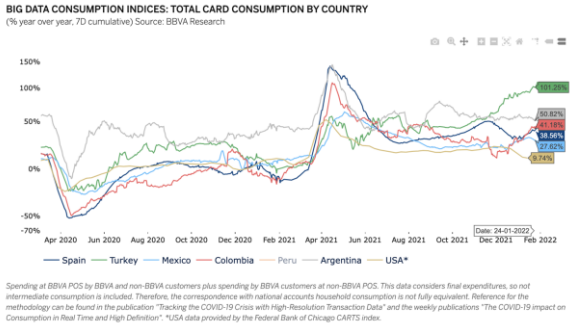
Olivier Blanchard
@ojblanchard1

Enormously impressed by the webinar by Raj Chetty today
bcf.princeton.edu/event-director...
An amazing combination of how to put big data together, test hypotheses, and draw highly relevant policy conclusions. This is 21st century economic research. I feel old, but excited. Watch it.

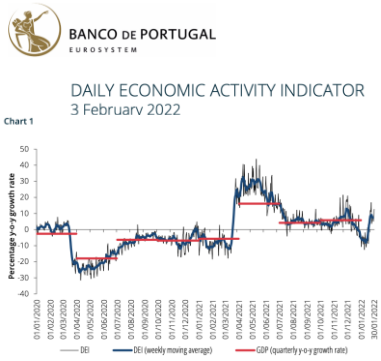
Traducir Tweet

The Economic Impacts of COVID-19: Evidence f
rom a New Public Database Built Using Private Sector Data

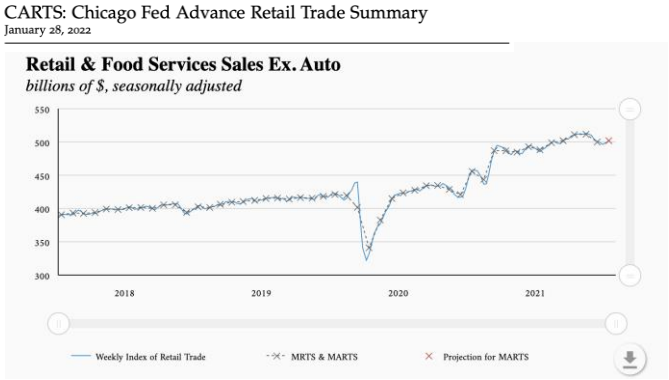
What can it be used for?: (i) Economics in Real Time…



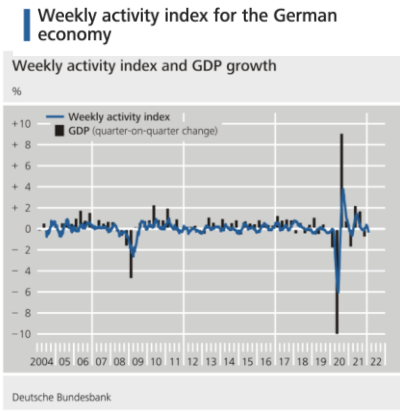
Source: BBVA Research



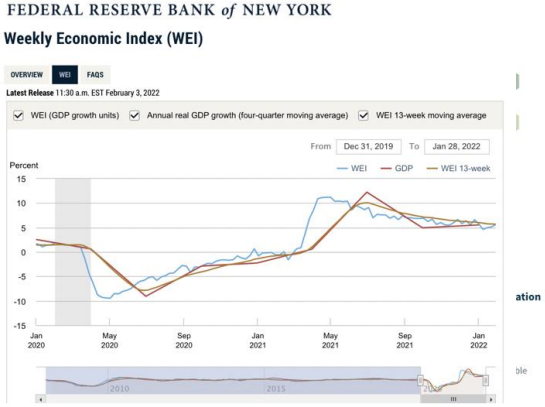
Source: Central Bank of Portugal



Source:Chicago Fed



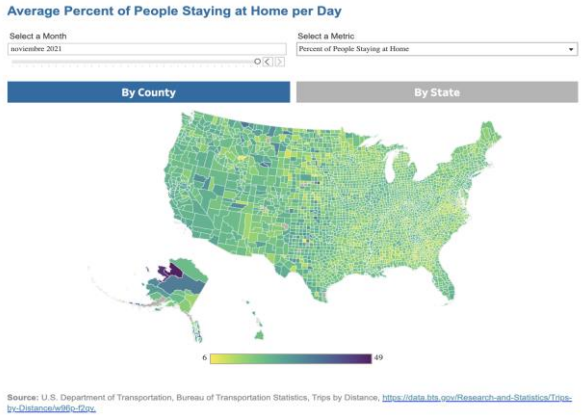
Source: Bundesbank



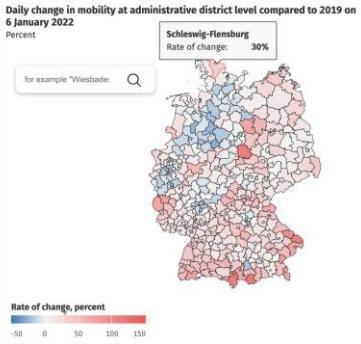
Source: NY Fed

Source: Affinity

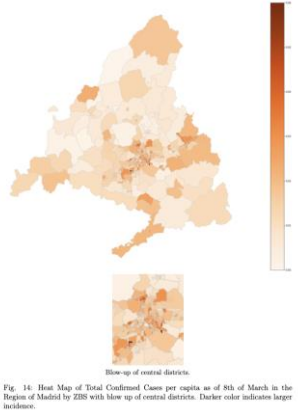
…(ii) High Definition Economics…

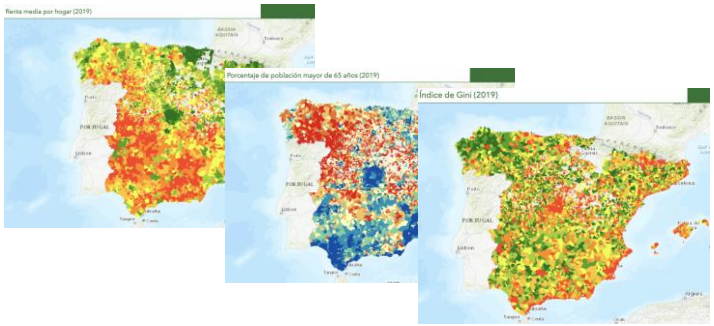


https://www.bts.gov/daily-travel

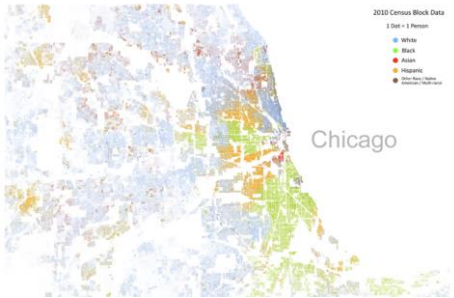


https://www.destatis.de/EN/Service/EXDAT/Datensaetze/mobility-indicators-mobilephone.html

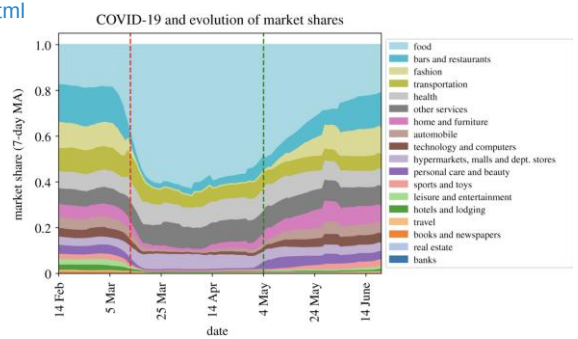




https://www.ine.es/experimental/atlas/experimental_atlas.htm



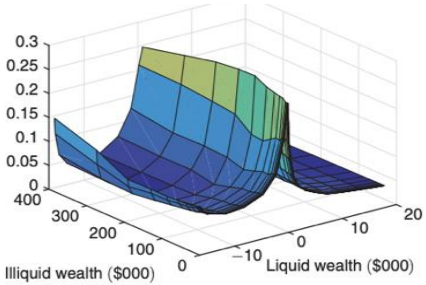
https://fivethirtyeight.com/features/the-most-diverse-cities-are-often-the-most-segregated/



https://www.econ.cam.ac.uk/research/cwpe-abstracts?cwpe=2030

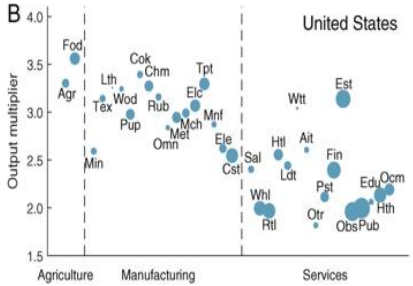
Potential application:"smart analysis for smart policies" on consumers, firms, cities, sustainability, etc

Marginal Propension to Consume (500): Liquid vs Illiquid Households



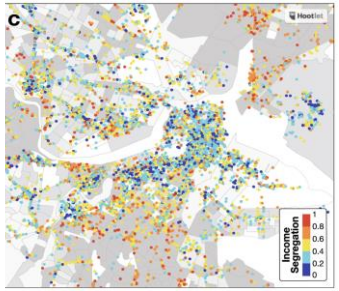
Monetary Policy according to Hnak. Kaplan, Moll & Violante (2018)

Output Multipliers in A Netwroked Economy



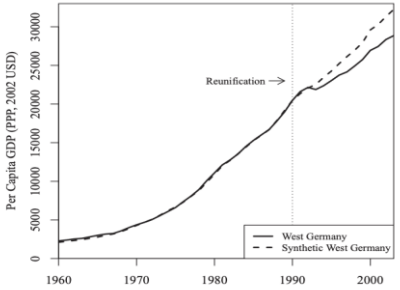
How production networks amplify economic growth. PNAS (2022)

Place and individual income segregation.



Mobility patterns are associated with experienced income segregation in large US cities. Moro et al 2021 Nature

The Effect of German Reunification (synthetic control)

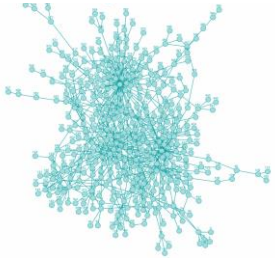


Comparative Politics and the Synthetic Control Method. Abadie et al (2018)

Effects of Covid on Education Progress By income

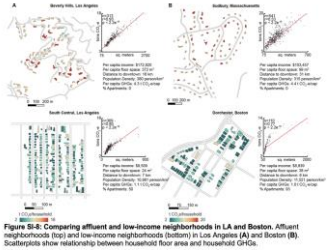
The Economic Impacts of COVID-19: Evidence f rom a New Public Database Built Using Private Sector Data

USA Input-Output Network



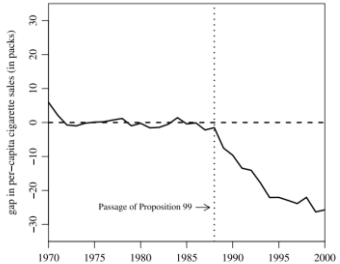
Acemoglu,D., Carvalho V, Ozdaglar,A, & Tahbaz-Salehi The network origins of aggregate Fluctuations. Econometrica

Household Co2 FootPrint According to different Characteristics



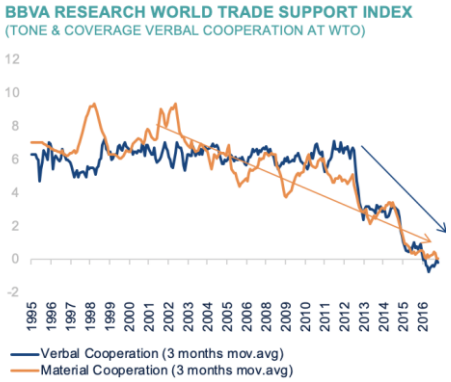
The Carbon Footprint of the US. PNAS (2020)

The Effect on Cigarrets consumption Ol California Legislation

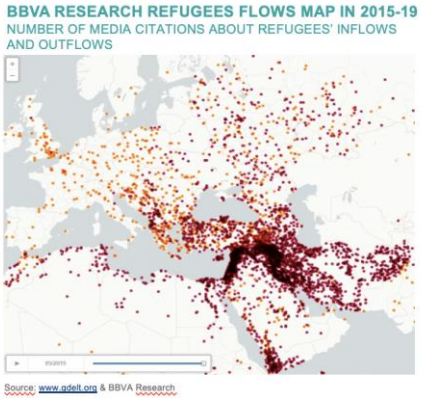


Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program

Turning text into Data: A view to understand narratives around key social, economic, and geopolitical issues (short term and structural)

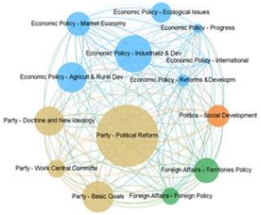


Map of mean UK labour market tightness
And Beveridge Curves

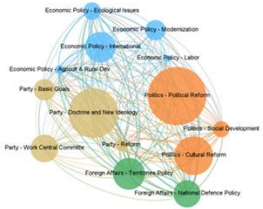


Understanding The China Strategy through…
The Chinese. Communist Party Files (2002-2017)

2002



2017



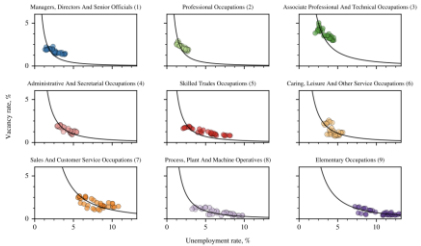
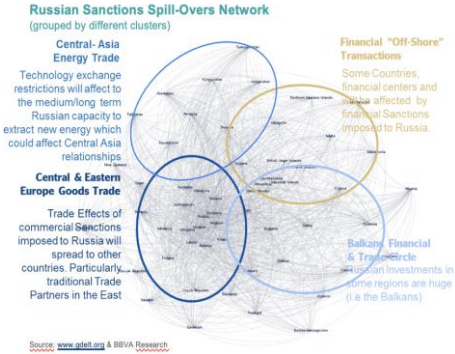


Figure 11: Beveridge curves (lines), estimated with Reed data, and Reed data (points) in u-v space for each 1-digit SOC code at quarterly frequency. Source: Reed, ONS.

Transforming naturally occurring text data into economic statistics: the case of online job vacancy postings (NBER)

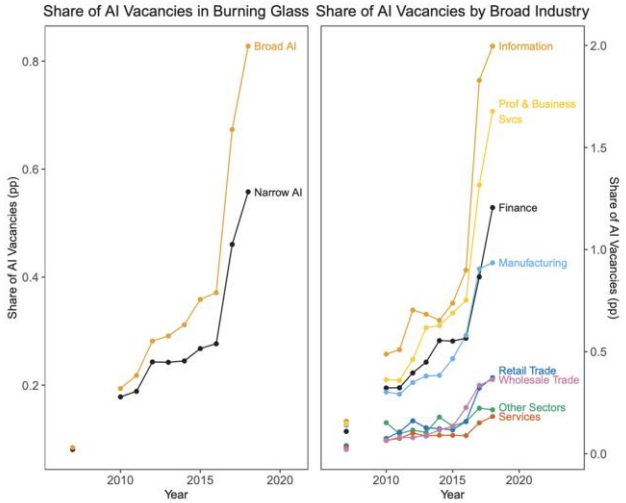


For all this reasons, it is a toolbox where expertise is on high demand

AI AND JOBS: EVIDENCE FROM ONLINE VACANCIES

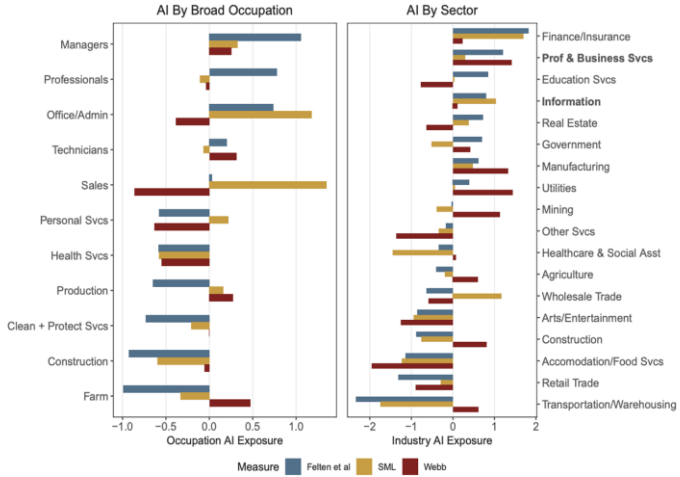
(Source: Acemoglu,D, Autor, D, Hazell, J & Restrepo NBER wp28257)

FIGURE 2: Share of AI Vacancies in Burning Glass



The left panel plots the share of vacancies in Burning Glass, that post a skill in the Broad or Narrow AI categories, as defined in the main text. The right panel plots the share of narrow AI vacancies in Burning Glass, by year, in each broad industry grouping.

FIGURE 3: AI Exposure by Broad Occupation and Sector



The left panel plots the average of the standardized measures of AI exposure across broad occupations. The right panel plots the average of the standardized measures of AI exposure across 2-digit NAICS sectors, by taking the mean across the 6 digit SOC occupations posted in each 2 digit NAICS sector, weighted by the number of vacancies posted by each sector in each occupation.

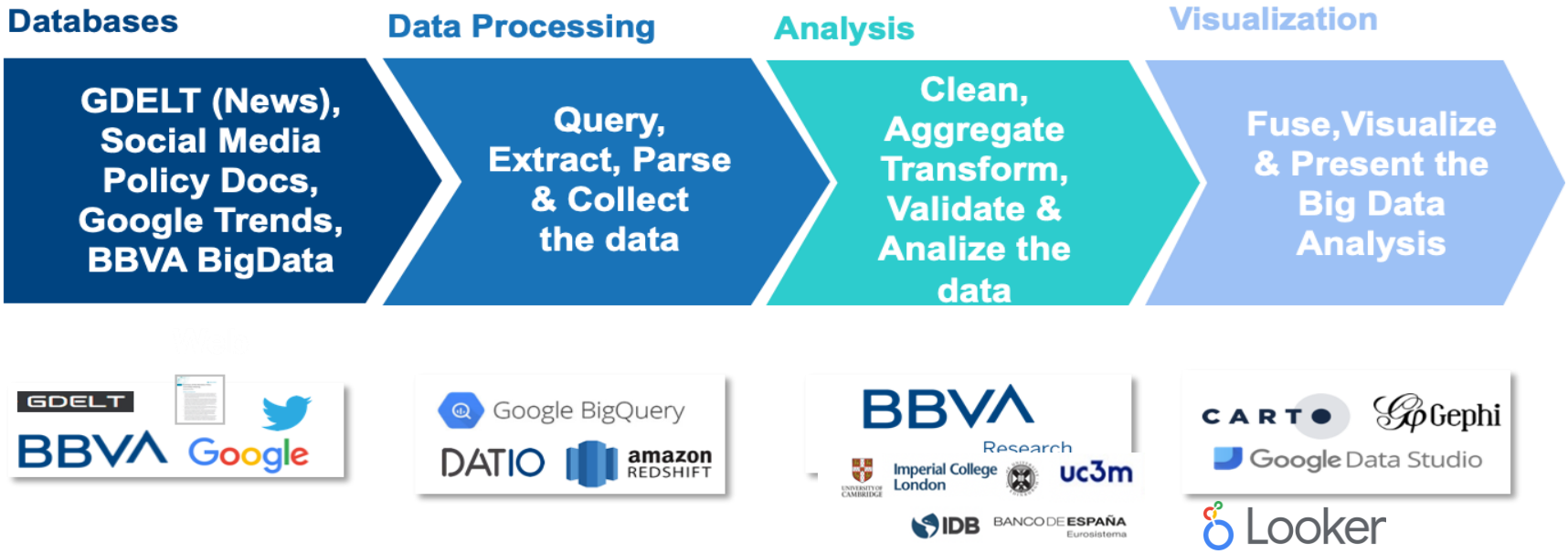
AI and Jobs: Evidence from Online Vacancies (NBER)

BBVA Research

02

How do we use Big Data at BBVA Research

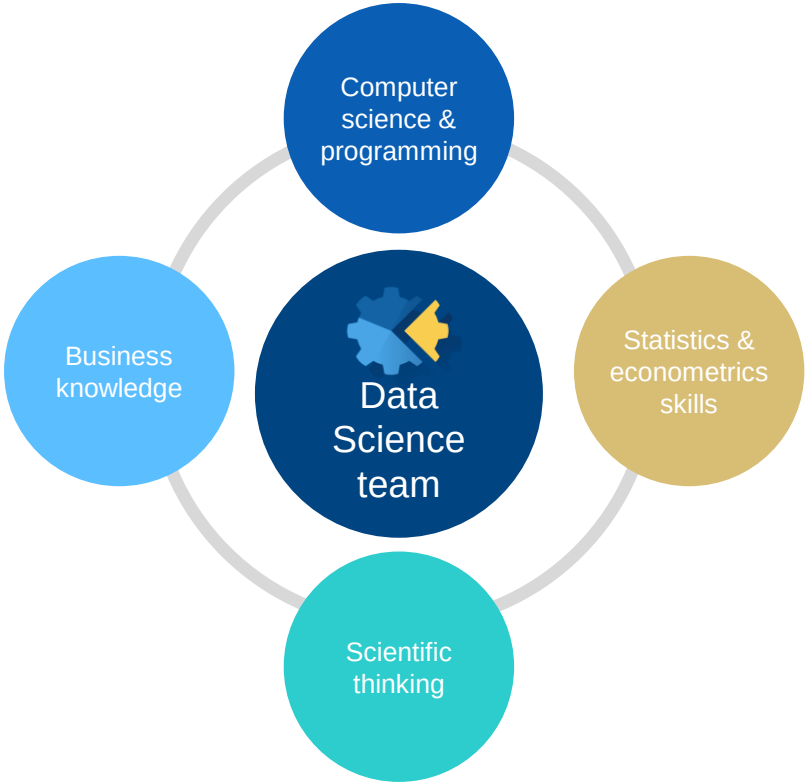
A Brief on BBVA Research Journey: our working process



Reassuring data management process

- Data Ingestion
- Data Protection
- Data Quality
- Data Representativity
- Data Analytics

We work with interdisciplinarity teams –inhouse and external– to be able to keep ourselves as close as possible to the frontier



Our external collaborators

UNIVERSITY OF CAMBRIDGE
INTERNATIONAL MONETARY FUND
Imperial College London
BANCO DE ESPAÑA Eurosistema
THE UNIVERSITY of EDINBURGH
IDB Inter-American Development Bank
uc3m Universidad Carlos III de Madrid
mit media lab
CLARITY AI
Google Cloud

Big Data at BBVA Research: our main lines of analysis working with internal and external databases

Text as Data

(Macroeconomics, Politics, Geopolitics, society, communication strategy,…)

National Accounts in Real Time & High Definition

(Macroeconomic & Sectorial analysis)

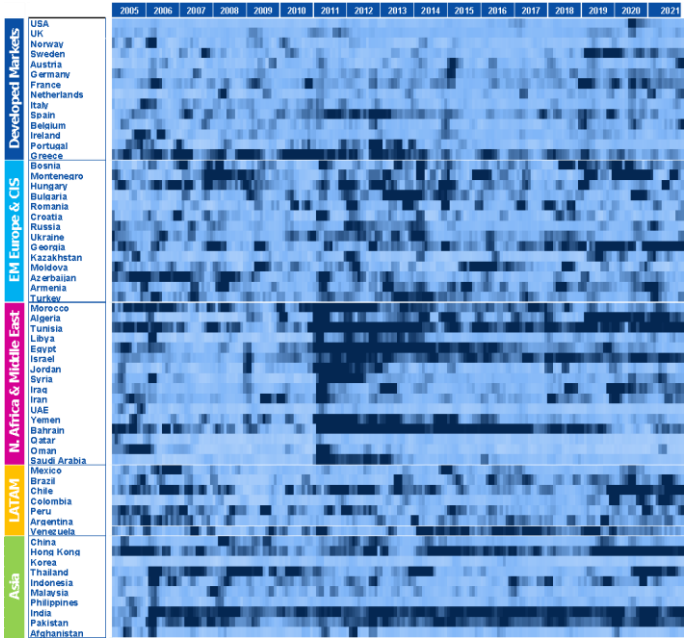
Structural & Policy Analysis In real Time

(Bank transactions Determinants, Monetary Policy, Saving Rates,e-commerce, Network analysis, sustainable lyfe styles,…)

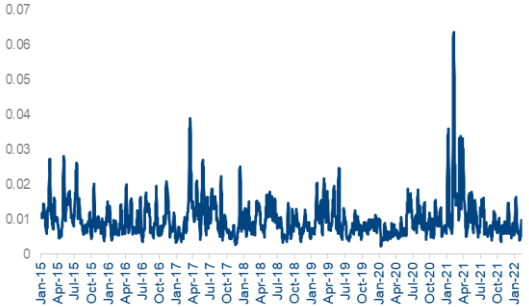
Analyzing geopolitics and social trends from news…

Text as Data

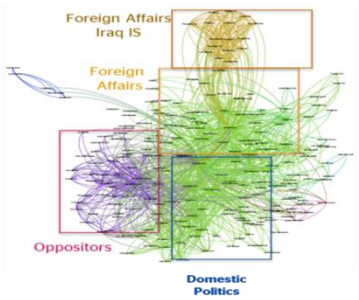
MONTHLY WORLD PROTEST INTENSITY MAP 2005 - 2022



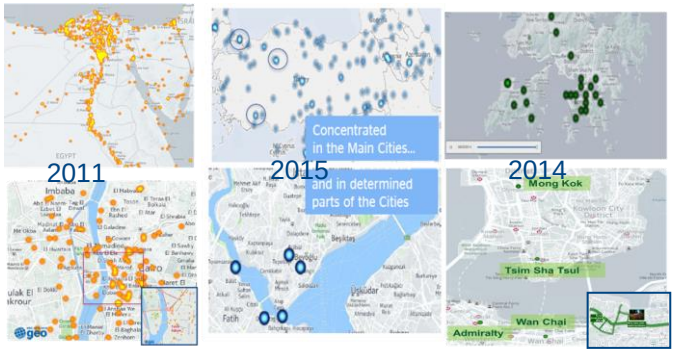
Turkey Protest Intensity Index



Turkey politics Network



Protest Intensity Index – Geolocated data



Source: www.gdelt.org & BBVA Research

… to economic issues from news and social media…

Text as Data

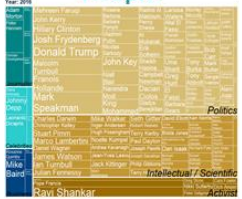
Understanding Sustainability framework using Big Data

What does the society receive from the media?

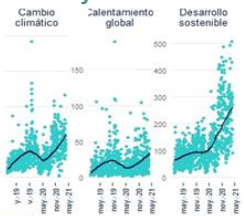
GDELT

MEDIA CLOUD

People

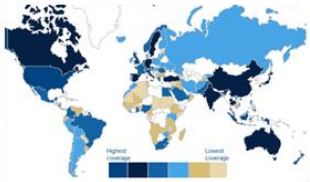


keywords



Organizations

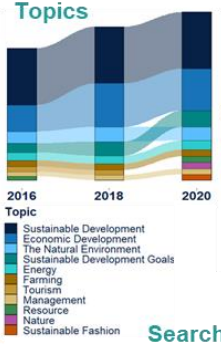
Coverage across countries



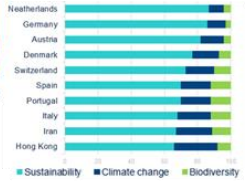
What the society is looking for?

Google Trends

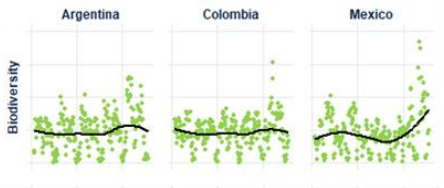
Topics



Searches distribution by countries



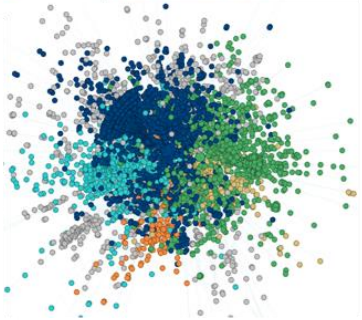
Searches by countries



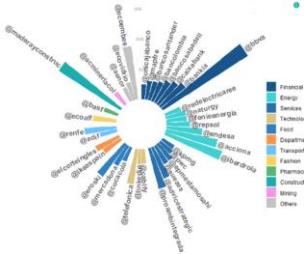
What the society is talking about?

Topics

Topics network



Companies



Coverage & sentiment

… to monetary policy analysis from official reports across countries

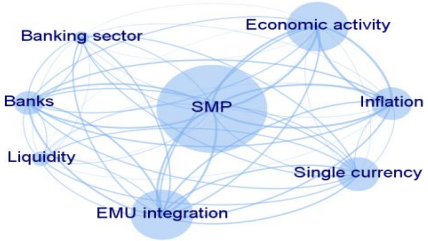
Text as Data

Monetary Policy in Developed Economies and Response in the Emerging Markets (Networks)

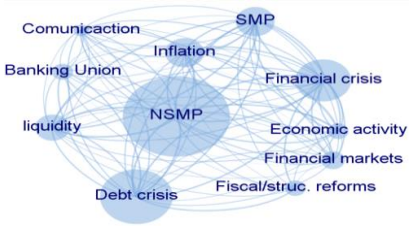
ECB

Identified topics

Pre Lehman
(1999-2007)



Crisis
(2007 -2013)



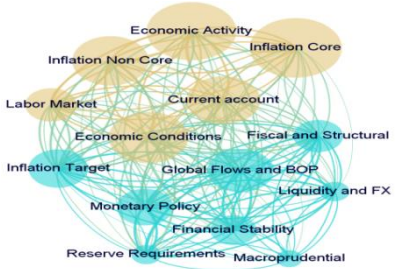
QE & Recalibration
(2013-2018)

CBRT

Identified topics

Inflation Target
(2006-2008)

Financial Crisis
(2009-2012)

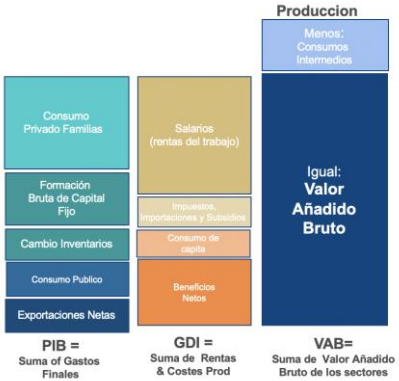


The Global Post-Crisis
(2012-2017)

Working with bank's data, Economics, Sectoral analysis (towards its use on issues such as sustainability)

National Accounts in Real Time & High Definition

National accounts in Real Time & HD



From cards and accounts movements…

… for Spain, Mexico, Turkey, Colombia, Argentina and Peru

Real Time Networks (Input Output tables)

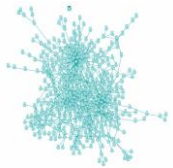
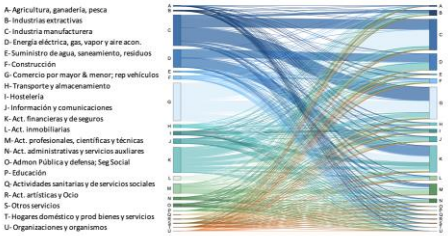


Tabla Input-Output Flujos Transferencias BBVA

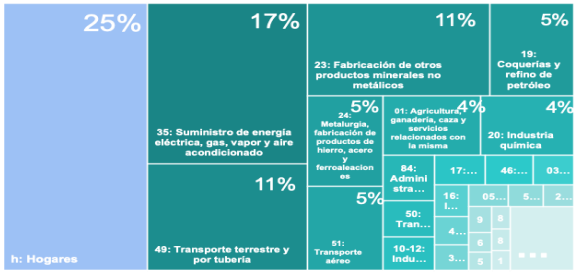
(CNAE, incluye sólo Clientes)



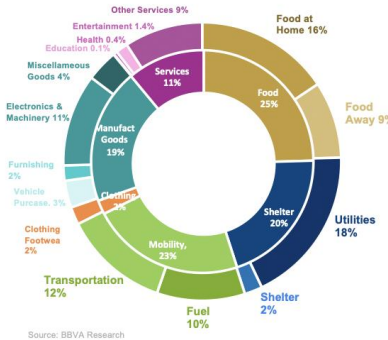
Fuente: BBVA Research

Analyzing the relationship between financial and real input-output tables

Sustainability in Real Time & HD



Spain: Distribution of household Co2 footprint by Items 2019 (% of total Co2 Emissions by Households)



From replicating aggregate series to providing granular data, not available(yet) using national accounts, more “micro” uses

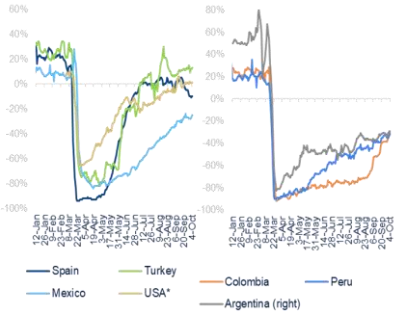
National Accounts in Real Time & High Definition

Consumption data

By sector

Health, Transport, Restaurants, Entertainment, Hotels, Food,…

Consumption in restaurants



Online / Offline

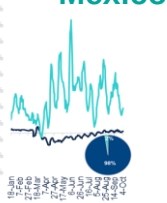
Performance of online vs face-to-face purchases

Spain

USA

Turkey

Mexico



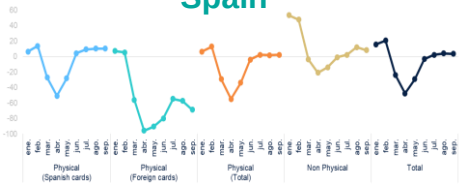
Card purchase / ATM withdrawal

Spain

Colombia

By nationality

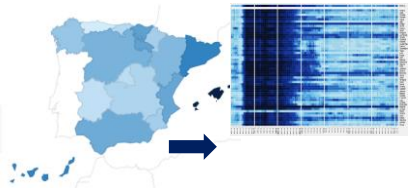
Spain

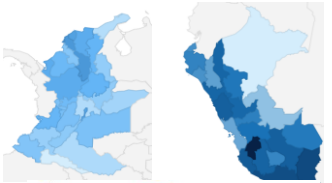


By geography

Regions, cities and even postal codes

Consumption by region



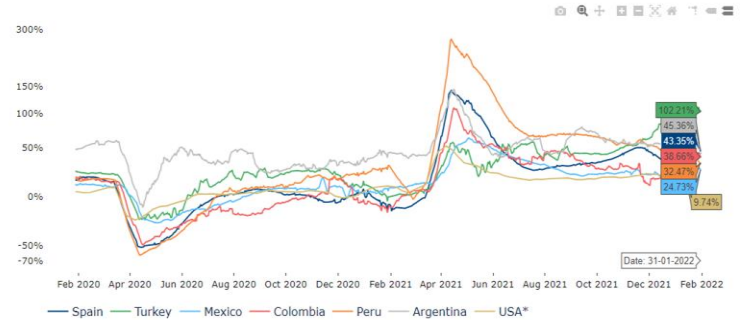


… reaching high frequency data in countries across categories

National Accounts in Real Time & High Definition

Big Data Consumption Indices: Total card consumption by country

(7D cum yoy)



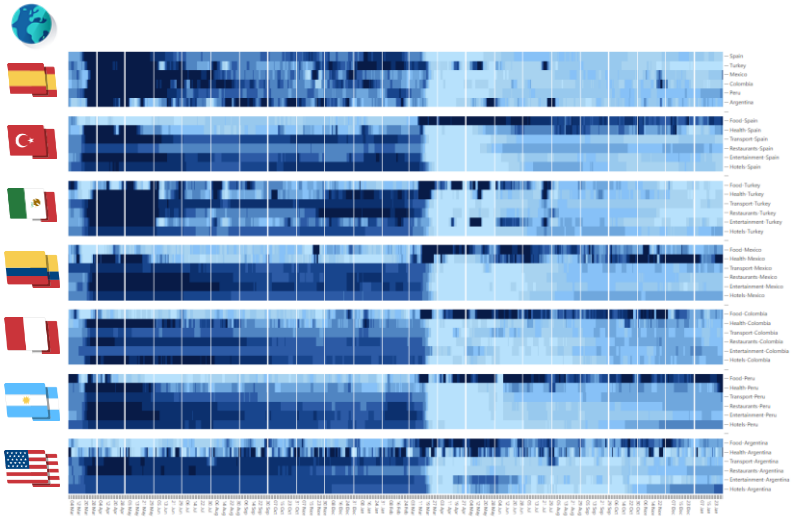
*Spending at BBVA POS by BBVA and non-BBVA customers plus spending by BBVA customers at non-BBVA POS. This data considers final expenditures, so not intermediate consumption is included. Therefore, the correspondence with national accounts household consumption is not fully equivalent. Reference for the methodology can be found in the publication "Tracking the COVID-19 Crisis with High-Resolution Transaction Data" and the weekly publications "The COVID-19 impact on Consumption in Real Time and High Definition". *USA data provided by the Federal Bank of Chicago CARTS index.*

Publication calendar for consumption according to the National Institutes of Statistics by country

Data may be analyzed between 1-3 months ahead than the official figures

Source: BBVA Research

Big Data Consumption Heat Map (7D cum yoy)



Series are coloured according to the history of each series. The darker blue represents the upper 90% percentile and the lighter one represents the lower 10% percentile. Reference for the methodology can be found in the publication "Tracking the COVID-19 Crisis with High-Resolution Transaction Data" and the weekly publications "The COVID-19 impact on Consumption in Real Time and High Definition".

Daily indicators are provided on our website with weekly updates

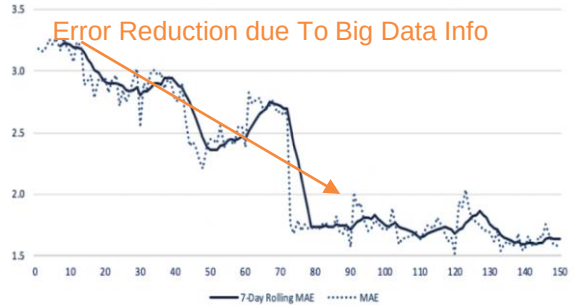
See Carvalho et al (2020) for further detail

Developing new models to exploit the advantage of real time but also information in high definition (sectoral Activity & Geography)

National Accounts in Real Time & High Definition

TURKEY: Big data information for Nowcasting and High Definition

Figure 7 Daily MAEs of equally weighted nowcast combinations between 2006Q1 and 2020Q3

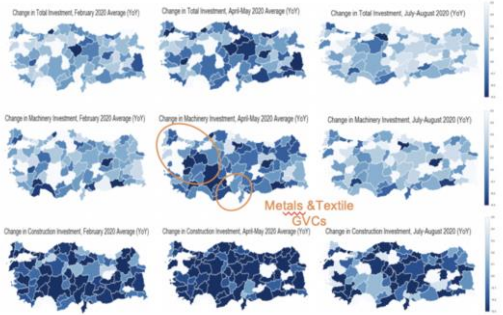


The Big Data Informatio give us 1-1/2 Months of advantage…

Figure 3 Big Data Investment Sectoral HeatMap
(% YoY Light Colours stand for positive growth rates and Dark Colours for negative rates)

… for a highly detailed performance of Fixed Assets …

Figure 4 Big Data Regional Investment Maps
(% YoY Light Colours stand for positive growth rates and Dark Colours for negative rates)



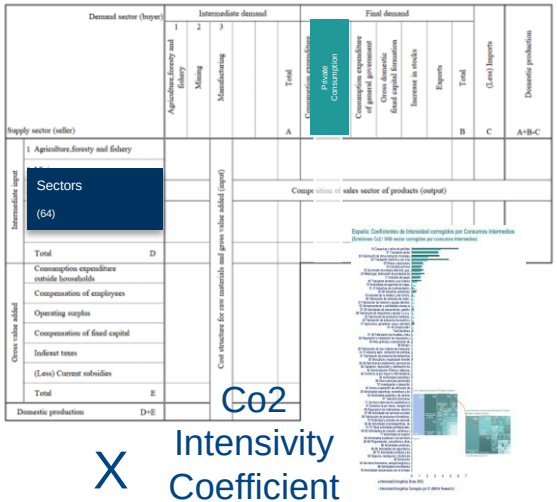
… at a Provincial (or higher) Level…
Given geolocalized information

Big Data Information and Nowcasting: Consumption and Investment from Bank Transactions in Turkey
(Arxiv version)

Estimating the household Co2 Footprint from the input-output analysis, distributing sectoral production to consumption categories

Sustainability in Real Time & High Definition

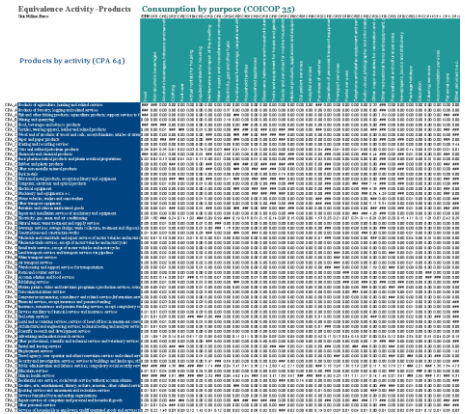
SPAIN: INPUT-OUTPUT TABLES



X Co2 Intensivity Coefficient (Co2 Kg/GVA)

A mix of IO tables & Sustainability cooefficients

SPAIN: EQUIVALENCE PRODUCTS-CONSUMPTION

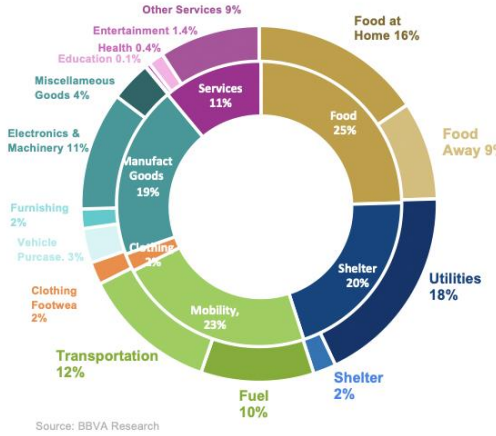


16 Categorias

64 Sectores

We translate Products in. IO to Consumption items of Households

SPAIN: DISTRIBUTION HOUSEHOLDS CO2 FOOTPRINT



We build up distributions of Co2 Footprints of Household Consumers

Analyzing Cash Vs Card Consumption Patterns in Mexico: A Machine Learning Approach

Structural & Policy Analysis in real time

Joint collaboration with Clarity AI (paper)

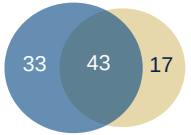
Card data

BBVA

card spending using BBVA cards transactions

- 3.2 billion transactions
- 32 states
- 344 types of consumption categories.

Data for 2017-2018
municipalities



Cash data

frogtek

purchases made in cash at convenience stores

- 62 million transactions
- 12 states in Mexico
- 171 types of products
- 1835 shops

CARD SPENDING BUDGET SHARE BY MUNICIPALITY (2017 vs 2018)

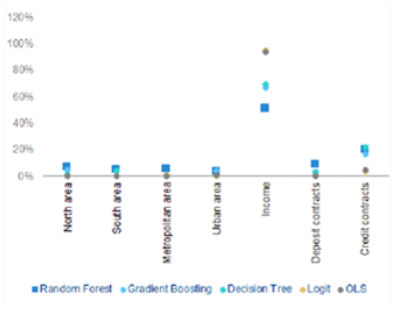
CASH SPENDING BUDGET SHARE BY MUNICIPALITY (2017 vs 2018)

Inside the black box of machine learning: achieving model explainability based on Shapley values, capturing non linear relationships to explain consumption by payment channel

Structural & Policy Analysis in real time

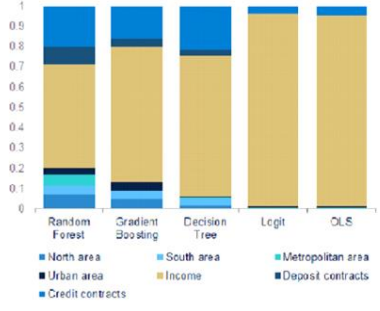
CONSUMPTION PATTERNS IN CASH

Figure 13. SHAPLEY SHARE COEFFICIENTS FOR CASH SPENDING MODELS BY VARIABLE



Source: own calculations. BBVA and Clarity. BBVA and Clarity. Models' order is based on models performance, from the best performance (random forest) to the worst one (OLS).

Figure 14. SHAPLEY SHARE COEFFICIENTS FOR CASH SPENDING MODELS BY MODEL



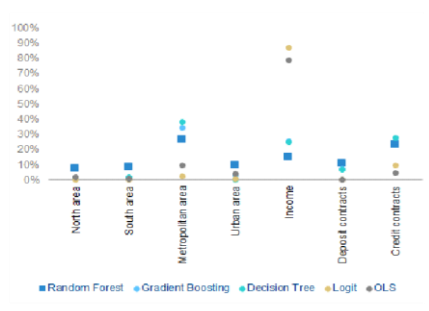
Source: own calculations. BBVA and Clarity. BBVA and Clarity. Models' order is based on models performance, from the best performance (random forest) to the worst one (OLS).

In line with the traditional theory, income remains the key variable in both linear and non linear model in the case of cash spending

[Cash Vs Card Consumption Patterns in Mexico: A Machine Learning Approach]

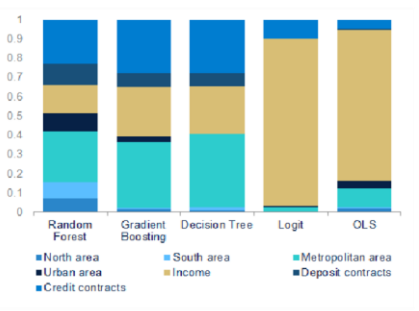
CONSUMPTION PATTERNS IN CARD

Figure 10. SHAPLEY SHARE COEFFICIENTS FOR CARD SPENDING MODELS BY VARIABLE



Source: own calculations. BBVA and Clarity. Models' order is based on models performance, from the best performance (random forest) to the worst one (OLS).

Figure 11. SHAPLEY SHARE COEFFICIENTS FOR CARD SPENDING MODELS BY MODEL



Source: own calculations. BBVA and Clarity. BBVA and Clarity. Models' order is based on models performance, from the best performance (random forest) to the worst one (OLS).

However, in card spending income is very relevant in the linear model, but urban and financial deepening are captured better in non-linear specifications

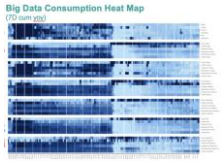
There are important advantages to using Big Data but we are in the early processes to guarantee *Robustness & Best Practices….*

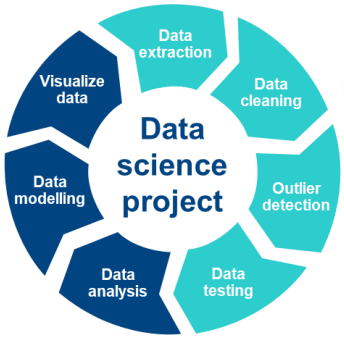
Pros — **Big Data** (Bank Transactions Data) — **Cons**

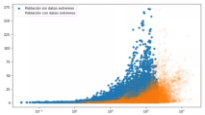
- Real Time
- High definition (GeoLocalizacion)
- Rapid Response
- Smart Policies
- …

- Balance Samples
- Stability
- Representativeness
- Seasonality & Outliers Treatment
- …

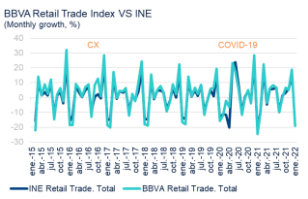
More than 60 % of the time devoted to Robustness (minimize the cons)

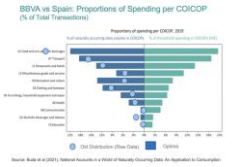






	Whole Series		Pre-COVID Series	
	RMSE	Δ RMSE	RMSE	Δ RMSE
Optimal	0.0250	0	0.0222	0
No Active Customer	0.1408	0.1158	0.1527	0.1305
Exclude Cash	0.0685	0.0435	0.0560	0.0338
Include Online	0.0272	0.0022	0.0263	0.0041
No Demographic Weighting	0.0264	0.0014	0.0238	0.0016
Non-consumption MCC	0.0254	0.0004	0.0224	0.0003
Raw Data	0.0835	0.0585	0.0677	0.0455

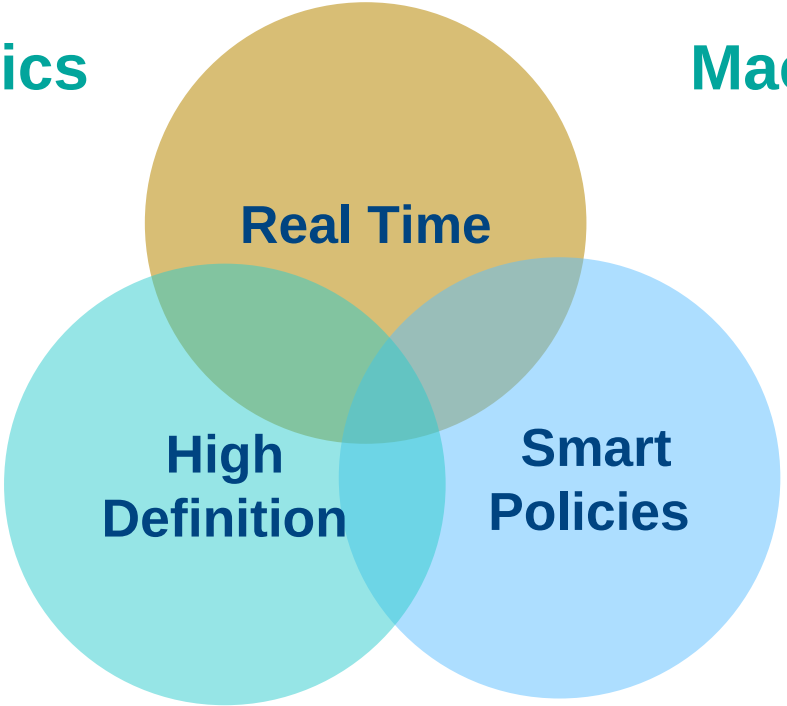




It is going to be one of the new frontiers of research, there's a huge potential to exploit

Microeconomics

Macroeconomics



Behavioural economics

Real time and granular analysis for policy response and evaluation

Disclaimer

The present document does not constitute an “Investment Recommendation”, as defined in Regulation (EU) No 596/2014 of the European Parliament and of the Council of 16 April 2014 on market abuse (“MAR”). In particular, this document does not constitute “Investment Research” nor “Marketing Material”, for the purposes of article 36 of the Regulation (EU) 2017/565 of 25 April 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council as regards organisational requirements and operating conditions for investment firms and defined terms for the purposes of that Directive (MIFID II).

Readers should be aware that under no circumstances should they base their investment decisions on the information contained in this document. Those persons or entities offering investment products to these potential investors are legally required to provide the information needed for them to take an appropriate investment decision.

This document has been prepared by BBVA Research Department. It is provided for information purposes only and expresses data or opinions regarding the date of issue of the report, prepared by BBVA or obtained from or based on sources we consider to be reliable, and have not been independently verified by BBVA. Therefore, BBVA offers no warranty, either express or implicit, regarding its accuracy, integrity or correctness.

This document and its contents are subject to changes without prior notice depending on variables such as the economic context or market fluctuations. BBVA is not responsible for updating these contents or for giving notice of such changes.

BBVA accepts no liability for any loss, direct or indirect, that may result from the use of this document or its contents.

This document and its contents do not constitute an offer, invitation or solicitation to purchase, divest or enter into any interest in financial assets or instruments. Neither shall this document nor its contents form the basis of any contract, commitment or decision of any kind.

BBVA Research

Big Data in Economics:
our journey in BBVA Research

Facultad de Ciencias Económicas y Empresariales
Universidad Complutense de Madrid

Jorge Sicilia
Chief economist BBVA

February 2022

Creating Opportunities