

Big Data

Beyond Cards: Measuring Spanish Consumption in Real Time and High Definition

Gergely Buda, Vasco Carvalho, Stephen Hansen, Alvaro Ortiz, Tomasa Rodrigo, and Sevi Rodriguez Mora

Summary

- We provide the first proof of concept that naturally occurring data from millions of financial transactions can be harnessed to estimate national accounts in real time and high definition.
- We show how to build up the Consumption National accounts from different means of payments in both levels and growth.
- We also show how relying only on Card transactions can lead to misleading results of consumption levels and growth given the upward bias of cards especially during post-covid recovery.
- We design the first ever distributional account system of consumption consistent with national accounts. We show how to understand Inequality in Consumption from different aspects: levels, growth, categories of consumption, age, and gender.
- We develop a set of consumption indicators in real time and high definition at national, regional, provincial and urban levels as well as at different categories as defined in national accounts. In addition, we can determine in detail the distribution of Co2 emissions in the consumption basket of Spanish households.
- Finally, we show how this information in real time and high definition constitutes a powerful tool for analysts and policymakers. Information in high definition constitutes a good base for better diagnoses and, furthermore, opens the door for the implementation of smart policies.

1.National Accounts in a World of Naturally Occurring Data

Since Covid-19 hit the world economy at the beginning of 2020, the reliance on high-frequency indicators have become more frequent. Our first paper Carvalho et al. (2021) was part of a fast-growing literature including those leveraging access to credit/debit cards.¹ In our new paper (Buda et al,2022) *National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption*², we move a step forward to complete the full National Accounts Household Consumption map by including the rest of means of payments.

Our data covers the universe of BBVA retail accounts in Spain and yield an unprecedented granular ledger, allowing us to track expenditure as it flows out of these accounts, transaction by transaction, for a total of 3 billion individual transactions by 1.8 million BBVA customers, from 2016 to 2021. Our paper makes four contributions:

¹ Our first paper Carvalho et al. (2021) was part of a fast-growing literature including those leveraging from access to credit/debit card see Gelman et al. (2014), Baker (2018), Aladangady et al. (2021), and Olafsson and Pagel (2018). Given the increasing availability of such data and in face of societal demands for high frequency, granular tracking of the economy during the COVID-19 pandemic, this literature expanded rapidly over the past two years; see, for some early contributions, Carvalho et al. (2021), Andersen et al. (2020), or Chetty et al. (2020), and Baker (2018) and Vavra (2021), for recent reviews taking stock of this literature. Some examples of tracking credit cards in Spain are also

² This brief is based in our recent paper Buda, G., Carvalho, V. M., Hansen, S., Mora, J. V. R., Ortiz, Á., Rodrigo, T. (2022) National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. <u>Cambridge Working Papers in Economics CWPE2244</u>.



- First, we construct a representative panel of household expenditure (Massive Survey) by adding to the debit and credit card transactions of the BBVA households (both online and offline) the rest of the payments. Thus, and beyond cards, we include all direct recurrent debits, all one-off transfers, and individual payments, as well as all cash withdrawals and the estimation of imputed rents. We process transactions for each account holder and each means of payment to: (i) categorize transactions across harmonized consumption spending categories, (ii) filter out nonconsumption expenditures (such as transfers to saving accounts, household-to-household transfers or tax payments), (iii) impute the consumption of housing services for all households, by exploiting information on actual rental, housing utilities, location and income for a subsample of BBVA households, and finally, (iv) construct a large sampling frame of households that is representative along demographic observables in particular, gender, age, and region so as to mimic the characteristics of the Spanish adult population to avoid biases of the BBVA Clients Database.
- Second, we construct from the bottom-up a series for quarterly aggregate final consumption expenditures of domestic households. We cross validate our accounts and compare it against that in the Spanish Quarterly National Accounts compiled by Spain's National Statistics Institute (INE).³ Despite methodological differences, we show that our naturally occurring aggregate consumption matches the official INE series remarkably well, both in levels and in growth rates.
- Third, we create the first distributional national accounts for consumption. Following the pioneering work of Piketty et al. (2018) in developing distributional national accounts for income, we have created, to the best of our knowledge, the first distributional national accounts for consumption. To show the value of these accounts, we present a micro-study of consumption dynamics.
- Last, but not least, we show some examples of how we can use the Big data information including in our accounts to track the economy in real time and high definition and how this information can be useful for the design of "Smart Policies" (i.e. Addressing policies there were more needed and/or maximizing the impact of macroeconomic policies).

2.Beyond Cards: Completing the map of total consumption

Since the Covid outbreak, we have been tracking Consumption in real time and high definition with the Consumption Big Data Indicators coming from Credit & Debit Cards transactions registered either in Physical point of sales (PoS) or online ones based on the methodology described in Carvalho et al. (2021). Although this proved to be very useful, it is incomplete. The information from Credit/Debit Cards supposes nearly 35% of Consumption, which makes this indicator to monitor consumption somehow incomplete. With our new methodology (Buda et al., 2022), we also include information for the rest of the means of payments used for consumption purposes which allow us to cover the 100% of Consumption in our Total Consumption index (Total). In addition to a higher level of consumption, the new indicators allow us to deal with different dynamics of the means of payment. The key differences are the following:

- **Card consumption (online):** This means of payment is undoubtedly the most dynamic and the most resilient to the Covid crisis. Although its weight in total consumption is still not high (9%), it presents a steady upward trend. The On-line payments are experiencing a secular transformation and we expect this positive trend to be alive.
- Card Consumption (Physical): Its share in total consumption is higher (26%) and, although less dynamic than the On-Line payments, its growth rate is higher than the rest. This new card consumption series has been also adapted to mimic household consumption in national accounts, as the one previously used included card

³These two series follow different methodologies. Our series aggregates directly from a nationally representative large-scale real time expenditure survey, as described above. Instead, official quarterly national accounts consumption, are largely based on quarterly firm sales survey data, with subsequent imputations regarding who is consuming (e.g. Spanish nationals vs. foreigners, households vs. firms) and what is consumed (e.g. investment or intermediate goods vs. consumption by households)



transactions which do not really correspond to Household consumption (Tax Payments, Securities, investments, etc...) and the consumption transactions in Spain from cards issued abroad (which should be considered as exports of services).

- **Direct debits:** These represent 20% of consumption and were not included in our first set of indicators. It includes transactions normally used for the payment of household supplies or utilities such as water, electricity, gas, and communications such as telephone bills. This consumption has been more resilient to the crisis and the post-Covid recovery has been less intense.
- **Rental payments:** This is an important component of consumption (near 25%) and more stable. To prevent and comply with the privacy of clients, the rental payments for households have been estimated using co-titulars of bank accounts, assuming that they form a household.
- **Money Transfers:** This is the payment with the smallest weight (5% of the total) and its use is linked to the purchase of durable goods (vehicles and white goods). Its response to the COVID crisis was huge, falling near 60%, but was followed by a positive recovery in this spending (similar to that of physical cards one).
- **Cash:** It still represents 15% of consumption and has been the biggest loser of the crisis. The share of cash in consumption has lost 10 pp over the period 2017 to 2021. Covid produced a strong adjustment that has had permanent effects.



One important feature of **the New Total Consumption Index is that consumption growth is more stable, less dynamic, and closer to the official household consumption**. The new Big Data Consumer indicator fits relatively well not only the level, but also the Growth of the Official Household consumption. Figure 3 and Figure 4 show how the Total consumption indicator (Turquoise) fits the official indicators as the National Quarterly Consumption and the higher frequency retail trade index much better than the information coming exclusively from credit and debit cards (Light Blue) The main reason for this higher accuracy lies in the ability of the new indicator to capture the dynamics of all the means of payments.





3. The Distributional Accounts of Consumption

Following the pioneering work of Piketty et al. (2018)⁴, there are already distributional national accounts for income for many countries. Combining existing national accounts aggregates, censuses, household surveys, and micro income tax data, this macroconsistent accounting methodology has arguably had a large impact in both academic and public discussions surrounding income inequality and its time evolution. Yet, to the best of our knowledge, distributional national accounts for consumption are nonexistent. Furthermore, **we present the first distributional accounting of consumption both in consumption levels and in growth**. Some key results of our analysis are as follows:

- First, we benchmark our analysis against consumption inequality as given by the Spanish Household Budget Survey. The comparison of the tails of the consumption distribution across the two datasets (Figure 5) again suggests that, in line with the literature on consumption surveys, **the Spanish Household Budget Survey is undersampling the top of the consumption distribution.**
- Second, we benchmark our results against existing distributional accounts for post-tax income in Spain, concluding that **macroconsistent consumption inequality is substantially smaller than its income counterpart** (where for example, 31% of total national post-tax income accrues to the top 10%) as shown in Figure 6.
- Finally, we show that it is also possible to break down this distributional analysis of aggregate consumption further, across consumption categories, demographics (age and gender) and time frequencies. We show that it is possible to reproduce analyses typically pursued in the consumption inequality literature, but this time in a way that is both consistent with the level and evolution of aggregate consumption in National Accounts.

⁴ Piketty's pioneering work has generated a growing literature on inequality, mostly focused on income inequality. In Spain an example of this stream using Big Data is Aspachs et al (2022).

P.



Source: Buda et al. (2022) and Household Budget Survey (INH)

Figure 7 is a good example of the **distribution of consumption by percentiles and by age**. The expenditure share on necessities (Food, Clothing, Housing, etc. in orange and brown colors) constitutes 57.4% of total consumption for the median adult in the consumption. Consistent with the concept of necessities, this share declines strongly over the consumption distribution, accounting for 67% of total consumption of adults in the bottom 10%, 49% of total consumption of the top 10%, and only 29% of the top 0.1%. Thus, although the total consumption of necessities increases with total consumption, it is somewhat smaller than that for total consumption. In contrast, the distribution of luxury consumption distribution only accounts for 24% of aggregate luxury spending, while the top 10% accounts for a disproportionately large 30%. As expected, luxury consumption is concentrated at the very top and, for example, accounts for 71% of consumption of the average adult at the very top 0.1% of the consumption.





 01 Food and Non-Alcoholic Beverages
 02 Alcoholic Beverages and Tobacco
 03 Communication

 02 Alcoholic Beverages and Tobacco
 09 Recreation and Culture
 09 Recreation and Culture

 03 Clothing and Footwar
 10 Elonation
 10 Elonation

 04 Housing, Water, Electricity, Gas
 11 Restaurants and Horeb

 05 Furnishings, Household Equipment
 12 Miscellaneous Goods and Uncategorized

 07 Transport
 07 Stransport

Source: INE and Buda et al. (2022)

25-30 35-40 35-40 45-50 56-55 50-65 56-60 60-65 56-60

0.

70 or

and Services

Source: Buda et al. (2022) and WID



Figure 8 shows the **distribution by age group of consumption categories**, where the familiar hump-shaped pattern of consumption is evident in our data. Quantitatively, the average Spanish adult between 35-40 years old consumed around 18500 euros during 2017, 10% more than the average adult consumer in Spain during the same year (and almost a quarter more than the median adult consumer). On the contrary, both those under 25 and those over 70, consume 9% less than the average adult in Spain. In sum, we observe a 20% increase in consumption over the life cycle from young adulthood till middle age, followed by a similarly sized decline in consumption into old age.

The development of Distributional Consumption Accounts can be very useful for different analytical and policy purposes. As these accounts are consistent with National Accounts categories of Consumption we can implement multitude of complementary analysis. We can develop the Environmental extended Distributional Accounts by applying the Co2 intensity coefficients (Kg Co2 by EUR of GDP) to the Household Consumption Categories (COICOP). This opens the door for richer analysis of Household Footprint analysis or decarbonization policy strategies.



Source: Buda et al (2022) applying the Multi Regional Environmental Extended Input Output coefficients of Co2 intensity for Spain.

In figure 9 we show the consumption levels, shares of consumption category by percentiles and age cohorts ordered in terms of Co2 intensity. The Darker colors stand for more intensive categories such as Transport, Food and Beverages, Furnishing and Household equipment and maintenance while the lighter colors reflect the more sustainable consumption categories such as Education, Communications, Health... A simple view reveals some interesting facts. First, the main source of inequality lies in Transport (Coicop 7) whose consumption increases exponentially in the higher percentiles of consumption in both levels and share. This contrasts with the consumption of Housing Utilities such as water, electricity, gas and other fuels. While this category of consumption is very evenly distributed in EUR levels (the typical pattern for first necessity goods) it represents a big share of the consumption basket among the low consumer individuals (normally the low income ones). Additionally, the third graph shows that mid age (30 to 45) concentrates the higher amounts of Co2 intensive consumers due basically, but not only, to transport. All of this rich information opens the door for better diagnosis and the potential to enhance the design of smart policies for Household de-carbonization strategies.



4. Consumption in Real Time, High Definition and the Potential Use to Design for Smart Policies

Although Naturally Occurring Data show their relevance for richer distributional analysis research, it also constitutes a powerful tool for economists for tracking the economy in real time at a very granular level or high definition. Figure 10 confirms our previous results: Our total consumption aggregate shows a good fit to the official national account growth rates (in dark blue horizontal lines for every quarter) in real time. It also confirms our previous result that relying only on credit cards would be misleading. Having a proxy in real time of consumption (55% of the GDP in 2021) provides an important advantage in terms of reaction for analysts and policy makers.

The high-definition component of our work can also be as relevant as the real-time one. Given that the information included in our work is geolocalized, we can show the evolution of real-time consumption at any geographical level, from national to zip code levels, assuring the representativeness of the data. We show an example in Figure 11, where the different patterns of consumption between Madrid and Barcelona can be observed. As shown, the recent deceleration of consumption is somehow more intense in Barcelona.



Source: Ine and Buda et al. (2022)

Source: INE and Buda et al. (2022)

Beyond the real-time geographic patterns, the granular information can be used to stress differences in behaviour at any geographical level in terms of time, age, income level, activity, and districts. The combination of all of this information becomes a powerful and interesting tool for policymakers and urban planners to elaborate detailed diagnostics of the situation, as well as to design smart policies.

Figures 12 and 13 are a good example of this potential. The first shows the impact in terms of consumption in Madrid and Barcelona by age and gender on the Restaurants and Hotels categories for 2020. The blue colors denote the depth of the adjustment, with the darker color reflecting a sharper adjustment and the lighter one reflecting a lighter one. As can be seen, the impact is not entirely homogeneous as older people adjusted their consumption much more than younger people. While the differences by gender are minimal, the geographical differences are clear, reflecting a lower adjustment in Madrid than in Barcelona.



In Figure 13, we go deeper into the geographical dimension of consumption levels. In the case of Madrid, the shift to dark colors in the southern zone is much more evident than in the northern zone (associated with higher income). On the other hand, in Barcelona, it is precisely the north zone where we see a greater adjustment.



Source: INE and Buda et al. (2022)



Source: INE and Buda et al. (2022)

5. Conclusions

Plentiful, naturally occurring transaction data can be used to generate complex, careful, accurate, and encompassing information on economic activity. Our article advocates the use of these unstructured, but readily available, data for both the construction of national aggregate and distributional accounts, as well as the study of the microstructure of economic activity.

Our proof-of-concept results imply that simple and transparent procedures such as organising the data around sound national accounting principles and ensuring a representative sample of a country's population - followed by bottom-up aggregation - track with remarkable precision not only the growth rate of consumption in national accounts, but also its level.

Further, the high-granularity or **high-definition data allows immediate decomposition across goods**, **demographics, space, or time frequencies.** As an example, we show the first distributional accounts analysis of aggregate consumption, providing a rich, macro-consistent description of consumption inequality and its time evolution. It can also be used to analyse the microstructure of the economy in many ways.

Last, but not least, the real-time and high-definition components of the data show a powerful economic tool to track the economy in real time not only at aggregate levels but also **zooming out across characteristics**, **gender**, **income categories**, **and geographical details**. The combination of this information constitutes the basis for better diagnoses and furthermore for the design of smart policies.

All this is possible because transaction data (once properly organised) can be deployed as a high-quality, largescale, real-time consumption survey containing both information on the consumption decisions of millions of individuals along with rich metadata, which tags billions of transactions.



References

- Andersen, A. L., Hansen, E. T., Johannesen, N., and Sheridan, A. (2020). Consumer Responses to the COVID-19 Crisis: Evidence from Bank Account Transaction Data. SSRN Scholarly Paper ID 3609814, Social Science Research Network, Rochester, NY.
- Aladangady, A., Aron-Dine, S., Dunn, W., Feiveson, L., Lengermann, P., and Sahm, C. (2021). From Transactions Data to Economic Statistics: Constructing Real-Time, High-Frequency, Geographic Measures of Consumer Spending. In Big Data for Twenty-First-Century Economic Statistics. University of Chicago Press.
- Oriol Aspachs, Ruben Durante, Alberto Graziano, Josep Mestres, Jose G Montalvo, Marta Reynal-Querol, Realtime inequality and the welfare state in motion: evidence from COVID-19 in Spain, *Economic Policy*, Volume 37, Issue 109, January 2022, Pages 165–199,
- Baker, S. R. (2018). Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data. Journal of Political Economy, 126(4):1504–1557.
- Buda, G., Carvalho, V. M., Hansen, S., Mora, J. V. R., Ortiz, Á., Rodrigo, T. (2022) National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consum
- Carvalho, V., Garcia, J., Hansen, S., Ortiz, A., Rodrigo, T., Rodriguez Mora, S., and Ruiz, P. (2021). Tracking the COVID-19 crisis with high-resolution transaction data. Royal Society Open Science, 8.
- Chetty, R., Friedman, J. N., Hendren, N., and Stepner, M. (2020). How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data. Working Paper.
- Piketty, T., Saez, E., and Zucman, G. (2018). Distributional National Accounts: Methods and Estimates for the United States. The Quarterly Journal of Economics, 133(2):553–609.
- Vavra, J. (2021). Tracking the Pandemic in Real Time: Administrative Micro Data in Business Cycles Enters the Spotlight. Journal of Economic Perspectives, 35(3):47–66.



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.

The content of this document is protected by intellectual property laws. Reproduction, transformation, distribution, public communication, making available, extraction, reuse, forwarding or use of any nature by any means or process is prohibited, except in cases where it is legally permitted or expressly authorised by BBVA on its website www.bbvaresearch.com.