National Accounts in a World of Naturally Occurring Data

A Proof of Concept for Consumption

G. Buda (BSE), V. Carvalho (Cambridge), S. Hansen (UCL)
A. Ortiz (BBVA Research), T. Rodrigo (BBVA Research) & J. Rodriguez (CUNEF)

Conference on Real-Time Data Analysis, Methods & Applications (Banco de España)
Table of contents

1. Introduction

2. Construction Data: Building a Massive Consumption Survey

3. Aggregate Results

4. Use Cases
Introduction
Relevance of Consumption and Problems with Surveys

- **Consumption is Relevant:**
  - Largest GDP component
  - Subject to extensive study
  (Heterogeneity at Heart of HANK Models)

- **Surveys suffer well-known problems:**
  - Don’t aggregate to national accounts consumption
  - Under-reporting not constant across income
  - Limited panel coverage
  - Low frequency
  - Declining response rates

<table>
<thead>
<tr>
<th>Survey Data</th>
<th>CPS(^a)</th>
<th>CPI Housing(^b)</th>
<th>CE Interview(^c)</th>
<th>MEPS HC(^d)</th>
<th>ACS-Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2012</td>
<td>90.4</td>
<td>66.2</td>
<td>71.3</td>
<td>61.3 (overall)</td>
<td>97.3 (weighted)</td>
</tr>
<tr>
<td>Jan 2014</td>
<td>89.5</td>
<td>70.8</td>
<td>67.0</td>
<td>52.8</td>
<td>96.7</td>
</tr>
<tr>
<td>Jan 2016</td>
<td>86.7</td>
<td>68.2</td>
<td>63.7</td>
<td>51.0</td>
<td>94.7</td>
</tr>
<tr>
<td>Jan 2018</td>
<td>84.6</td>
<td>65.4</td>
<td>58.6</td>
<td>46.8</td>
<td>92.0</td>
</tr>
<tr>
<td>Jan 2019</td>
<td>83.1</td>
<td>63.3</td>
<td>57.6</td>
<td>46.0</td>
<td>86.0</td>
</tr>
<tr>
<td>Jan 2020</td>
<td>81.7</td>
<td>63.9</td>
<td>53.2</td>
<td>NA</td>
<td>71.2</td>
</tr>
<tr>
<td>Jan 2021</td>
<td>78.2</td>
<td>52.4</td>
<td>43.7</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Jan 2022</td>
<td>73.3</td>
<td>52.6</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**SOURCE:** Response rates were found on the websites of the U.S. Bureau of Labor Statistics (for CPS, CPI Housing, and CE Interview columns, see https://www.bls.gov/osmr/response-rates/household-survey-response-rates.htm), the Agency for Healthcare Research and Quality (for MEPS HC column, see https://meps.ahrq.gov/mepsweb/survey_comp/hc_response_rate.jsp), and U.S. Census Bureau (for ACS-Annual column, see https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/response-rates/).


\(^d\)Medical Expenditure Panel Survey, Household Component, Agency for Healthcare Research and Quality.
This Paper

- We apply **National Accounting principles (ESA-2010)** to Transactions from major private sector bank (BBVA)

- Results in a with **Consumption Panel** characterized by:
  - A Massive Survey: Almost two million participants
  - Aggregates to national accounts consumption at comparable frequencies
  - Updated Real time (daily frequency since 2015) & Geo-localized information
  - Breakdown of consumption into COICOP categories (officially yearly, now real time!!)

- With **Many Applications**
  - Consumption in Real Time & HD (1st Distributional Accounts of Consumption)
  - Rich characterization of consumption growth
  - Inflation Heterogeneous impact, High-frequency response to monetary policy shocks & Sustainability other papers
Construction Data: Building a Massive Consumption Survey
Transactions vs Household Budget Survey: Pros and Cons

**Advantages** of Transactions vs Official Survey (HBS):

- High Frequency
- Observed rather than reported spending
- Larger data both in cross section and time series

**Disadvantages** of Transactions vs Official Survey (HBS):

- Individual $\neq$ Households
- Spending $\neq$ Consumption
- Bank Clients $\neq$ Population
The Data Sample: Actual and Progress

- **Sample:** From 2015-Q1 to present
- 1.8 millions of "**Active Costumers**" (3bn individual transactions & 200 Bn euros).
  - In the sample latest 3 Years (*less restrictive & Incorporating Youngers*)
  - More than 10 transactions per quarter
  - Excluding clients with big Transfers among Banks (individual or Households).
  - Active Costumer drop to 1.8 million from more than 10 million clients (*Likely Increase*)

- Considering BBVA clients linked in a "**Household Units**" whenever:
  - They have co-signed a financial contract (bank account, loan, mortgage...)
  - They reside in the same postal code

- Households include both active and non-active clients
From Individuals to Households

Households Proxy vs Official Data

To form the official distribution of household sizes, we use INE’s Continuous Household Survey (https://www.ine.es/dyna/INEbase/en/operacion.htm?c=Dataest&c_id=126433617696&mem=results&did=1264336172981) and extract from each surveyed household the number of adults. We focus on adults since minors are unlikely to be BBVA customers.
Consumption ≠ Spending

- Not all Individual-to-Firm Transactions are Consumption

- We first limit attention to Individual-to-firms/organizations with tax ID and try to assign transactions to 12 official COICOP Categories

- We use ESA-2010 principles to design appropriate filters:
  - Expenditure by Non-Residents ≠ Consumption (Export Services)
  - Real Estate purchases & House repairs ≠ Consumption (Investment)
  - Direct Tax payments ≠ Consumption (Government Revenue)
Structure of Payments Data: Non Housing Expenditures

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

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

**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.

**Cash**
- Both cash and over the counter.
- Assume is consumption.
- Assumptions on distribution.
Estimating Housing Expenditures

- A major but largely imputed component
- We build household rental payments using search of text description
  - Minimum 100 EUR & 70 months
  - 32127 Households
- We regression on:
  - Income (proxied by BBVA, 6 month average)
  - Utility Payments (Direct Debits)
  - Geography: 327 Regions (postal codes)
- Good fit in-sample & out-of-sample.
- Imputed to Whole

### Estimating Rental Payments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spending on House Utilities</td>
<td>0.0884</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0.0362</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td></td>
</tr>
<tr>
<td>N of Contract Groups</td>
<td>16,977</td>
<td>15,512</td>
</tr>
<tr>
<td>N of Observations</td>
<td>1,134,735</td>
<td>15,512</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.3911</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.3765</td>
<td></td>
</tr>
<tr>
<td>Within $R^2$</td>
<td>0.1200</td>
<td></td>
</tr>
<tr>
<td>Root MSE</td>
<td>204.6144</td>
<td>221.64</td>
</tr>
</tbody>
</table>
### Consumption Categories (COICOP): Transactions (Daily) vs Official (Yearly)

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Food and Non-Alcoholic Beverages</td>
</tr>
<tr>
<td>02</td>
<td>Alcoholic Beverages, Tobacco, and Narcotics</td>
</tr>
<tr>
<td>03</td>
<td>Clothing and Footwear</td>
</tr>
<tr>
<td>04</td>
<td>Housing, Water, Electricity, Gas, and Other Fuels</td>
</tr>
<tr>
<td>05</td>
<td>Furnishings, Household Equipment, and Routine Household Maintenance</td>
</tr>
<tr>
<td>06</td>
<td>Health</td>
</tr>
<tr>
<td>07</td>
<td>Transport</td>
</tr>
<tr>
<td>08</td>
<td>Communication</td>
</tr>
<tr>
<td>09</td>
<td>Recreation and Culture</td>
</tr>
<tr>
<td>10</td>
<td>Education</td>
</tr>
<tr>
<td>11</td>
<td>Restaurants and Hotels</td>
</tr>
<tr>
<td>12</td>
<td>Miscellaneous Goods and Services</td>
</tr>
</tbody>
</table>

**Table 1: COICOP Consumption Categories (Two-Digit)**
Weighting: From BBVA to National

Household Weighting

Demographic Weighting

\[ c_i = \frac{\sum_{j \in A(i)} c_j^{NH} + c_h(i)}{A(i) + 0.5O(i)} \]

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

**A(i):** “Active” customers in i’s household

**O(i):** “Inactive” customers in i’s household

**(g, a, q):** gender, age, within-region neighborhood income quintile cell

\( x^{INE} \): count of Spaniards in cell in census

\( x^{BBVA} \): count of Spaniards in cell in BBVA

(Under Revision)
Consumption by Means of Payments is not Homogeneous:

- Cash: Distribution by assumptions
- Off-Line Cards: Balanced
- On-line Cards: Third point
- Direct Debit: Utilities, Insurance
- Transfers: Durables Spending (Cars, White Goods..)
Aggregate Results
Aggregate Naturally Occurring Consumption vs. National Accounts

Levels

Growth (QoQ)

These figures compare quarterly aggregate household consumption according to official INE data and to naturally occurring data. To seasonally adjust both series, we use the Jdemetra+ application and apply X-13ARIMA-SEATS. The plot on the left shows the total level of consumption. The plot on the right displays the growth rate in aggregate consumption from quarter $t-1$ to quarter $t$. 
Consumption Categories (COICOP): Naturally Occurring vs Official

![Graph illustrating consumption categories comparison between Naturally Occurring Data and Official INE HBS.](image)

- **01 Food and Non-Alcoholic Beverages**
- **02 Alcoholic Beverages and Tobacco**
- **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**

Legend:
- △ INE HBS
- ○ Naturally Occurring Data

**Notes:**
- National Accounts in a World of Naturally Occurring Data: A proof of Concept for Consumption
Total Consumption: Not Only Cards

Spain: Real Household Consumption vs BBVA BigData
(Sadj %-QoQ Growth Rate CPI deflated)

Spain: Retail Sales Index vs BBVA Big Data
(Sadj mom Growth Rate. Big Data CPI deflated)

Source: Buda et Al (2022) BBVA Research & INE
Total Consumption: Not Only Cards

Spain: Real Household Consumption vs BBVA Big Data
(Sadj %QoQ Growth Rate CPI deflated)

Spain: Retail Sales Index vs BBVA Big Data
(Sadj %mom Growth Rate. Big Data CPI deflated)

Source: Buda et Al (2022) BBVA Research & INE
Total Consumption: The High Definition Component

Urban Big Data Consumption: Barcelona & Madrid
(Moving Average 28D. YoY Nominal)

Consumption in Restaurants & Hotels in 2020
(Avg Consumption per person December 2020 & December 2019, %)

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

Source: Buda et Al (2022) BBVA Research
Aggregate Results: Take Home

National accounts are vital economic statistics but face pressure:

- Declining survey participation
- Budget cuts for national statistics agencies
- Pressure for more timely and granular data

In many countries they are sparse to non-existent (Silungwe et. al. 22):

- 1/3 of countries do not publish quarterly accounts (1/2 in Africa)
- Only 4 European and 5 Asian countries produce quarterly acc. within 30 days
- 1/4 of countries have no Household Budget Survey (1/2 in Africa)

Proof of Concept: Widely available financial transaction data is competitive as a national accounts aggregate consumption measure.
Use Cases
Consumption Inequality

**Distribution of Consumption BBVA vs. Spanish Household Budget Survey**
(yearly consumption per adult distribution 2017)

- INE Household Budget Survey (Microdata)
- Naturally Occurring Data

**Consumption distribution disaggregated by COICOP**
(consumption categories by percentiles adult distribution 2017)
Consumption Distributional Accounts

Consumption distribution by COICOP Categories (Share)
(consumption categories by percentiles adult distribution 2017)

Consumption distribution by Category & Age
(consumption categories by age cohorts 2017)

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
Uncategorized
Sustainability: Inequality on Households Carbon Footprint

Distribution of CO2 Emissions by Spanish Households
(as % of total expenditure 2017)

There is a Relevant Margin for Heterogeneous Policies

Some Policies can have higher costs for low income families

Distribution of CO2 Emissions by Age
(as % of total expenditure 2017)

National Accounts in a World of Naturally Occurring Data: A proof od Concept for Consumption
Consumption Growth Analysis

- Distribution of Consumption Growth does not look Gaussian not Homogeneous
  - Thick Tails
  - Albeit less than income
  - Very Non-Linear & Lumpy process
  - Very strong mean reversion

- Potential Causes
  - Income Process
  - Lumpiness of purchases
Heterogeneous Effects of Inflation Shock

Effect of “Inflation Shock” on Household by Income & Age
(Unexpected Inflation Shock by Age & Percentiles of Income, in % of Income)

- Relative Consumption Channel
- Fisher Channel
- Nominal Income Channel
- Total Impact (% Income)

Source: Forthcoming Cardoso et al (2023)

Indebted Mid-Age Balanced the impact (Fisher Effect)

Older & High Income Cohorts relative higher impact (No Debt)
Monetary Policy: Short & Variable Lags + Heterogeneous response

Local Projections Model Responses to Monetary Policy Shock

Aggregate Consumption

Zoom-in on first 31 days

Days after shock

Days after shock

Aggregate sales

Zoom-in on first 31 days

Days after shock

Days after shock

Aggregate Employment

Zoom-in on first 31 days

Days after shock

Days after shock

Local Projections Model Responses of Consumption to Monetary Policy Shock

Clothing and footwear

Furnishings, equip. and maint.

Transport

Recreation and culture

Education

Restaurants and hotels

Alcoholic bev., tab. and retail.

Health

Communication

Food and non-alcoholic bev.

Housing services and utilities

National Accounts in a World of Naturally Occurring Data: A proof od Concept for Consumption
Spain: Big Data Total Consumption BBVA & Tax Agency Total Real Sales
(SA. Adjusted, Cumulative 91 days % Quarterly Growth Rate. Big Data Consumption CPI deflated)

Five Year Sample 2018-2023

One Year Sample: October 2022- October 2023

Source: Own source and AEAT
References


- Buda, Gergely, Stephen Hansen, T Rodrigo, VM Carvalho, Á Ortiz, and JVR Mora. (2022). “The Spanish Households’ Carbon Footprint Inequality in High Definition & Real Time”. BBVA Research Economic Watch


Thanks...