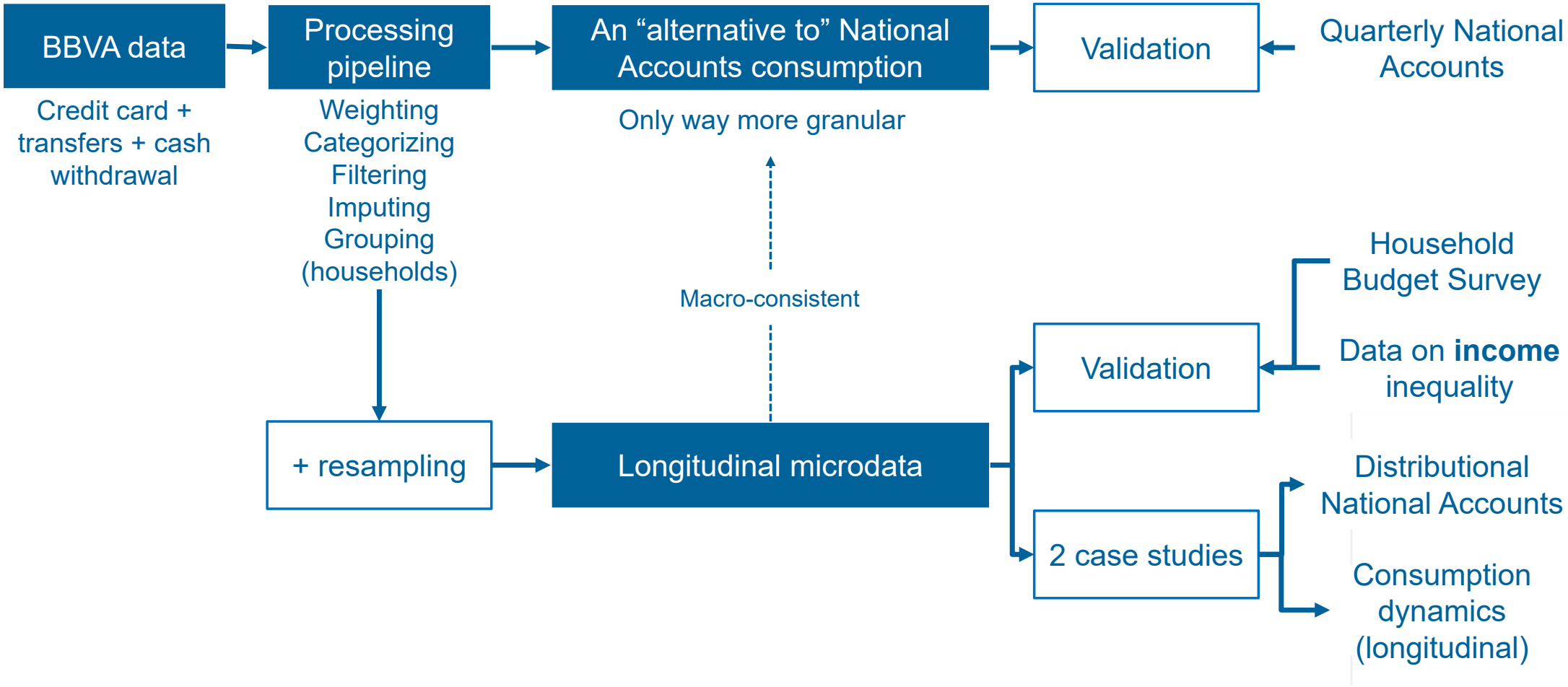
NATIONAL ACCOUNTS IN A WORLD OF NATURALLY OCCURRING DATA: A PROOF OF CONCEPT FOR CONSUMPTION

Discussion by Nicolas Woloszko

Advanced analytics: new methods and applications for macroeconomic policy
2022, 22 July, Bank of England, London.

OECD
BETTER POLICIES FOR BETTER LIVES

Summary

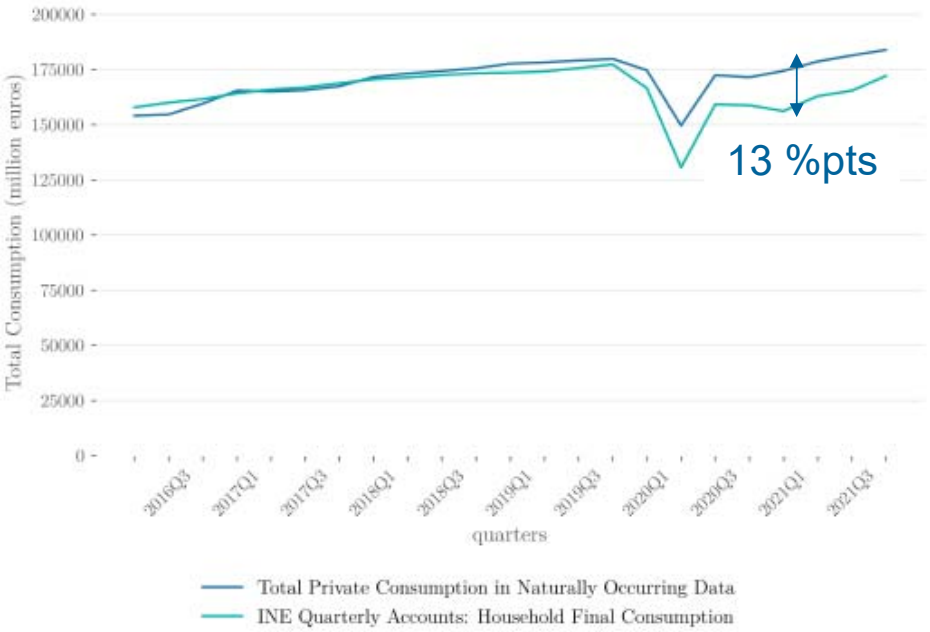

BBVA data
Credit card + transfers + cash withdrawal
Processing pipeline
Weighting
Categorizing
Filtering
Imputing
Grouping
(households)
An “alternative to” National Accounts consumption
Only way more granular
Validation
Quarterly National Accounts
Macro-consistent
+ resampling
Longitudinal microdata
Validation
Household Budget Survey
Data on **income** inequality
2 case studies
Distributional National Accounts
Consumption dynamics (longitudinal)

Evidence fall short of the ambitious claim

The claims are very strong: **“the promise of reshaping economic measurement”**, **“an alternative”** to official data.

But the evidence leaves room for additional questions:

Figure 5



(1) Macro data: mixed quantitative evidence

- **Goodness of fit decays after 2020 (13%pts)**
- The paper does not provide any comment
- Occurrence of **concept drift** ?

(2) Micro data: only qualitative evidence

- Comparisons with income inequality: a very broad stylised fact
- Paper claims to “beat” the other benchmark (HBS) - probably rightly so

We lack unambiguous quantitative evidence

Overfitting exist even outside machine learning

- Validation is performed over few data points, and the processing pipeline has many steps and many hyperparameters: risk of overfitting the validation set.

  We do not know how many times you tried before you succeeded…

- Additional evidence could include:
  - Could the country coverage be broadened ?
  - Validation of intermediate steps:
    - Already performed on household size distribution, could there be other opportunities ?
  - There are issues with HBS levels and top shares
    - Validate the structure of expenses per income decile: should match at least in the middle + bottom of the distribution
    - Two-step approach: (1) adjust the HBS survey data for undersampling using (Ruiz & Woloszko 2016), and (2) make the comparison

Forcing the sale ?

The language is unusually hyperbolic and laudatory for a research paper (more of an internal note ?)

- The title phrase "**Naturally occurring data**" is misleading:
  - BBVA data does not occur naturally, it results from business operations which serve a specific purpose
  - They are processed with an ad hoc pipeline
- Comment on Figure 5:

> We obtain striking results. In spite of the distinct methodologies employed to obtain aggregate consumption measures, naturally occurring and official data line up exceedingly well in both levels and growth rates at quarterly frequency. This is a strong validation of our approach and suggests that even

- Does the validation of the macro aggregates imply the validity of the micro data ?

> National Statistical Institute. Importantly, this result immediately implies that our underlying micro-data can be additionally deployed to build distributional national accounts for consumption, character-

Still, the product is super rich and offers many possible applications

- Real-time + distributional CPI weights
  - Real time data on prices can be scrapped (R. Rigobon & A. Cavallo)
  - Weights come with 5-year release delay from official sources !
  - Quite important role (R. Chetty et al.) especially around large shocks

- Impact of carbon tax (based on longitudinal data)
  - How does household expense structure reacts to relative price changes ?
  - Track substitution dynamics resulting from (carbon) taxes

An “alternative” to National Accounts ?

I see two problems:

- Sampling biases (possibly addressed, but Figure 5 puts emphasis on risk of concept drift)
- Possible conflicts of interests

I would be more comfortable if all banks entered data sharing agreements with NSOs (unless there is a monopoly, such as with e.g. Google Trends data)

Should there be **law** to enforce **public-private data sharing**, when private data is thought of as a **public good** ?