National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption

Gergely Buda (BSE) , Vasco M. Carvalho (University of Cambridge & CEPR)
Stephen Hansen (Imperial College & CEPR), Alvaro Ortiz (BBVA Research)
Tomasa Rodrigo (BBVA Research) & José V. Rodríguez Mora (University of Edinburgh & CEPR)

Summary

- Introduction

- Building a Consumption Survey

- National Accounts

- Distributional Accounts

- Potential: Economics of Real Time, High Definition & Smart Policies

- Conclusion

Introduction: Why Use Transaction Data

• **Modern payment systems** generate an enormous amount of real-time data on activity which can be **incorporated into national accounting measures** (Bean 2016).

• Advantages of such data include **timeliness, granularity, and cost** to statistical agencies (albeit not to private sector).

• This is **important for lower-income countries**, where transaction data may be the only reliable source of information for building national accounts.

• Growing **interest in non-traditional data for tracking the economy**, especially in the wake of COVID-19.

• But few, if any, attempts in the academic literature to **build national accounting objects from first principles using large-scale payment data.**

Introduction: First Proof of Concept on Naturally Ocurring Data

First proof of concept that naturally occurring transaction data, arising through the decentralized activity of millions of economic agents, can be organized via national accounting rules and then harnessed to produce a large-scale, high-quality and highly-detailed consumption survey

- Universe of BBVA retail accounts in Spain by BBVA

- Allowing us to track expenditure as it flows out of these accounts, transaction by transaction

- 3 billion individual transactions by 1.8 million BBVA customers, from 2016 to 2021

Contributions: Four contributions of the Paper

1. How to **construct representative panel of household expenditure**. Massive survey.

2. Show that it **aggregates to Quarterly National Accounts**

3. Create **Distributional National Accounts for Consumption**

4. Study the **Micro-Structure of consumption dynamics** (just an example !!):

Building a Consumption Survey

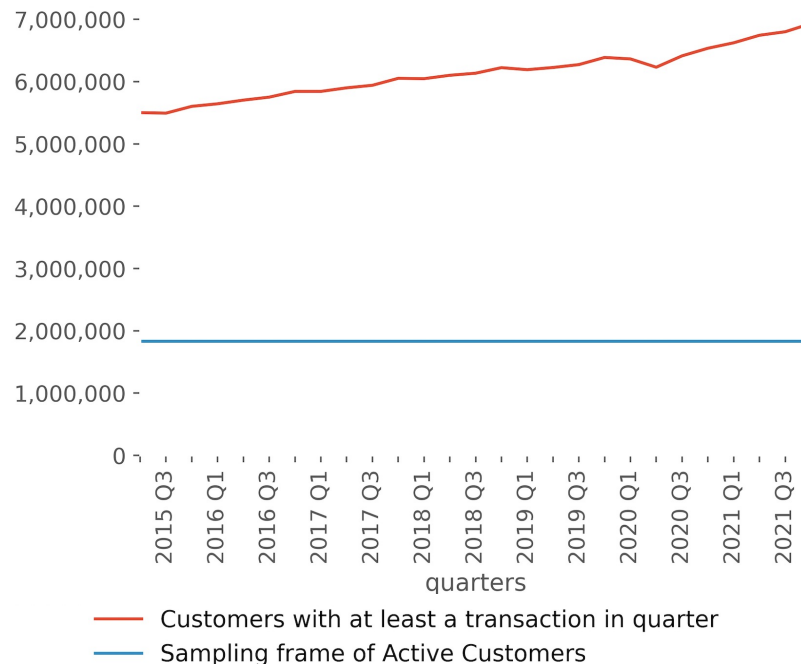# Building a Consumption Survey: Key Problems

- Two problems for translating transaction data into a representation of the population:

    o The client pool of a **Bank ≠ Population**: **Biases**

    o Spending is not the same than consumption: **Spending ≠ Consumption**

Building a Consumption Survey: Active Customers

Sample Frame

- 10,270,041 unique customers (2015-2021)
- Most spend infrequently or for short periods only.
- Define "Active Customers" as making at least 10 consumption related transactions in each quarter.
- 1,827,866

Customers with transactions and sampling frame of active customers

Legend:
— Customers with at least a transaction in quarter
— Sampling frame of Active Customers

Building a Consumption Survey: Non-Housing Expenditures

Classification of Non-Housing Consumption Spending

- If a transaction is explicitly categorized in one of the 12 COICOPS.

- Follow national accounting principles wherever possible

Card Data
- Merchant Client Code (MCC) of the counterparty firm.
- Manual Mapping to COICOPS
- Multi-product retailers. Assigned by external data on distributions.

Direct Debit
- ~ 100 internal labels.
- Manual Mapping
- When this is unclear, we read field, determine firm and use either MCC (if possible) or NACE code of firm to assign COICOP.

Transfers
- String match counter-party name to commercial registry.
- If counter-party is located as a firm, we assign as above.

Cash
- Both cash and over the counter.
- Assume is consumption.
- Assumptions on distribution.

Spending Category	Volume of Transactions	Number of Transactions
Offline Card Transactions	60,319 million	1,772 million
Online Card Transactions	11,858 million	313 million
Direct Debits	66,036 million	752 million
Cash Withdrawal	64,592 million	359 million
Transfers excl. rent	11,148 million	15 million

Building a Consumption Survey: Housing Consumption (Rents)

Determination of
Housing Consumption Spending

- We locate payment of rental for housing services.
  - Reading of free-text field in direct debits and transfers.
  - Minimum 100 EUR
  - Exclude parking, etc.
  - Payments made in 70 months.
  - 32,127 households.
- Use household covariates to predict monthly rent
  - Income (from BBVA table, six month average)
  - Utility Payments (direct debits)
  - Geography: 327 regions (consolidating postal codes)

Variable	Model	Test set
Spending on House Utilities	0.0884	
	(0.0008)	
Income	0.0362	
	(0.0011)	
N of Contract Groups	16,977	15,512
N of Observations	1,134,735	15,512
R^2	0.3911	
Adjusted R^2	0.3765	
Within R^2	0.1200	
Root MSE	204.6144	221.64

Out-of-sample behavior is reasonable with households that are 50-70 months in data.

Use covariates to IMPUTE housing consumption for the rest of the househods (the vast majority)

Building a Consumption Survey: Estimating Households

Households Proxy vs Official Data
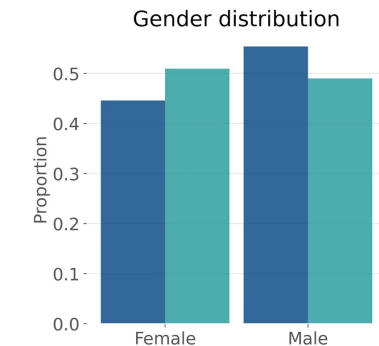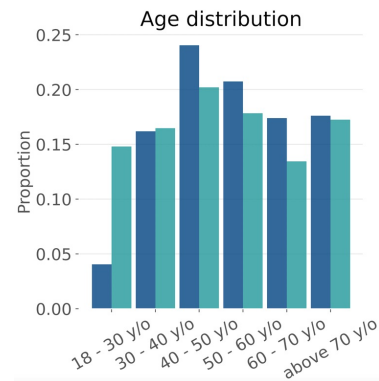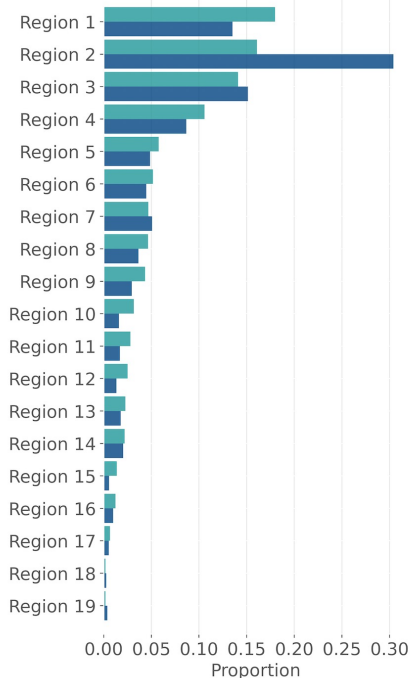(Customers with transactions and sampling frame of active customers)

- Link clients into perceived house-hold groups.
- Individuals with whom they share a contract and live in same postal code.
- 1,589,280 household groups

INE Continuous Household Survey - Microdata
Cosigner Groups with Active Customers in Naturally Occurring Data
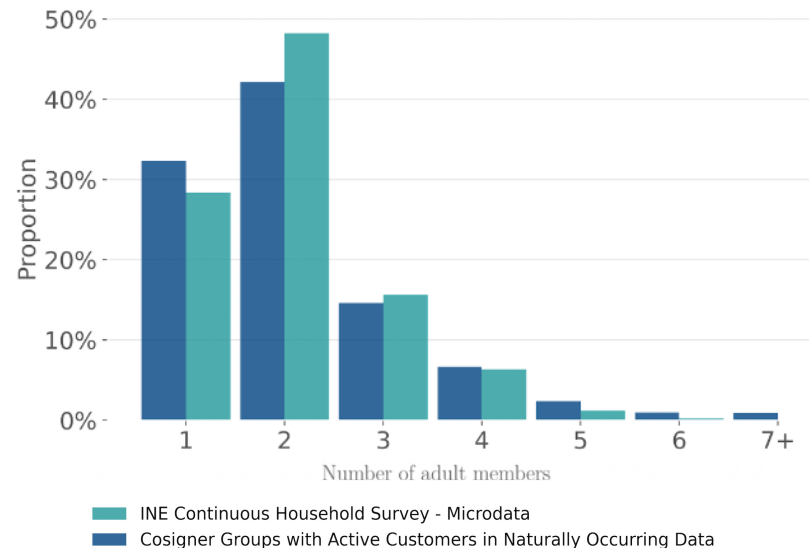
	HBS 2016	HBS 2017	HBS 2018	HBS 2019	HBS 2020	BBVA Sample
Households	22,011	22,043	21,395	20,817	19,170	**1,589,280**
Adults	47,420	47,055	45,328	43,988	40,285	**1,827,866**

Building a Consumption Survey: Demographics of Active Clients

Demographics of Active Customers

$$c_{g,a,r} = c_{g,a,r}^{hh} \left(\frac{x_{g,a,r}^{INE}}{x_{g,a,r}^{BBVA}} \right)$$

Regional distribution (Autonomous Communities)

Age distribution

Gender distribution

INE Spain Demographics   Customer Demographics in Naturally Occurring Data

Building a Consumption Survey: Weighting & Sampling

Weighting & Sampling

- We observe **Spending not Consumption** within household

- Assume **equal spending among active clients within Households**, and **half the weight of non Active Clients**

- Define cells **Gender, Age and Region (5 geographical units per province)**

- Adjust Demographic Weights of cells to **Make them Representative**

$$c_{g,a,r} = c_{g,a,r}^{\text{hh}} \left(\frac{x_{g,a,r}^{\text{INE}}}{x_{g,a,r}^{\text{BBVA}}} \right)$$

- In occasion we need to create national simple

- We draw $x_{g,a,r}^{INE}$ times from the pool of active client IDs within cell

- Sampling with Replacement

Building a Consumption Survey: Payment-Consumption Connection

Connecting Payments with Consumption

COICOP Shares by Payment Method

Share of Spending in COICOP by Payment Method

Legend:
- 01 Food and Non-Alcoholic Beverages
- 02 Alcoholic Beverages and Tobacco
- 03 Clothing and Footwear
- 04 Housing, Water, Electricity, Gas
- 05 Furnishings, Household Equipment
- 06 Health
- 07 Transport
- 08 Communication
- 09 Recreation and Culture
- 10 Education
- 11 Restaurants and Hotels
- 12 Miscellaneous Goods and Services
- Uncategorizable
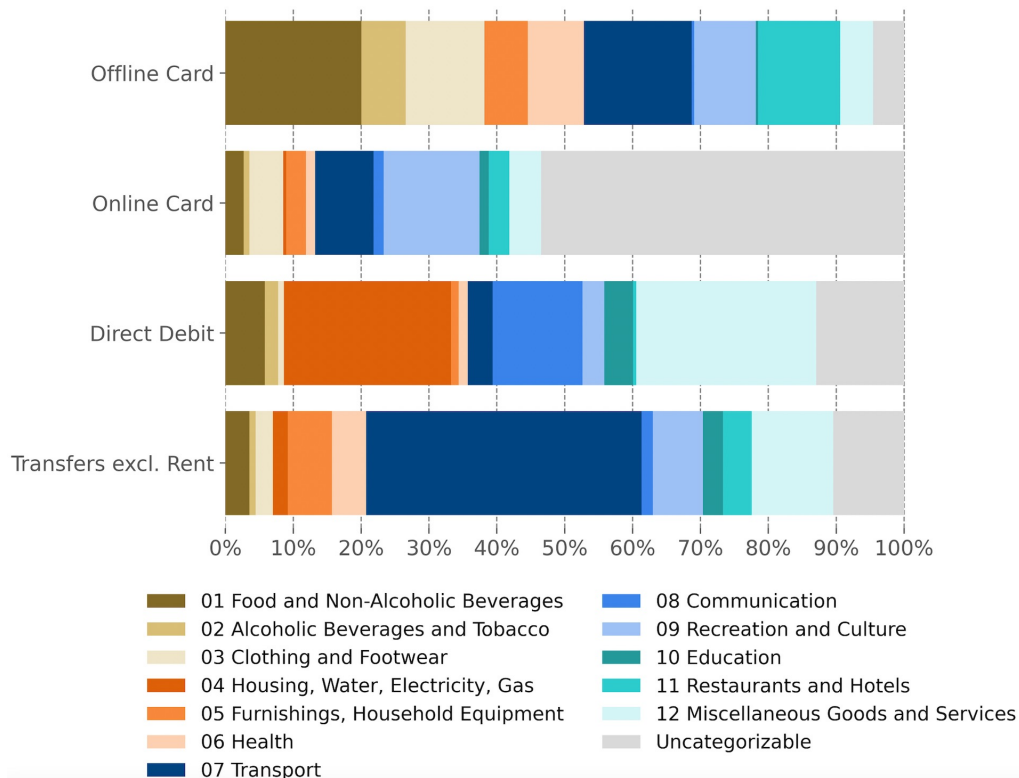
Aggregating To National Accounts

National Accounts: Similar Levels

Total Consumption Private: Naturally Ocurring vs Official
(level Mill Euros)

— Total Private Consumption in Naturally Occurring Data
— INE Quarterly Accounts: Household Final Consumption
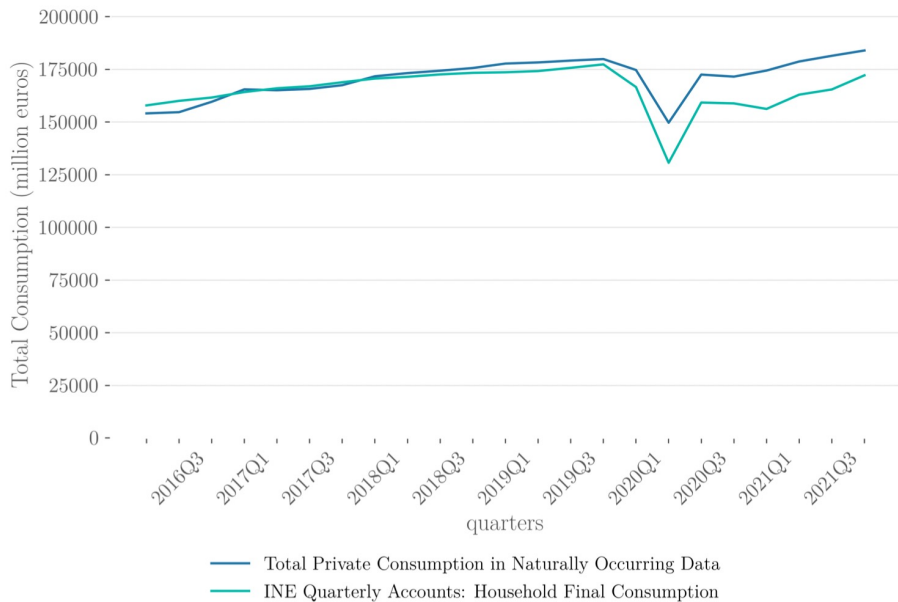
Total Private Consumption: Naturally Ocurring vs Of
(QoQ Sadj Growth Rate)

— Total Private Consumption in Naturally Occurring Data
— INE Household Final Consumption
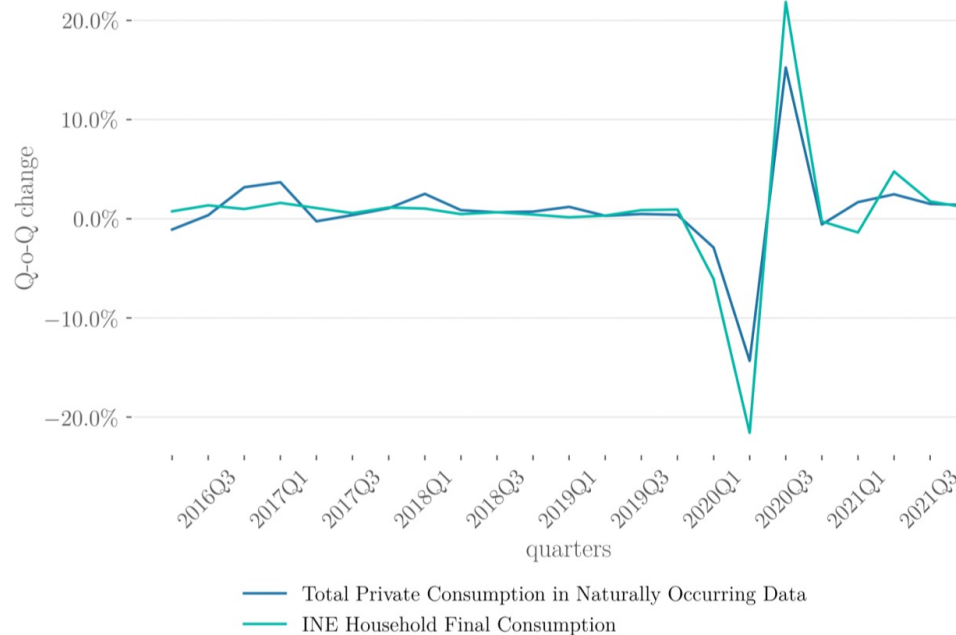
National Accounts: COICOP Distributions in HBS and N. Accounts

Distribution of Spending across COICOP Categories

(a) BBVA vs HBS
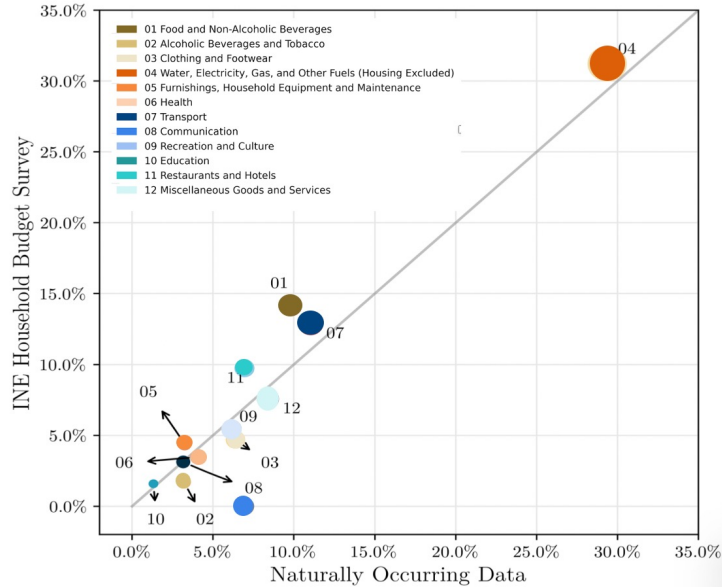
(b) BBVA vs National Accounts
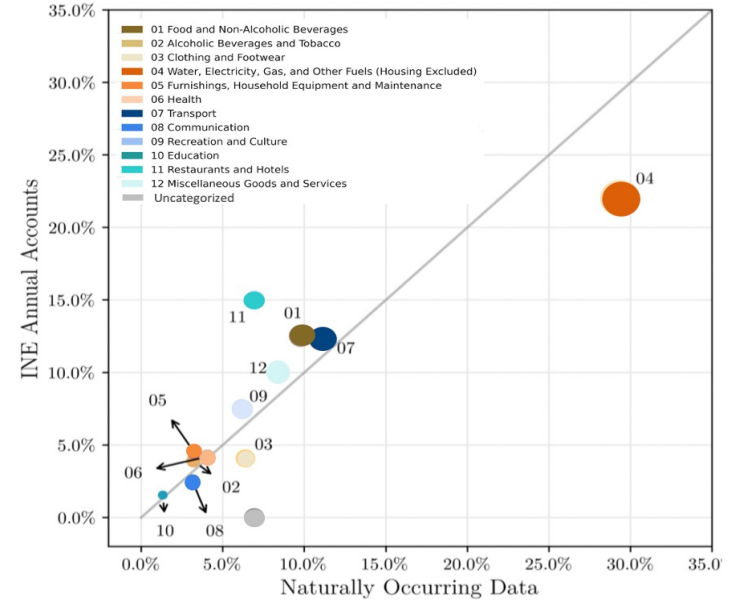
National Accounts: Richer Dynamics (Daily) by Payment & Category

Proportions of Consumption by Payment Method
(Cumulative 3M, % of Total)

Legend:
- Cash
- Housing Rents
- Direct Debit
- Card Offline
- Card Online
- Money Transfer
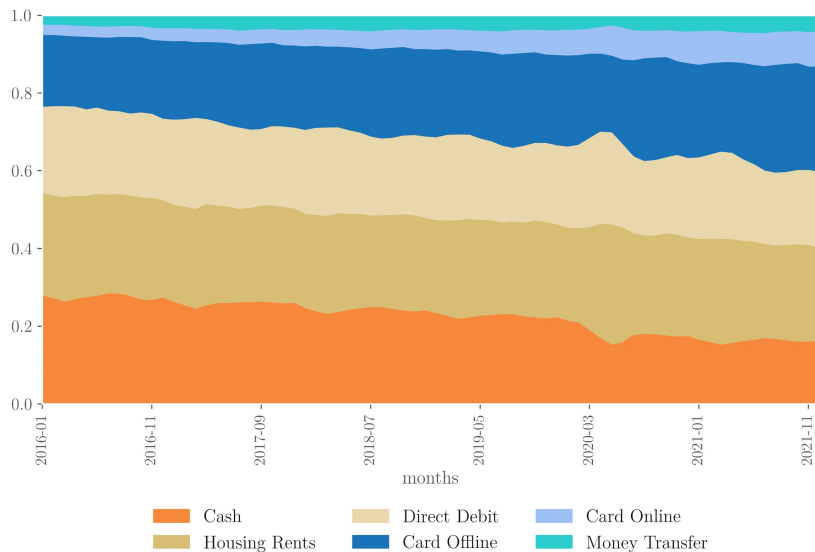
Proportions of Consumption by COICOP
(Cumulative 3M, % of Total)

Legend:
- 01 Food and Non-Alcoholic Beverages
- 02 Alcoholic Beverages, Tobacco, and Narcotics
- 03 Clothing and Footwear
- 04 Housing, Water, Electricity, Gas, and Other Fuels
- 05 Furnishings, Household Equipment and Maintenance
- 06 Health
- 07 Transport
- 08 Communication
- 09 Recreation and Culture
- 10 Education
- 11 Restaurants and Hotels
- 12 Miscellaneous Goods and Services
- Non-categorizable

Creating Distributional National Accounts For Consumption

Distribut. Accounts: Consumption´s Macro-Consistent distribution

Yearly adult consumption by levels of consumption in 2017;
(Weighted sampling procedure, 2017 Euro by percentiles of consumption)

- Macro-consistent, distribution of Consumption.

- It aggregates into NA

- Follows Picketty et al 2018.
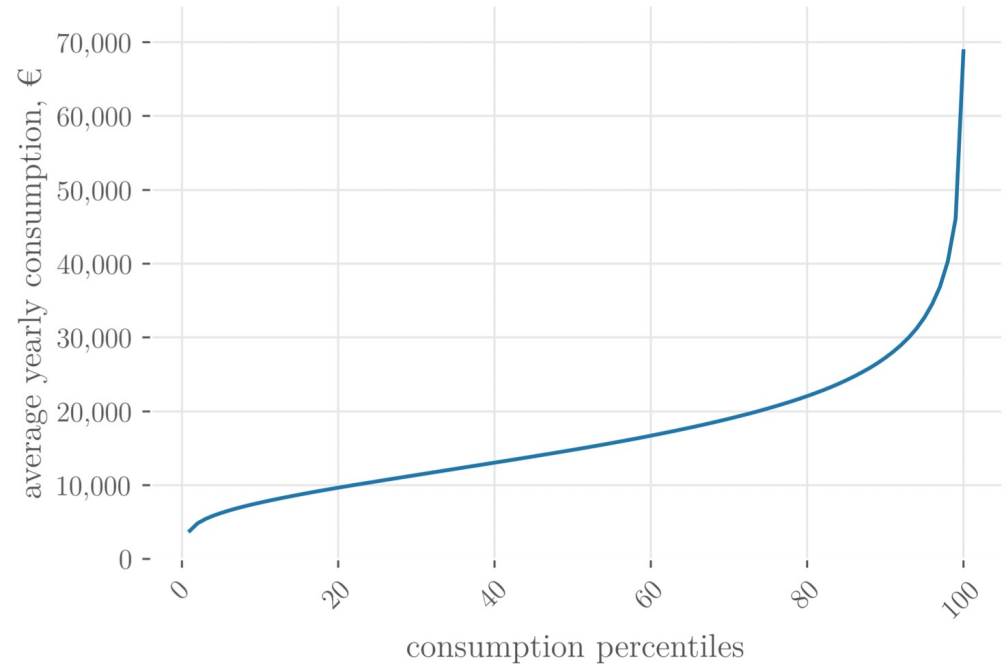
Distributional Accounts: Differences Consumption and Income

Consumption Distribution per Spanish adult in 201 BBVA vs. WID 2017 post-tax income distribution.
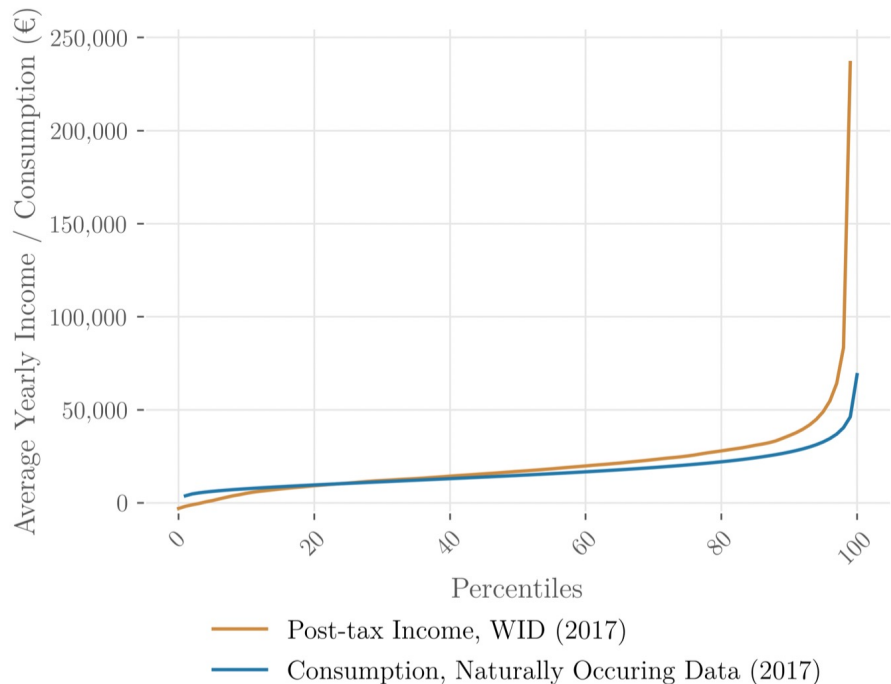
Lorenz Curves of BBVA 2017 consumption and WID 2017 post-tax income in Spain.
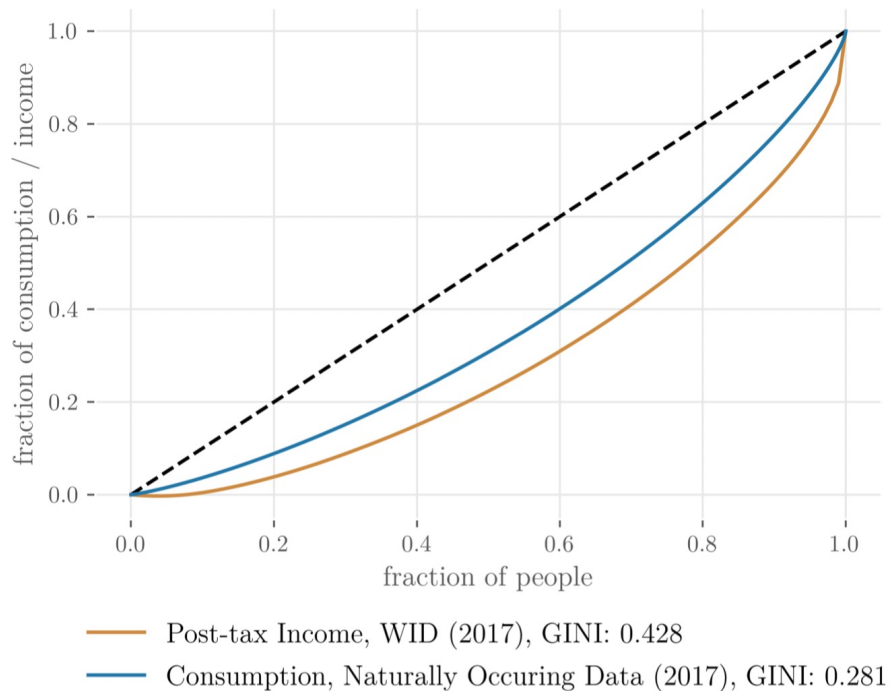
Distributional Accounts: Right Tail Difference (in Level)

Comparison Distribution of Consumption with Household Business Survey (HBS)

- Remember: difference in level.

- Right tail difference: Bigger share of consumption among the people who consumes most.

Distribution of Consumption per adult in BBVA vs. Spanish Household Budget Survey consumption per adult distribution 2017
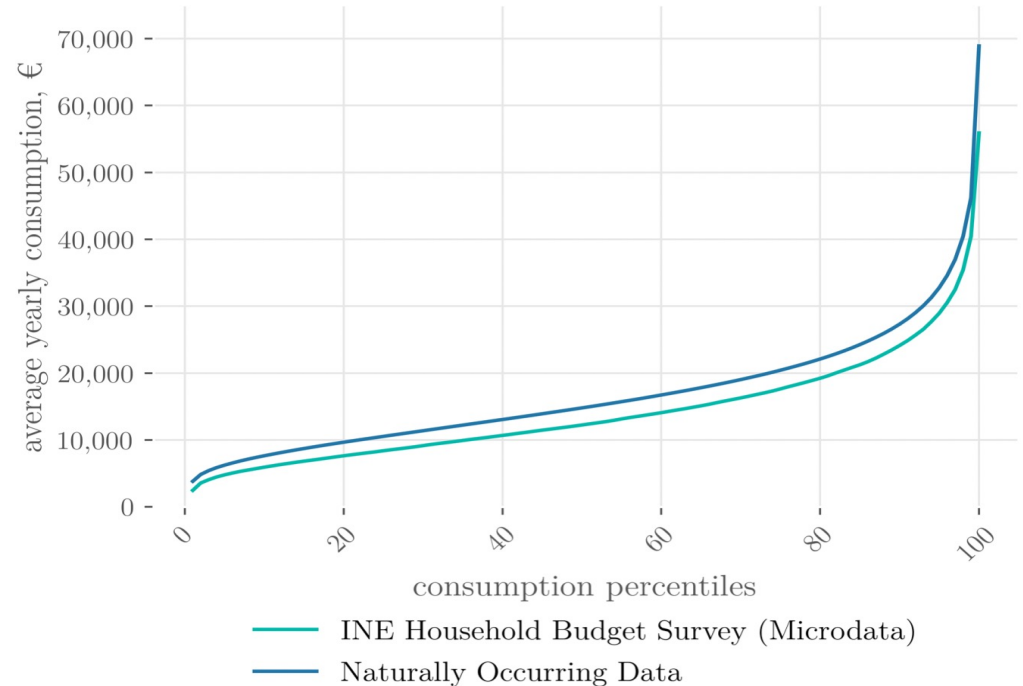
Distributional Accounts: Different Categories by Consumers

Distribution across
COICOPs and People

Consumption distribution disaggregated by COICOP consumption categories

Distributional Accounts: Time Frequency Matters

Lorenz Curves of the distribution of consumption across different time frequencies.

- Across Time frequencies
- Consumption is a flow.
- Unite of Time aggregation matters.
- Lumpy.
- Critical in survey design.
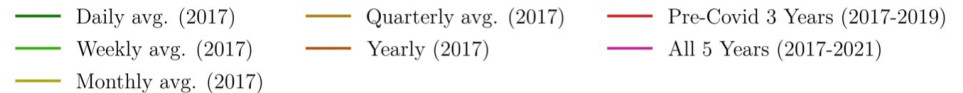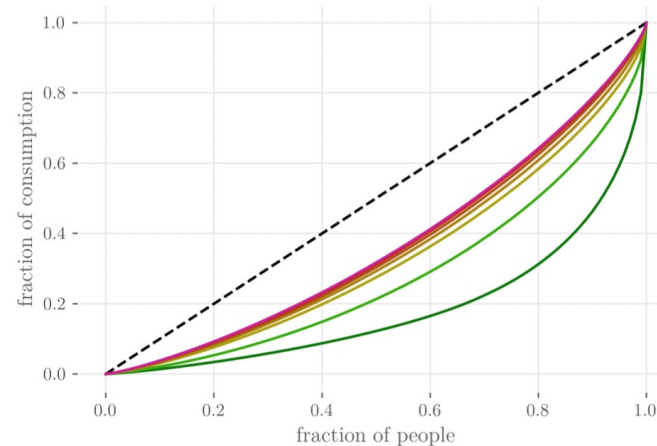
Frequency	Gini index
Daily avg. (2017)*	0.629
Weekly avg. (2017)	0.439
Monthly avg. (2017)	0.338
Quarterly avg. (2017)	0.307
Yearly (2017)	0.281
Pre-Covid 3 Years (2017-2019)	0.273
All 5 Years (2017-2021)	0.265

*30 days sampled randomly

Distributional Accounts: Consumption Growth Distribution

Distribution of Consumption Growth

- Consumption Growth Inequality not so apparent in normal years (2017-2019)

- But increasingly apparent during Covid (2019-2020)…

- ..and Post-Covid Recovery (2020-2021)

Growth Rate of Consumption per Percentiles
### (cross-sections)

Legend:
- Pre COVID 2017 to 2019
- 2017 to 2019 mean growth
- Post COVID 2019 to 2020
- 2019 to 2020 mean growth
- Post COVID 2020 to 2021
- 2020 to 2021 mean growth
- Whole Period 2017 to 2021
- 2017 to 2021 mean growth
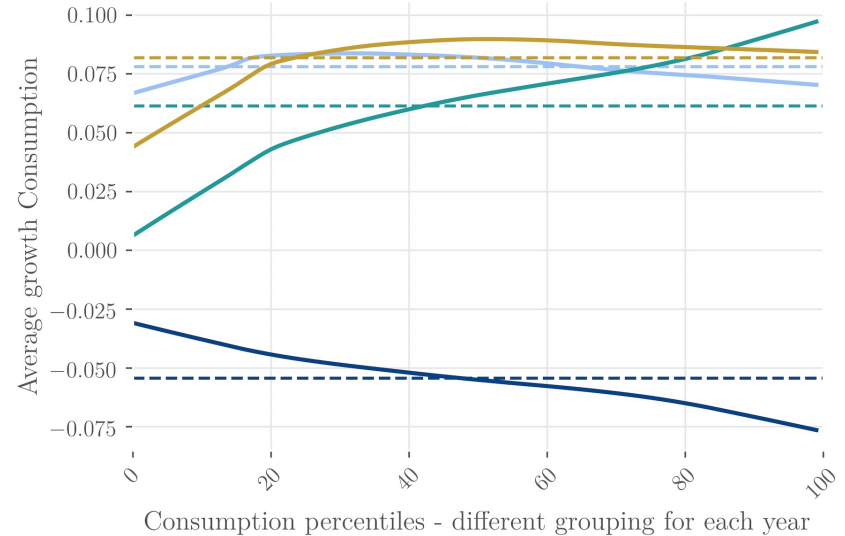
Distributional Accounts: Consumption Growth Dist. by Categories

Relying only on Cards can lead to biases towards Consumption inequality

(only increase in a couple of COICOPS in Normal years)
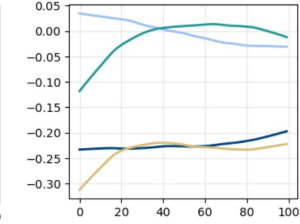
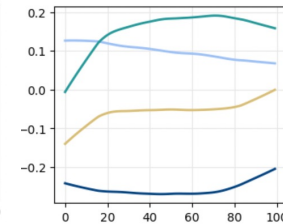### Growth Rate of Consumption per Percentiles by COICOP Consumption Group
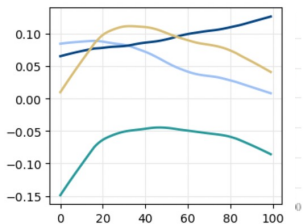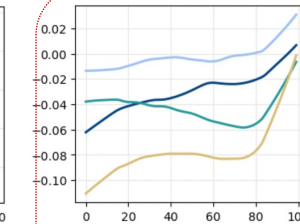

(a) Food and Non-Alcoholic Beverages
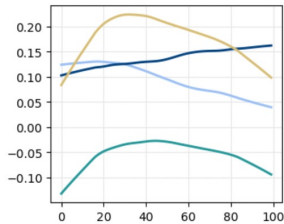
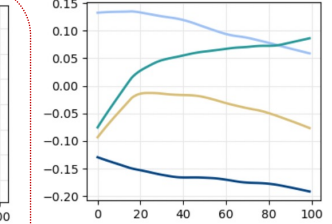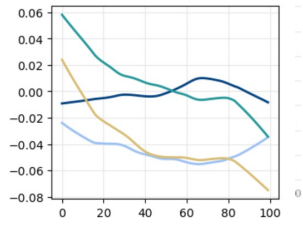(b) Alcoholic Beverages, Tobacco, and Narcotics

(c) Clothing and Footwear
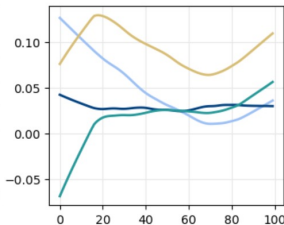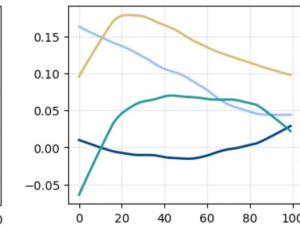
(g) Transport

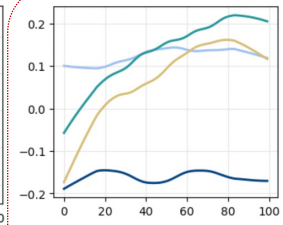(h) Communication

(i) Recreation and Culture

(d) Utilities (excluding imputed housing rent)

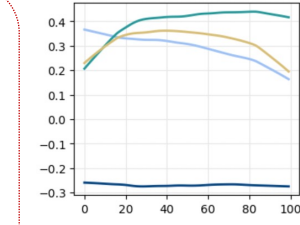(e) Furnishings, Household Equipment, and Routine House-
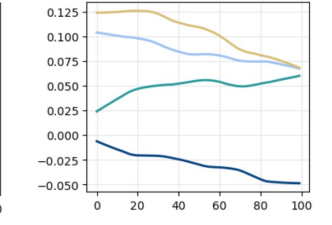
(f) Health

(j) Education

(k) Restaurants and Hotels

(l) Miscellaneous Goods and Services
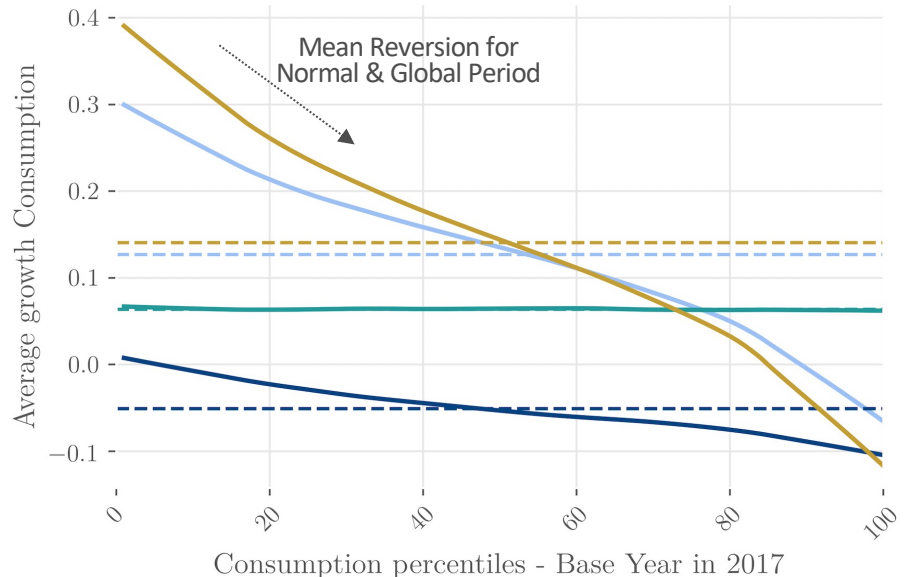
Pre COVID 2017 to 2019
Post COVID 2019 to 2020
Post COVID 2020 to 2021
Whole Period 2017 to 2021

Individual Consumption Dynamics: Panel Structure Individual Level

Distribution of Consumption Growth

- Panel Structure at individual level.

- Very different from Cross-Sections
- In one year (pre-COVID), massive mean reversion
- 2020, 2021, much flatter... becase of rapid reversion during first year.

Growth rates conditional on initial consumption percentile
(Growth Between Years for Consumption percentiles of 2017))

Mean Reversion for Normal & Global Period

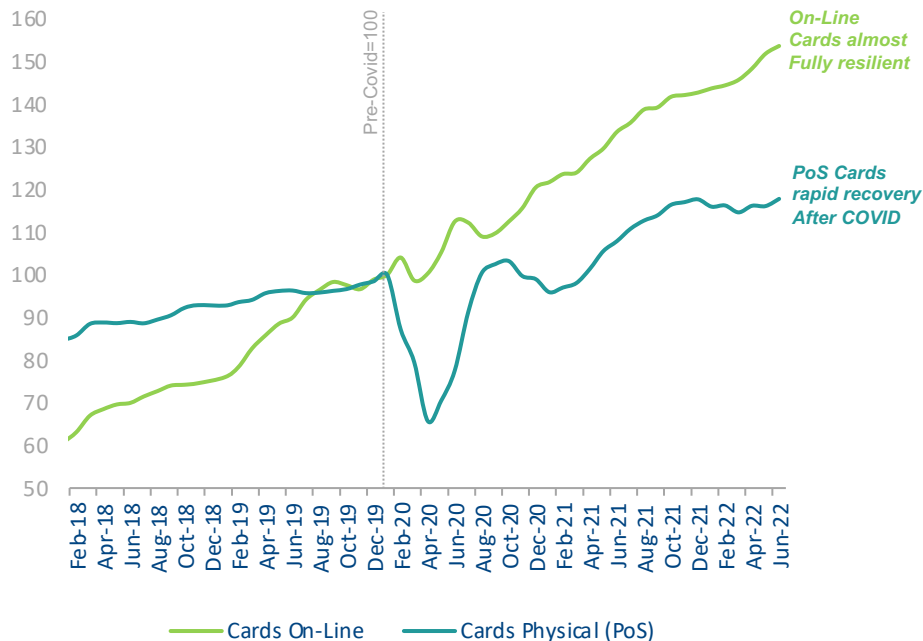|	
Pre COVID 2017 to 2019	Post COVID 2020 to 2021
2017 to 2019 mean growth	2020 to 2021 mean growth
Post COVID 2019 to 2020	Whole Period 2017 to 2021
2019 to 2020 mean growth	2017 to 2021 mean growth

Potential Uses:

The Economy in Real Time, High Definition and Design of Smart Policies

Cards payments (PoS & OnLine) proved to be very useful to track activity

Spain: Evolution of Payments for Consumption
(SAdj. Cum 3M January 2020=100)

On-Line
Cards almost
Fully resilient

PoS Cards
rapid recovery
After COVID

Pre-Covid=100

Cards On-Line   Cards Physical (PoS)

Spain: Distribution of Payments for Consumption
(Cumulative June 2020-2021, % of Total)

Cards
On-Line
11%

Cards
Physical
27%

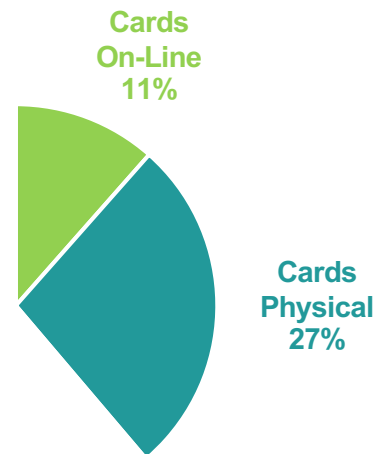# But this view is partial (38%) and biased (upwards) after the COVID..

Spain: Evolution of Payments for Consumption
(SAdj. Cum 3M January 2020=100)
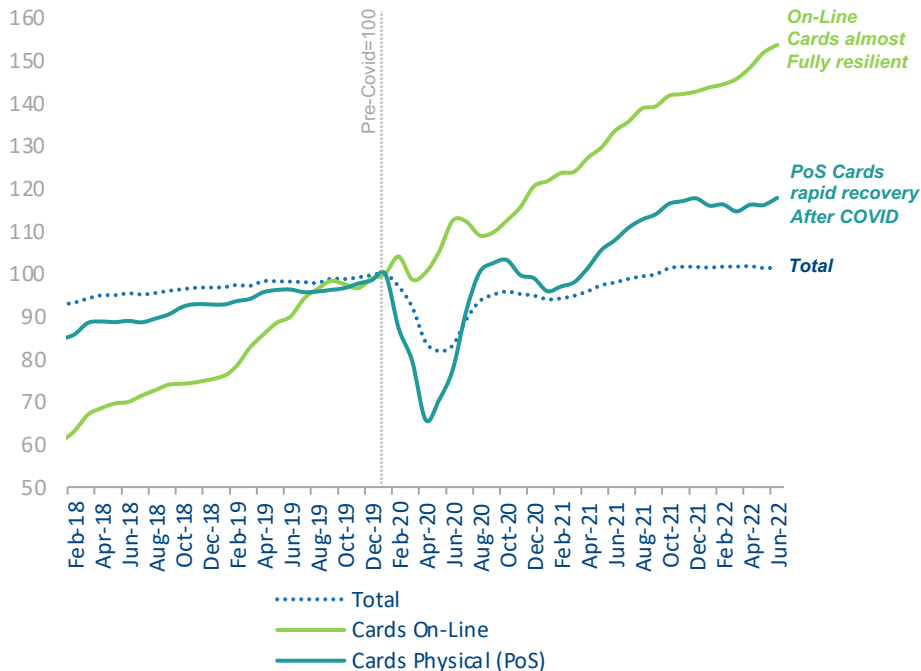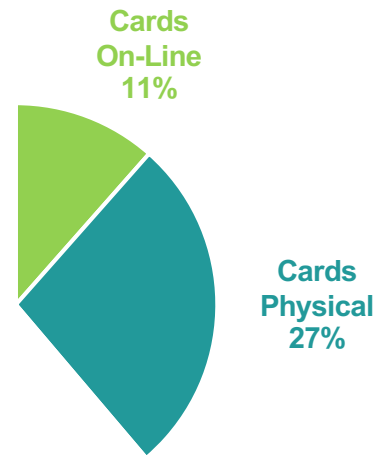
Spain: Distribution of Payments for Consumption
(Cumulative June 2020-2021, % of Total)

On-Line
Cards almost
Fully resilient

PoS Cards
rapid recovery
After COVID

Total

Pre-Covid=100

Legend:
- Total
- Cards On-Line
- Cards Physical (PoS)

Cards
On-Line
11%

Cards
Physical
27%

Source: BBVA Research & Buda et Al (2022).National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. Forthcoming

Cards can be misleading... Leading to an optimistic view

Spain: Evolution of Payments for Consumption
(SAdj. Cum 3M January 2020=100)
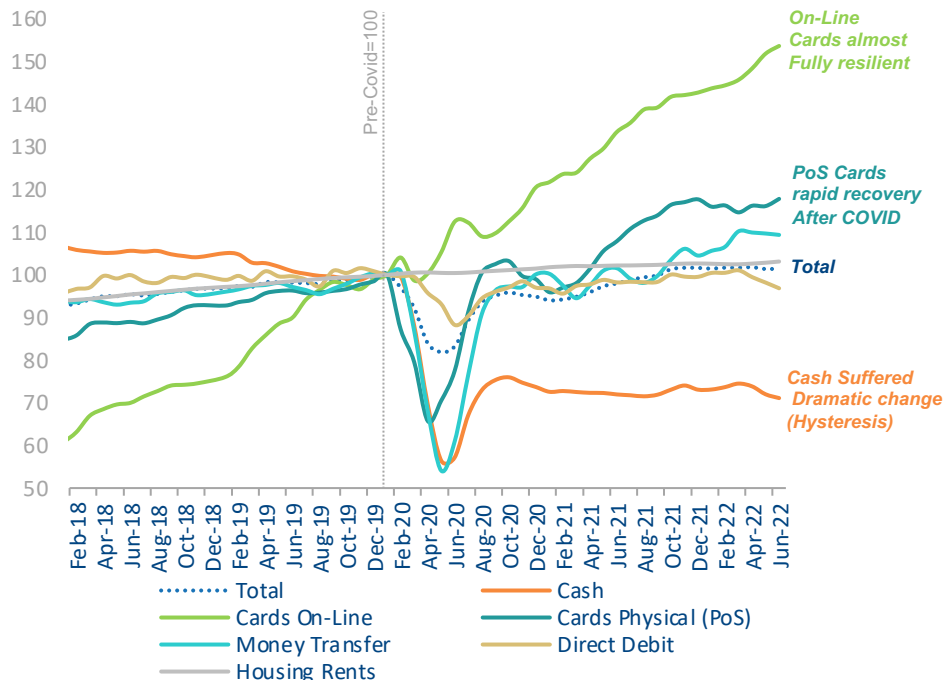
Pre-Covid=100

On-Line Cards almost Fully resilient

PoS Cards rapid recovery After COVID

Total

Cash Suffered Dramatic change (Hysteresis)

Legend:
- Total
- Cash
- Cards On-Line
- Cards Physical (PoS)
- Money Transfer
- Direct Debit
- Housing Rents

Spain: Distribution of Payments for Consumption
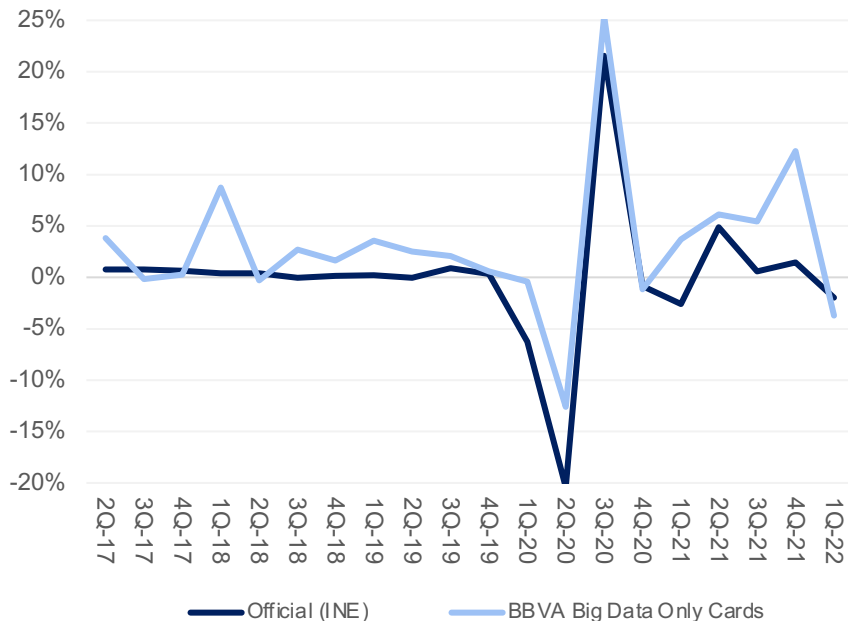### (Cumulative June 2020-2021, % of Total)

- Cash 14%
- Cards On-Line 11%
- Cards Physical 27%
- Money Transfer 5%
- Housing Rents 22%
- Direct Debit 31%

Source: BBVA Research & Buda et Al (2022).National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. Forthcoming
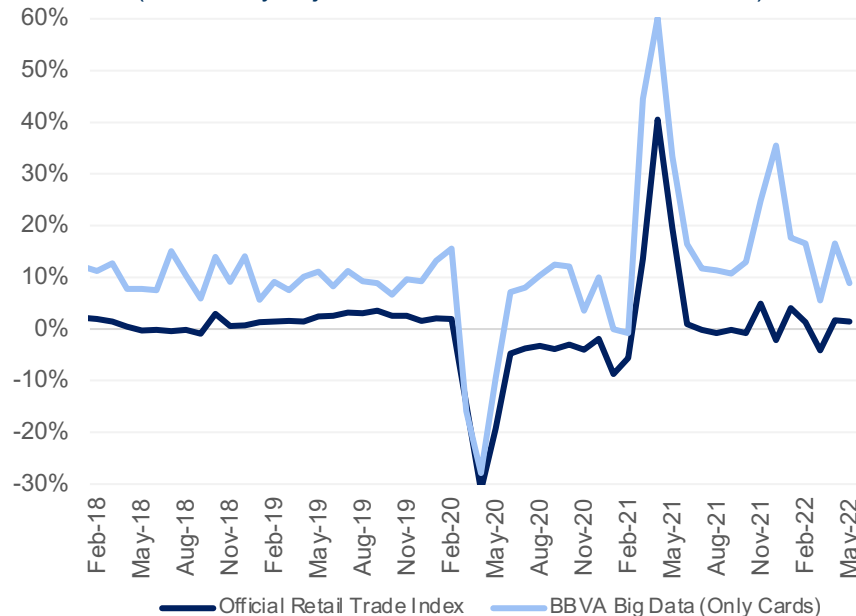
... And this bias is transmitted to the fit with official Growth rates

Spain: Real Private Household Consumption
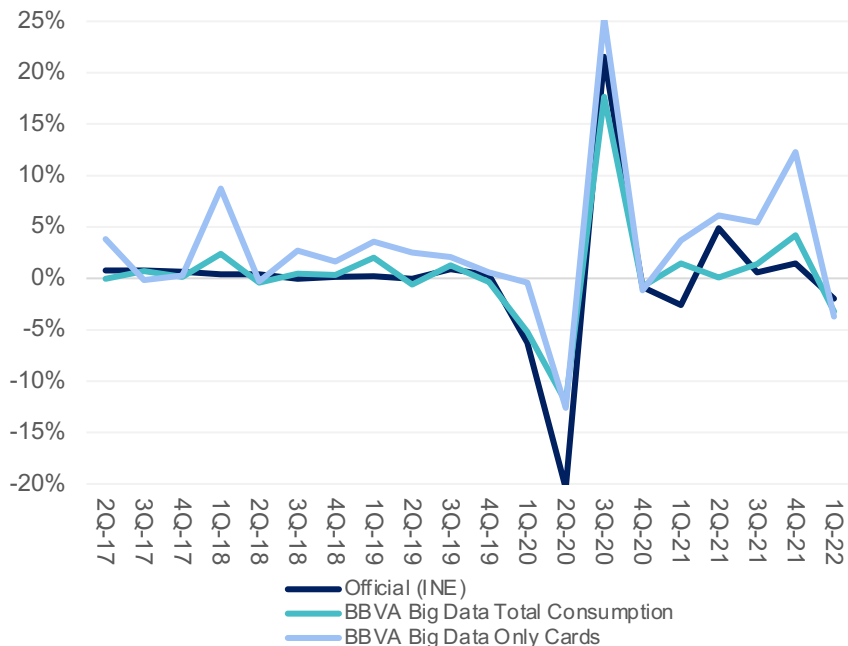(Seasonally Adjusted QoQ Growth Rate CPI deflated)

— Official (INE)   — BBVA Big Data Only Cards

Spain: Real Retail Trade Index
(Seasonally Adjusted YoY Growth Rate CPI Deflated)

— Official Retail Trade Index   — BBVA Big Data (Only Cards)

Source: BBVA Research & Buda et Al (2022).National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. Forthcoming
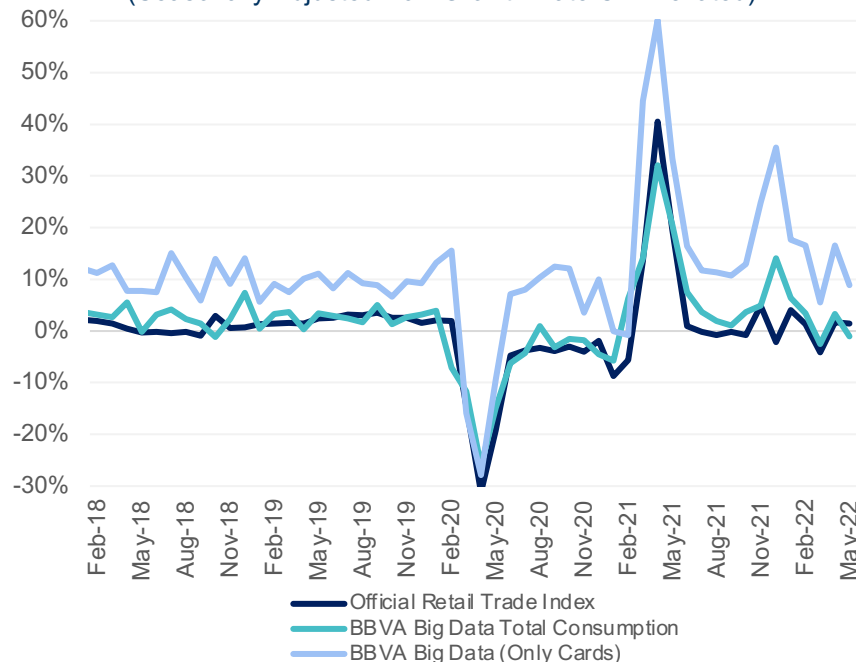
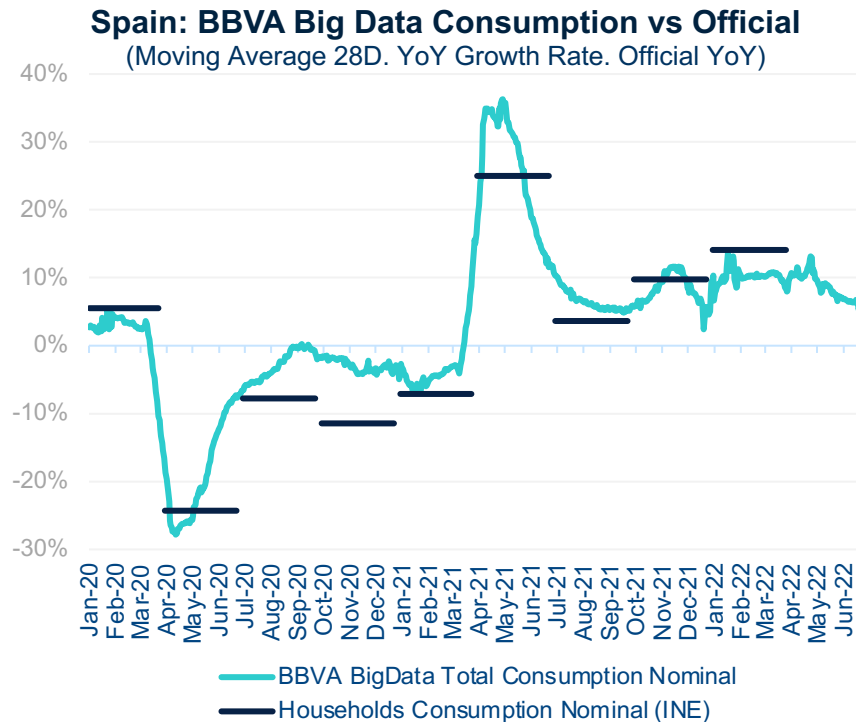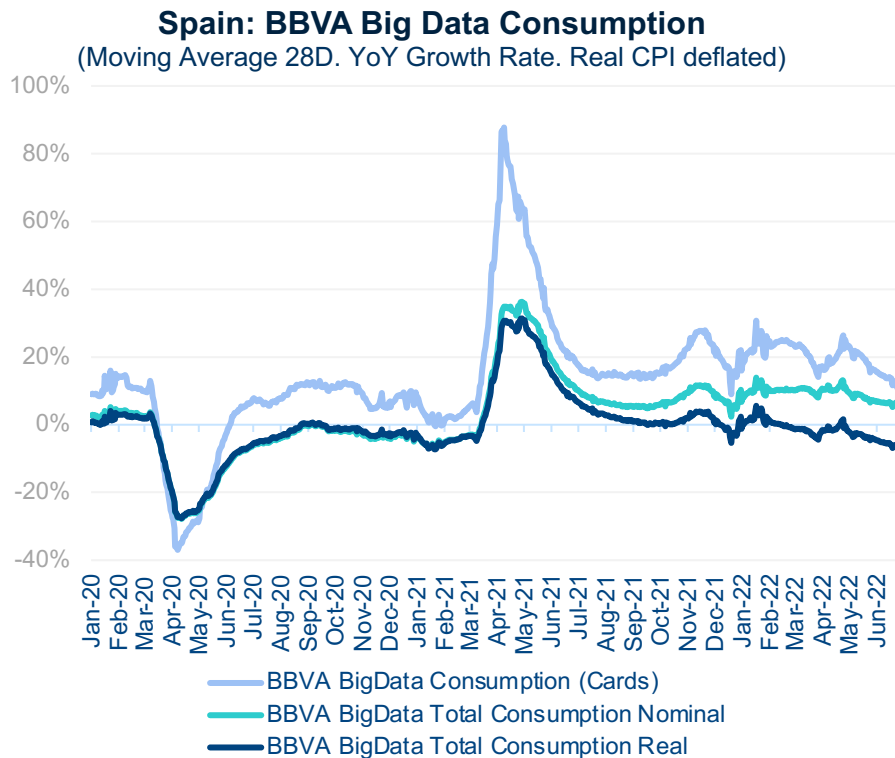# But fortunately … we can correct with Naturally Ocurring Data

Spain: Real Private Household Consumption
(Seasonally Adjusted QoQ Growth Rate CPI deflated)

Spain: Real Retail Trade Index
(Seasonally Adjusted YoY Growth Rate CPI Deflated)

Official (INE)
BBVA Big Data Total Consumption
BBVA Big Data Only Cards

Official Retail Trade Index
BBVA Big Data Total Consumption
BBVA Big Data (Only Cards)

Source: BBVA Research & Buda et Al (2022).National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. Forthcoming

... not only in monthly and quarterly but also in Real Time

Spain: BBVA Big Data Consumption
(Moving Average 28D. YoY Growth Rate. Real CPI deflated)

- BBVA BigData Consumption (Cards)
- BBVA BigData Total Consumption Nominal
- BBVA BigData Total Consumption Real

Spain: BBVA Big Data Consumption vs Official
(Moving Average 28D. YoY Growth Rate. Official YoY)

- BBVA BigData Total Consumption Nominal
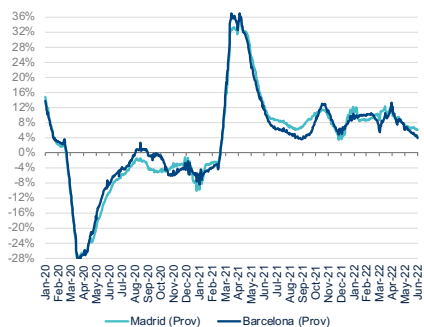- Households Consumption Nominal (INE)

Source: BBVA Research & Buda et Al (2022).National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. Forthcoming

At Geographical Level: National, Regional, Provincial, Urban and Zip Code...
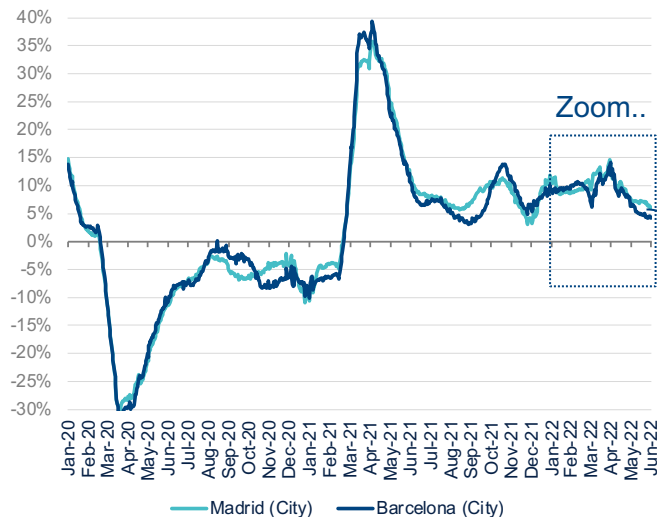
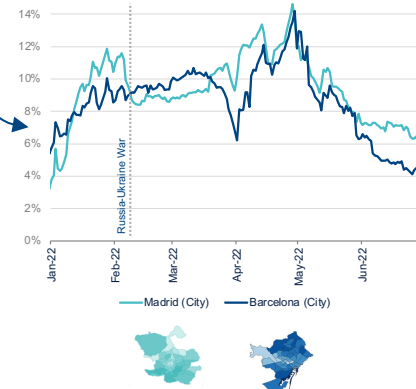**National & regional**
(Moving Average 28D. YoY. Real)
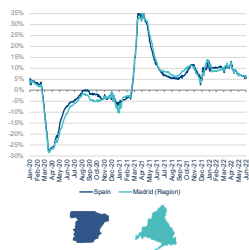
Provincial: Barcelona vs Madrid
(Moving Average 28D. YoY)

Urban: Barcelona & Madrid
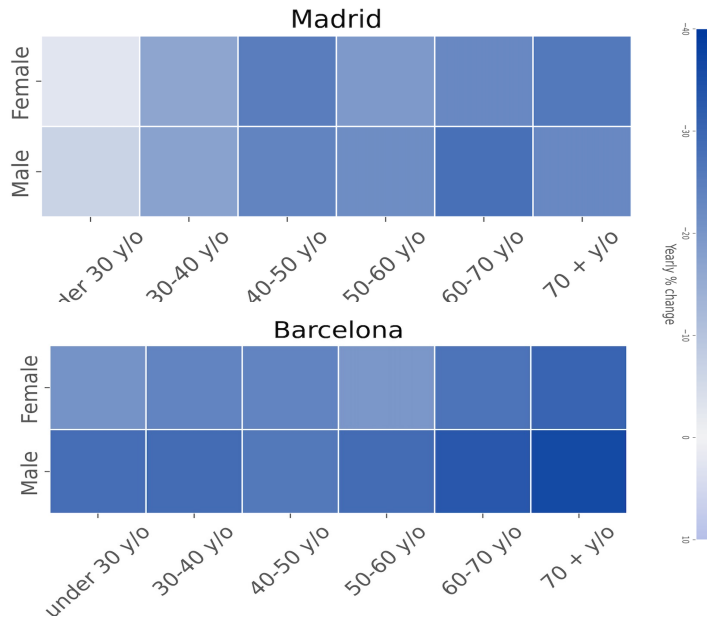(Moving Average 28D. YoY)

Zoom..

Urban: Barcelona & Madrid 2022
(Moving Average 28D. YoY)

… and "High Definition" (i.e zip code level by Activity, Age, Gender …)
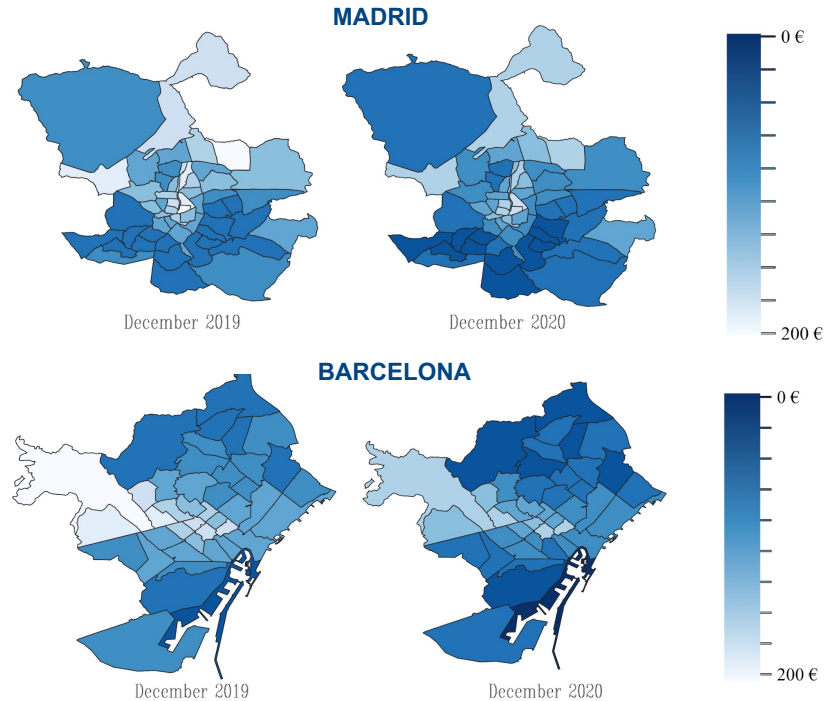

Change in Consumption Restaurants & Hotels in 2020
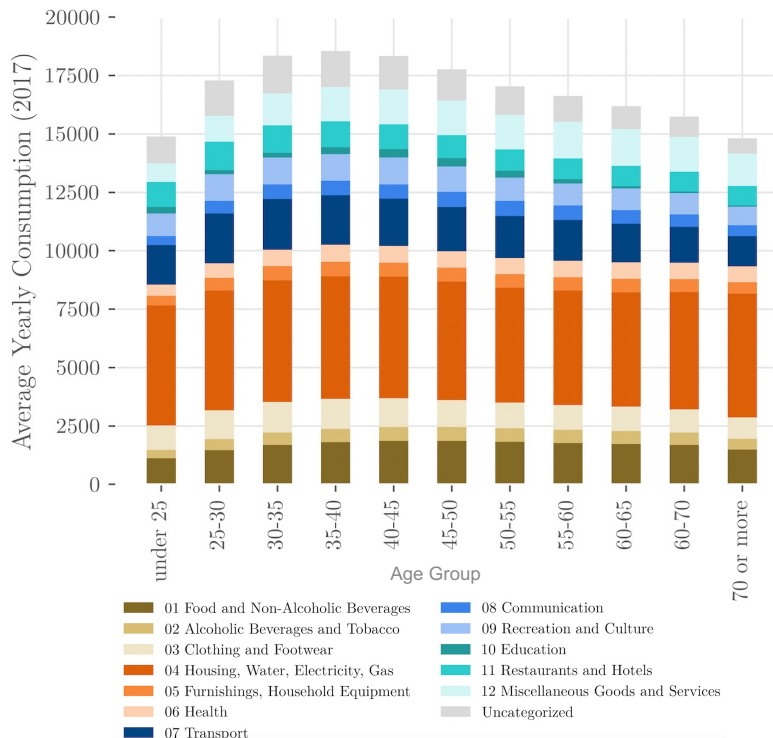
(December 2020 vs December 2019, % YoY)

Average Consumption pp in Restaurants & Hotels in 2020
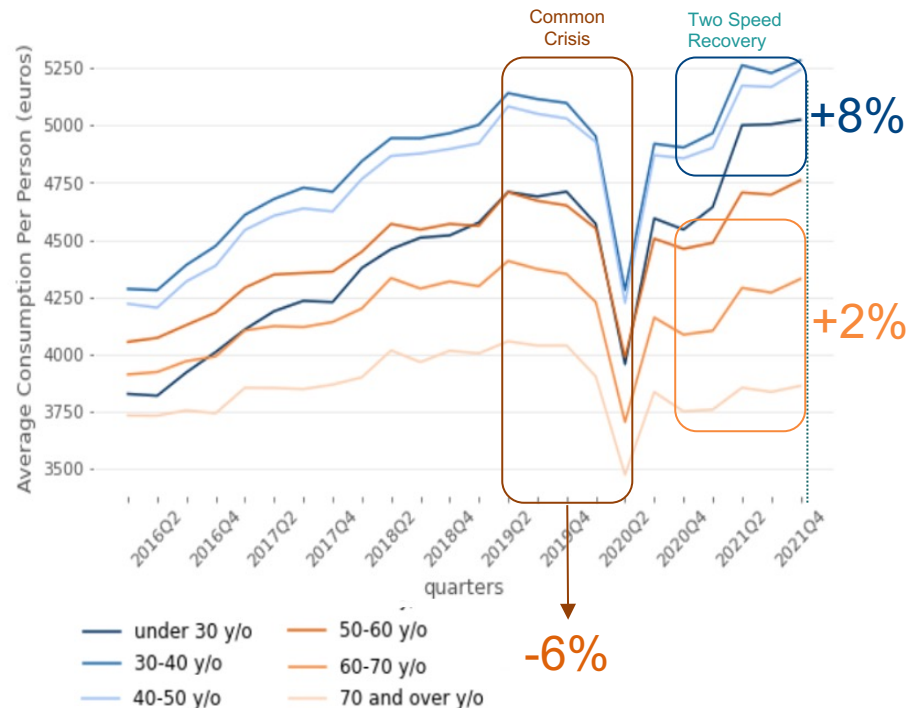
(December 2020 & December 2019, % YoY)

Source: Buda et Al (2022).National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. Forthcoming

...Naturally Ocurring Data are also relevant for the design of SmartPolicies

Spain: Proportion of Consumption per Age Groups (2017

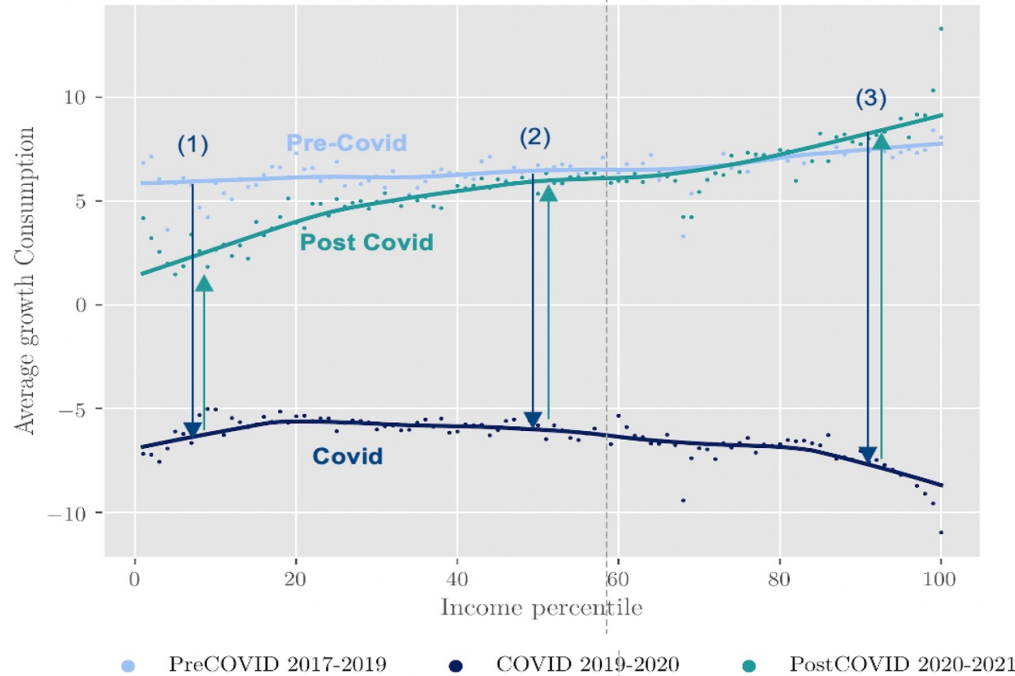**Spain: Levels of Consumption per person by Age Groups**

Source: Buda et Al (2022).National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. Forthcoming

... as High Granularity can give us a "detailed" vision of reality

Spain: Consumption growth by Income Percentile

…Just a partial recovery for the Lower income people

…A sharp adjustment in Consumption Growth for "High Income" and the lower ones …

…Remember they are growth rates.. so Consumption Level recovery Incomplete at the end of 2021

Conclusions

Conclusions

- The vast amount of data naturally occurring within financial institutions can be harnessed to produce high quality consumption survey.

- Unlike standard consumption surveys, a simple aggregation of the survey's data results in National Accounts levels

- But with arbitrary frequency, and incredibly more dense coverage.

- Not only the survey micro data generates distributional accounts for consumption... it allows an individual panel structure that allows for careful study of consumption dynamics

- Of course, it allows for using covariates (income, ...) to understand the determination of consumption at micro and macro levels.

> We aim to use similar procedures to determine Investment, I-O tables, external sector, internal flows, etc...

National Accounts in a World of Naturally Occurring Data:
A Proof of Concept for Consumption

Gergely Buda (BSE)
Vasco M. Carvalho (University of Cambridge & CEPR)
Stephen Hansen (Imperial College & CEPR)
Alvaro Ortiz (BBVA Research)
Tomasa Rodrigo (BBVA Research)
José V. Rodríguez Mora (University of Edinburgh & CEPR)