

EU RO NA

**EUROSTAT REVIEW
ON NATIONAL ACCOUNTS
AND MACROECONOMIC
INDICATORS
2021**

**EUROSTAT REVIEW
ON NATIONAL ACCOUNTS
AND MACROECONOMIC
INDICATORS**

2021

Manuscript completed in July 2021

The European Commission is not liable for any consequence stemming from the reuse of this publication.

Luxembourg: Publications Office of the European Union, 2021

© European Union, 2021



The reuse policy of European Commission documents is implemented based on Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39).

Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.

Theme: Economy and finance

Collection: Statistical working papers

PDF:

ISBN 978-92-76-40305-0 ISSN 1977-978X doi:10.2785/935251 KS-GP-21-001-EN-N

Contents

Editorial	5
Indian household balance sheet: accounting issues and wealth estimation Akash Malhotra	7
A note on service(s) in recent European Union legal texts: is online digital content necessarily a service? Damien Broussolle	41
A systematic approach for data validation using data-driven visualisations and interactive reporting Jules Zaccardi and Enrico Infante	67

Aims and scope

EURONA is an open access, peer-reviewed, scholarly journal dedicated to National Accounts and Macroeconomic Indicators. EURONA aims at providing a platform for researchers, scholars, producers and users of macroeconomic statistics to exchange their research findings, thereby facilitating and promoting the advancement of National Accounts and Macroeconomic Indicators.

EURONA publishes empirical and theoretical articles within the scope of National Accounts and Macroeconomic Indicators, as well as articles on important policy uses of these statistics. They may relate to both users' and producers' interests, present subjects of general relevance or investigate specific topics.

EURONA is non-partisan and applies the highest standards to its content, by emphasising research integrity, high ethical standards, validity of the findings and cutting edge results. EURONA gives room to all viewpoints.

The articles published in EURONA do not necessarily reflect the views or policies of the European Commission.

Website: <https://ec.europa.eu/eurostat/web/national-accounts/publications/eurona>

Contact: ESTAT-EURONA@ec.europa.eu

Editors

Paul Konijn, Eurostat

Nicola Massarelli, Eurostat

Giuliano Amerini, Eurostat

Editorial board

John Verrinder, Eurostat

Albert Braakmann, Statistisches Bundesamt

Gerard Eding, Centraal Bureau voor de Statistiek

Rosmundur Gudnason, Statistics Iceland

Robert Inklaar, University of Groningen, the Netherlands

Sanjiv Mahajan, Office for National Statistics

Gabriel Quiros, International Monetary Fund

Philippe Stauffer, Federal Statistical Office

Peter van de Ven, formerly Organisation for Economic Co-operation and Development

Editorial

I am pleased to introduce this issue of EURONA. It has three stimulating contributions on a variety of topics.

In the first article, Akash Malhotra presents household financial balance sheets for India for the period from 1970 to 2018. The data shed a new light on the developments of Indian household finance. The author compares a set of new data with existing estimates and also compares key indicators with those for other countries and regions of the world. The author concludes that frequent compilation of household balance sheets is instrumental in monitoring the financial situation of Indian households and possible financial risks they face.

In the second article of this issue, Damien Broussolle explores the concept of (a) service(s), in particular those with digital content, in relation to the use of the concept in national accounts compared with the use in European Union (EU) legal texts. The author argues that lawyers could learn from economists to develop a coherent concept based on ownership rights rather than on the tangible nature of goods or services. Naturally, the issue is important within the context of the ongoing digitalisation of EU economies. The author was inspired by the works of T. Peter Hill on this subject and the article is a tribute to him.

Jules Zaccardi and Enrico Infante present, in the third and last article, a systematic visualisation approach to data validation. They have developed an interactive tool for the validation of annual financial accounts, but the tool can be used for other official statistics as well. It aims to contribute to the efficiency of the validation process as well as to the overall quality of the data. The tool is illustrated in the paper with examples from financial accounts.

I hope that you enjoy reading these contributions.

Paul Konijn

Editor of EURONA

1

Indian household balance sheet: accounting issues and wealth estimation

AKASH MALHOTRA (*)

Abstract: This paper presents a novel estimate of the Indian household balance sheet (HBS) starting from 1970/71 to 2017/18 and studies the evolution of Indian household finance in an international context. Comparative analysis suggests that the Global Wealth Databook (an annual publication of Credit Suisse Research Institute), has been significantly underestimating both the stock of financial assets and liabilities. By contrast, the Reserve Bank of India, in its recent publication, is marginally overestimating the stock of financial assets and institutional liabilities. An analysis of the net financial position of the Indian household sector reveals that India is presently facing what may be called a five balance sheet challenge — critical HBS leverage ratios have continued to climb even during the recovery period post-2010 (following the global financial and economic crisis). The author recommends close monitoring and more frequent releases of the Indian HBS.

JEL codes: D14, E01, E21, G51, O57

Keywords: household balance sheet, wealth accumulation, household debt, financial sectoral accounts, India

(*) Centre for Economic Studies and Planning, Jawaharlal Nehru University, New Delhi, 110067, India.
E-mail: akash_malhotra@iitb.ac.in.

1. Introduction

Sectoral balance sheets offer a key stock perspective. They shed light on the structure of sectoral finance, wealth accumulation, and the associated default risks, along with other statistical and macroeconomic issues. National accounts are supposed to have balance sheets for each sector. Unfortunately, at the time of writing, India does not have an official balance sheet for the household sector. In most high-income countries, household balance sheets (HBS) are available quarterly, or at least annually (for example, the OECD publishes these data for its member countries in *Households' financial assets and liabilities*). There is certainly a very high user demand for such data (Shorrocks et al. (2019)). Recently, an increasing number of economies have begun completing their national accounts by publishing sectoral balance sheets, at least for financial assets and liabilities (Shorrocks et al. (2019)). Alternative estimates of the HBS for China have been released, namely by Li (2018) and Piketty et al. (2019). India has endorsed the second phase of the G20 Data Gaps Initiative (DGI-2) and accordingly, NSC (2018) made recommendations related to timeliness and the more frequent release of financial accounts, state-level coverage, and the inclusion of new aspects such as flows versus stocks in the sectoral accounts, among other methodological improvements. This paper is an attempt to move in the same direction — here I compile a balance sheet for the Indian household sector which presents the evolution of the outstanding stock of household wealth and debt.

Compared with debates on income growth, household wealth accumulation in India has received less consideration from academics as well as policymakers. This is partly because, in policy circles and academic debates alike, income — particularly GDP figures — commands more attention than wealth. But it is also because of the dearth of appropriate data on household wealth or debt in the country. Even though India has a long history of collecting data on household debt through its decennial wealth surveys, indebtedness in the household sector continues to be underestimated, or at least to be perceived as less of a problem. Presently, the decennial all India debt and investment survey (AIDIS) serves as the lone source of primary data for household wealth in India. The Credit Suisse Research Institute has been publishing an annual *Global Wealth Databook* (GWD) since 2010. This contains estimates for the stock of household wealth and debt starting from the year 2000 for a large number of countries, including for India based on the AIDIS data. The estimates of the HBS provided in the GWD are widely-used in policy and journalistic circles. However, as I illustrate in this paper, the GWD has been misjudging the financial position of Indian households. For the first time ⁽²⁾, a preliminary estimate of the HBS has been released by the Reserve Bank of India (RBI (2019)) for the period ⁽³⁾ 2011/12 to 2017/18. However, as I show in this paper, the HBS presented in RBI (2019) is plagued with accounting errors and, as a consequence, overestimates the stock of household financial assets and liabilities alike.

Preparing balance sheets for any sector requires data on its stocks of assets and liabilities to be available. Such data are not available in the Indian national accounts statistics (NAS) simply because the data collection efforts of the Indian Ministry of Statistics and Programme

⁽²⁾ Prior to RBI (2019), a partial balance sheet for the household sector was released with RBI (2018), which reported data on outstanding positions for a select few financial instruments but did not report the stock of gross financial assets and liabilities or the net financial position of the household sector.

⁽³⁾ The Indian Fiscal Year begins on 1st April and ends on 31st March of the next calendar year. For instance, fiscal year 2011/12 represents the year starting on 1 April 2011 and ending on 31 March 2012.

Implementation (MoSPI) focus on data needed to compile the current accounts, rather than on data relevant to accumulation accounts. The approach I use here, to derive the HBS, is a perpetual inventory approach that relies on accumulating the flow-of-funds (FoF) while making appropriate technical adjustments and assumptions regarding the initial value of stocks. The FoF data are available back to at least 1970/71. The availability of FoF data is going to form the crux of the methodology employed for the compilation of the HBS here. All things considered, this paper attempts to create a paradigm for future studies aimed at improving the reliability of the Indian HBS data and estimation methods. The rest of the paper is organised as follows: Section 2 describes the data sources and major accounting issues related to the HBS compilation for India. Section 3 highlights the deficiencies in the extant HBS estimates and analyses the development of the net financial position of the Indian household sector over time and the composition of household wealth and debt, based on the HBS computed for this paper. Section 4 compares the structure of Indian household finance with that found in other regions/economies of the world and diagnoses the risks arising from increasing household indebtedness. Section 5 presents some concluding reflections and suggestions pertinent to future avenues of research.

2. Data and methodology: major accounting issues

According to the Indian System of National Accounts (CSO (2012)), the 'household' sector in India comprises of individuals, unincorporated establishments (like sole proprietorships and partnerships), non-profit institutions serving households (like educational institutions, charitable trusts, non-governmental organisations (NGOs) political parties, and so on) and all non-government non-corporate enterprises (like farms and non-farm businesses). Thus, in addition to individuals, the household sector also includes all enterprises/economic units that are not covered in the other three domestic sectors⁽⁴⁾ of the economy. Technically, this interpretation is different from the definition of 'households' as proposed by the United Nations System of National Accounts — SNA 2008 (United Nations (2009)), which includes only individuals and groups of persons sharing the same living accommodation or pooling some or all of their income and wealth, wherein each member of the household has some claim upon the collective resources of the household; it thereby excludes non-profit institutions. In this article, I will continue to follow the Central Statistics Office (CSO (2012)) definition of 'households', unless stated otherwise⁽⁵⁾. In NAS, the gross financial savings made by the household sector are estimated by net changes in the financial position of households for a list of financial assets: currency, deposits, trade debt, shares & debentures, claims on government, insurance funds, and provident & pension funds. The annual flows for liabilities are also reported instrument-wise: bank advances, loans and advances by cooperative banks and societies, loans by financial corporations & non-banking companies (for example, non-banking financial corporations (NBFCs)), loans and advances from government, and loans from insurance companies. Since the household sector is not an organised sector and direct

⁽⁴⁾ The Indian FoF categorises the economy into four domestic sectors: *financial corporations, non-financial corporations, general government, households*; and one external sector: *rest of the world* (RBI, 2015).

⁽⁵⁾ The size of the household sector produced from the CSO (2012) definition is likely to be larger than that from the SNA 2008 definition.

estimates of its balance sheets are not available, the financial flows for various instruments are either estimated through (i) a residual approach, in other words, after duly accounting for such instruments held by public and private corporate sectors, or (ii) through business/enterprise information collected from the accounts of counterpart institutions transacting with households, or (iii) through existing information on sectoral distributions (direct or survey). The only two instruments for which data on flows are directly available are provident & pension funds and life insurance funds:

$$\Delta(\text{provident \& pension funds}) = \text{contribution} + \text{interest} - \text{withdrawal}$$

$$\Delta(\text{life insurance funds}) = \text{income} - \text{expenditure}$$

Unless stated otherwise, the data on various balance sheet items presented in this paper have been sourced from NAS Statement 5.3: changes in *Financial assets and liabilities of the household sector*. Presently, there is an internal arrangement to supply input data for the preparation of Statement 5.3 from the RBI to the CSO. For a detailed exposition of the methodology employed by the RBI in the compilation of FoF accounts, the reader may refer to RBI (2015). In line with the extant practice, changes in financial assets and liabilities of the household sector are estimated in the form of financial flows based on counterparty sector data, in other words, from the central bank, commercial banks, NBFCs, insurance companies, housing finance companies, mutual funds, the general government sector, non-financial corporates, and so on. It may be noted that counterparties do not always provide assets and liabilities specific to the household sector to the RBI. In such cases, flows are estimated using multiple rates and ratios based on various sectoral reports. At times, this also involves the logical judgement of RBI staff. In a personal correspondence with the Reserve Bank of India (dated 26 July 2018) under the *Right to Information Act, 2005*, I was let known that the stock data are neither published nor preserved by the RBI exactly for the above reasons.

However, in my opinion the above-mentioned circumstances do not pose any serious limitation to the reconstruction of a time series of stock data by accumulating the flow data over time, as the FoF account for the household sector, or for that matter the entire FoF matrix, does incorporate transactions, revaluations and other changes in the volume of assets (OCVA). However, this disaggregation was not compiled separately by the RBI until recently ⁽⁶⁾. Essentially, the FoF accounts, as available in NAS, represent annual changes in stock, in other words, period to period changes in the outstanding amounts of financial assets and liabilities. This allows me to use the perpetual inventory approach to derive the HBS from the household FoF data. The approach relies on accumulating flows over time while making appropriate technical adjustments and assumptions regarding the initial value of stocks.

⁽⁶⁾ With the release of balance sheet data (2011/12 to 2017/18) for institutional sectors, RBI (2019) attempted the bifurcation of financial flows into transactions and valuation changes for mutual funds, insurance, provident & pension funds, households and the central bank.

Table 1: India's household balance sheets, 2003/04 to 2017/18 (¹)
(INR billion, current prices)

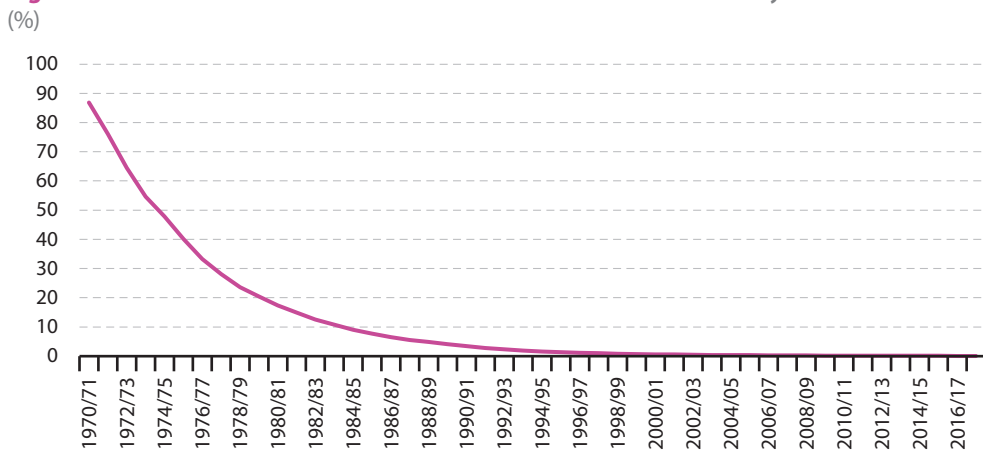
	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18
1. Financial assets	29 255	33 727	39 569	47 216	54 940	62 208	72 106	82 905	92 232	102 873	114 780	127 353	142 315	156 699	175 395
1.1. Currency	2 905	3 274	3 796	4 468	5 281	6 202	7 172	8 543	9 606	10 721	11 716	13 049	15 055	11 890	16 598
1.2. Deposits	12 094	13 844	16 504	20 908	24 949	29 359	33 508	39 110	44 515	50 577	57 247	63 370	69 815	79 495	84 848
1.2.1. Bank deposits	11 006	12 757	15 414	19 707	23 597	27 775	31 756	37 239	42 499	48 250	54 643	60 436	66 659	76 045	81 147
1.2.2. Non-banking deposits	1 207	1 207	1 213	1 258	1 271	1 419	1 604	1 655	1 755	2 034	2 262	2 552	2 732	2 983	3 192
1.2.3. Trade debt (owned) (¹)	-119	-120	-122	-57	81	166	148	216	261	293	341	383	424	467	509
1.3. Shares & debentures	1 633	1 683	2 017	2 522	3 262	3 211	3 660	3 677	3 842	4 012	4 202	4 405	4 689	5 132	5 762
1.4. Claims on government	4 054	5 119	5 990	6 182	5 899	5 623	6 058	6 354	6 135	6 064	6 294	6 304	6 983	7 614	8 436
1.5. Insurance funds	3 464	4 144	4 979	6 128	7 826	9 355	11 953	14 054	16 011	17 810	19 855	22 848	25 490	29 033	32 537
1.6. Provident & pension funds	5 105	5 663	6 283	7 008	7 723	8 457	9 756	11 167	12 124	13 689	15 467	17 376	20 283	23 535	27 214
2. Institutional financial liabilities	4 908	6 108	7 946	10 772	12 654	14 290	16 324	19 104	22 005	25 309	28 896	32 664	36 518	41 205	48 610
2.1. Bank advances	4 087	5 207	6 962	9 699	11 494	13 041	14 986	17 676	20 401	23 488	26 512	29 336	32 030	35 490	40 429
2.2. Advances from government	122	117	112	106	103	101	100	99	102	104	110	112	116	116	124
2.3. Advances from other financial institutions	640	723	807	900	987	1 075	1 164	1 252	1 425	1 640	2 197	3 139	4 295	5 522	7 980
2.4. Advances from co-operative non-credit societies	59	62	65	67	69	72	74	77	77	77	77	77	77	77	77
3. Non-institutional liabilities	2 618	3 144	3 940	5 138	5 798	6 278	6 865	7 675	8 426	9 692	11 065	12 508	13 984	15 778	18 614
4.a. Net institutional financial wealth (1-2)	24 348	27 619	31 623	36 444	42 286	47 919	55 782	63 801	70 227	77 563	85 884	94 689	105 797	115 494	126 784
4.b. Net institutional financial wealth (Moore's estimate in 1951)	24 448	27 719	31 723	36 544	42 386	48 019	55 882	63 901	70 327	77 664	85 984	94 789	105 897	115 594	126 884
5. Net total financial wealth (4.b-3)	21 830	24 575	27 784	31 406	36 588	41 740	49 017	56 226	61 901	67 972	74 919	82 281	91 913	99 816	108 270

(¹) A positive (or negative) sign before the net stock position of trade debt implies that the household sector, as a whole, is acting as a net sundry creditor (or debtor) to the other sectors.
Source: author's calculations

2.1 Initial value assumption: robustness checks

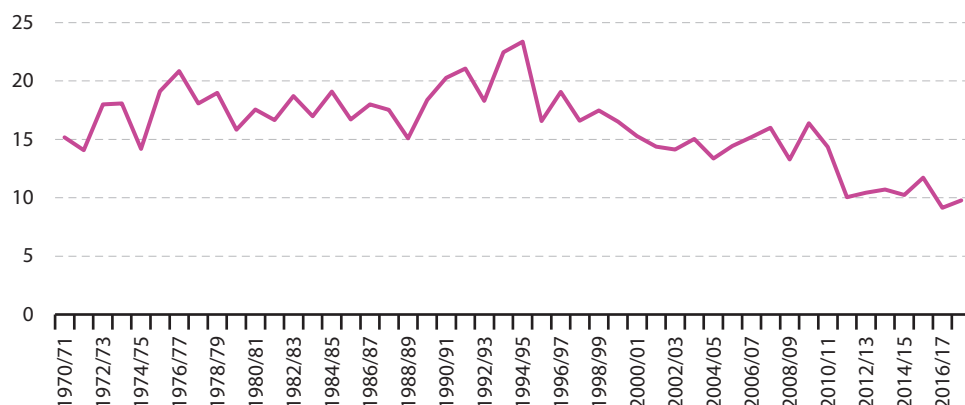
The FoF data for net institutional financial savings (= gross financial savings – changes in institutional financial liabilities) are available starting from 1950/51. However, constrained by the availability of segregated FoF data for most of the individual financial items (except for currency, provident & pension funds, and shares & debentures), I begin cumulating FoF flows starting from 1970/71. A link to the computed HBS is provided in the Appendix at the end of this article and an excerpt from the same is presented in Table 1. The initial stock value for all financial items is assumed to be zero on 31 March 1970. Using the 1969/70 stock data as a benchmark position, the subsequent stock data are constructed by incrementing the flow data collected from the net changes reported in FoF data for each item. Fortunately, there exist FoF data for net institutional financial savings, currency, and provident & pension funds starting from 1950/51 in older CSO records and estimates of stock data on 31 March 1951 for these items in Moore (2007). These allow me to test the validity of my zero initial value assumption and compute the magnitude of deviations caused, if any. Using Moore's estimate for the stock of institutional financial wealth on 31 March 1951 and FoF data for net institutional financial savings from 1950/51 to 2017/18, I am able to compile an alternate time series for the stock of financial wealth (net of institutional liabilities) held by Indian households, referred to as *net institutional financial wealth (Moore's estimate in 1951)* in Table 1 (4.b.). As apparent from Table 1, the two series (4.a. and 4.b.) converge and the stocks of net institutional financial wealth as estimated by the two series differ from each other only by 0.08 % in 2017/18. This indicates that my assumption of ascribing zero value to the stock of institutional financial wealth on 31 March 1970 does not cause significant deviations in the recent past as corroborated by Figure 1, which shows that deviations become less than 3 % after 1990/91 and fall below 1 % after 1996/97. Therefore, for the rest of the paper, my period of analysis will be from 1990/91, at the earliest (?) for all HBS items with constraints on FoF time series data availability, in other words, items for which FoF data are not available before 1970/71 in NAS.

Figure 1: Deviation in net institutional financial wealth as estimated by the two series



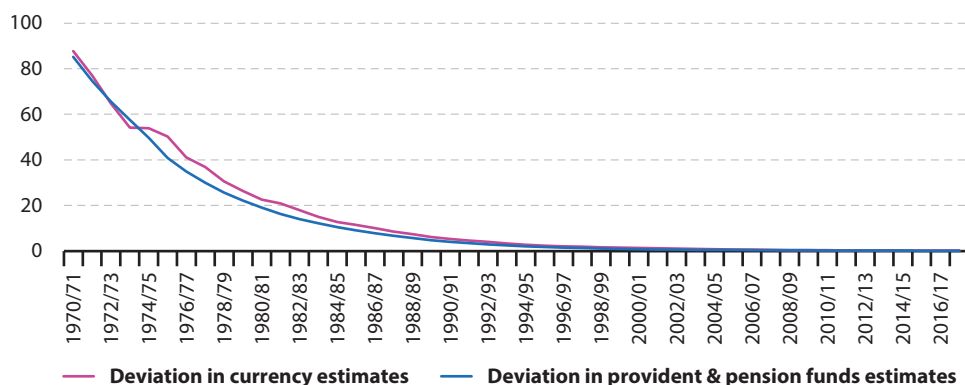
(?) Determining whether an error lies within a tolerable margin is, ultimately, a subjective exercise. Therefore, I leave the choice of the earliest year (starting from which the compiled HBS could be considered reliable) to the judgement of future users of this HBS data. Accordingly, deviations reported in Figures 1 and 3 shall come in handy while making such judgement calls.

Figure 2: Rate of accumulation of net Institutional financial wealth (Moore's estimate in 1951)
(%)



An apparent reason for the validity of my assumption is the consistently high rate of financial wealth (net of institutional liabilities) accumulation by Indian households which has remained above or close to 10 % (in nominal terms) since 1970/71, as illustrated in Figure 2. A consistently high accumulation rate implies that the absolute value of net institutional financial savings in 2017/18 (INR 11 290 billion) dwarfs the net institutional financial savings made by Indian households prior to 1970/71 (less than INR 15 billion annually), thereby trivialising the contribution of savings made a long time ago. Notice that the rate of financial wealth accumulation dipped significantly after the 2008 global financial and economic crisis, from levels around 15 % to levels around 10 %, and has since remained at this lower level.

Figure 3: Deviation in currency and provident & pension funds as estimated by the two series
(%)

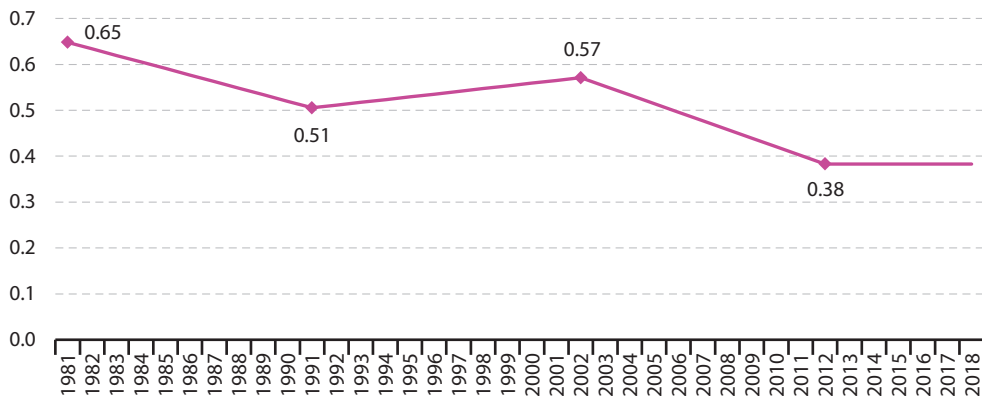


The same pattern is observed for currency and for provident & pension funds (P&PF) items in the HBS; for both these items, the two estimates have converged in recent years. The deviations for both currency and for P&PF have been less than 5 % since 1990/91 (see Figure 3). In 2017/18, the deviation in the two stock estimates for currency was 0.15 % and for P&PF it was 0.10 %. The convergence observed in these series further supports my assumption of ascribing zero value (on 31 March 1970) to the stock of those financial assets/liabilities for which flow data are not available prior to 1970/71.

2.2 Estimating non-institutional credit

In this article, I do not attempt to estimate the non-financial component of the HBS, mainly due to the paucity of reliable flow/stock data for physical assets ⁽⁶⁾, namely land holdings, dwellings, precious metals, automobiles and other consumer durables. I do, however, estimate the stock of outstanding cash loans borrowed by households from non-institutional credit agencies such as moneylenders, landlords, traders, input suppliers, and so on. The FoF data made available in NAS do not include credit from non-institutional sources and an estimate of the stock of debt obtained from FoF data represents only the debt raised from institutional agencies. To produce a reasonable estimate of the stock of non-institutional debt, I compute the ratio of non-institutional to institutional household debt at the all-India level from various rounds ⁽⁷⁾ of AIDIS and apply these ratios on the stock of institutional debt estimated from the RBI's FoF data. The under-reporting of household debt in India's wealth survey is widely recognised in the literature (Shorrocks et al. (2019)) and is discussed in Section 3.2 in detail. However, it is likely that survey participants would under-report institutional and non-institutional debt by roughly the same factor, as there is no apparent reason for significant inequality in the magnitude of under-reporting among these two types of debt. The ratios for years between various rounds of AIDIS surveys — 1981 to 1991, 1991 to 2002, and from 2002 to 2012 — are estimated from three separate linear interpolations, and the ratios for years post-2012 are ascribed the same value as that in 2012 (see Figure 4). This imputation is admittedly crude but better than simply disregarding the entire non-institutional debt owed by households. The net total financial wealth, as reported in Table 1 (5.), is computed by subtracting the outstanding institutional and non-institutional household debt from the stock of financial assets.

Figure 4: Ratio of non-institutional to institutional household debt



Source: AIDIS (1981, 1991, 2002 and 2012 rounds)

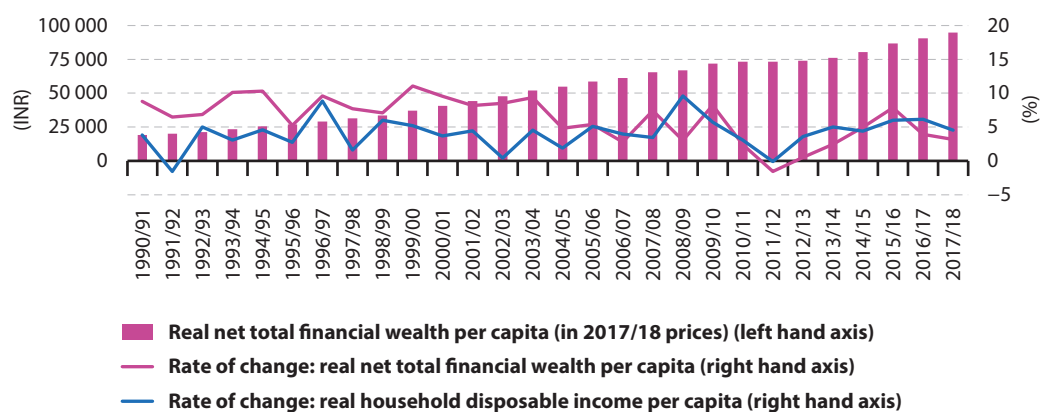
⁽⁶⁾ Note that survey estimates for *household* holdings in these non-financial assets are available in various rounds of NSSO surveys and AIDIS which are conducted from time to time.

⁽⁷⁾ I use data from the last four rounds of AIDIS, namely 1981, 1991, 2002 and 2012. The AIDIS data represent the value of stocks on 30 June of the corresponding year. I ascribe the computed ratios for a particular round of AIDIS to 31 March of the corresponding year. For example, the ratio computed from the 1991 AIDIS is ascribed to the stock at the end of fiscal year 1990/91.

3. Indian households' net financial position: wealth and composition analysis

The financial position (net of total liabilities) of the Indian household sector in 2017/18 was INR 108 trillion⁽⁹⁾. The net total financial wealth per capita, in 2017/18 prices and as adjusted by the GDP deflator, grew significantly from INR 16 720 in 1990/91 to INR 82 272 in 2017/18 (see Figure 5). In the last three decades, the per capita financial wealth (net of total liabilities) grew on average by 6.2 % annually in real terms. This accumulation was supported partly by an average 4.1 % annual growth in per capita real household disposable income (see Figure 5). After adjusting for changes in price levels, the stock of financial assets owned by Indian households rose from INR 21 trillion in 1990/91 to INR 175 trillion in 2017/18, whereas the outstanding stock of household debt increased from INR 8 trillion to INR 67 trillion in the same period (see Figure 6).

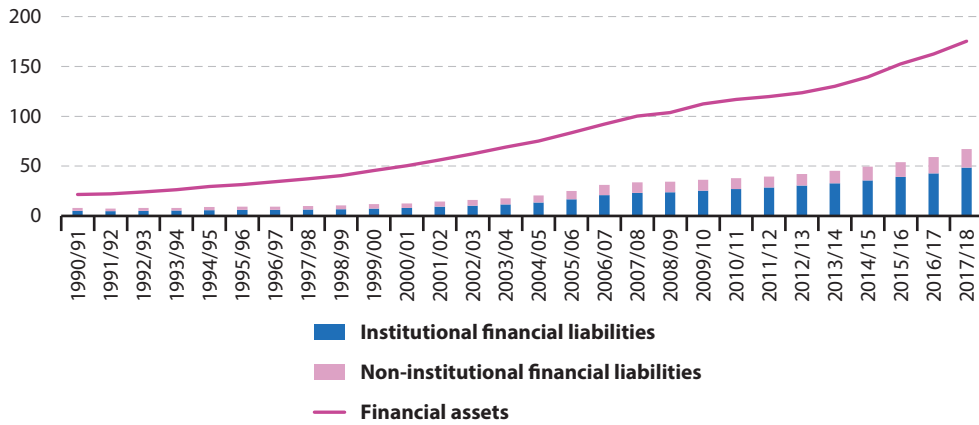
Figure 5: Evolution of per capita net total financial wealth (NTFW) and household disposable income (HDI)



⁽⁹⁾ A trillion is a million million or 1 000 000 000 000.

Figure 6: Stock of financial assets and liabilities (in 2017/18 prices)

(INR trillion, in 2017/18 prices)

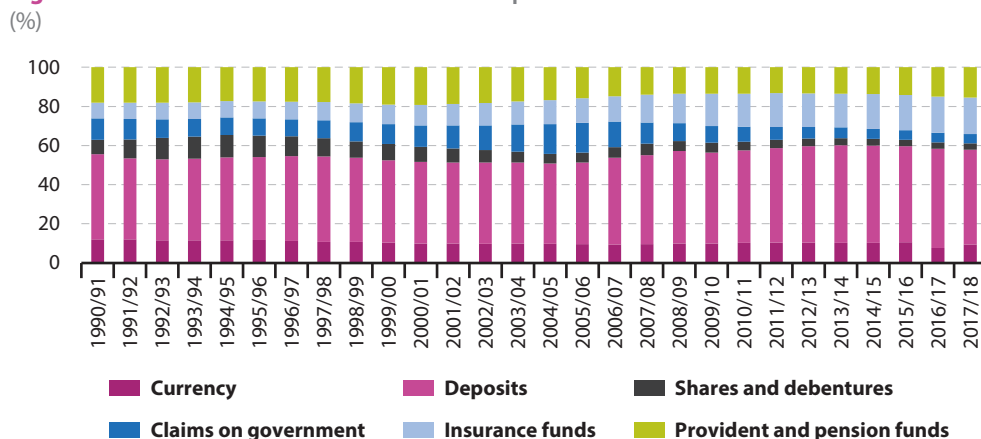
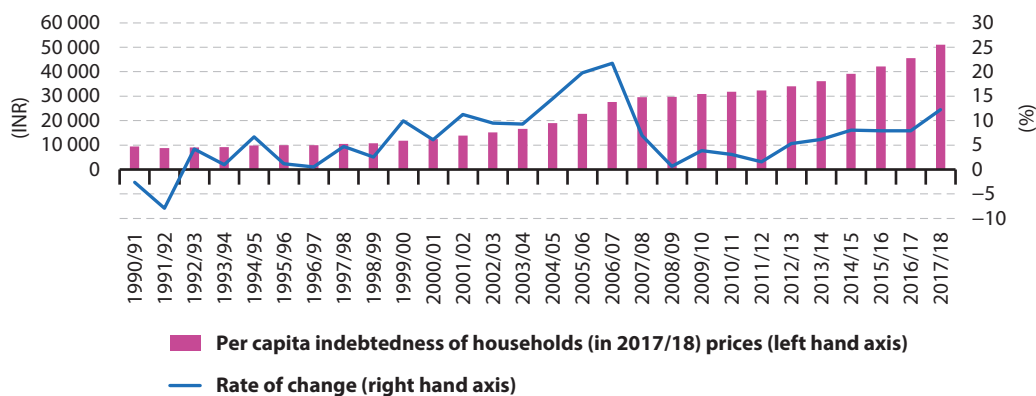


The pattern of ownership of financial assets varied widely owing to a few asset classes as depicted in Figure 7. The combined share of currency and deposits, the most preferred financial instruments, remained more or less constant (55 % on an average) over time. The share of life insurance funds ⁽¹⁾ increased consistently from 8.1 % in 1990/91 to 18.6 % in 2017/18, whereas the share of provident & pension funds ⁽²⁾ in the total stock of financial assets owned by the household sector declined gradually from 18.1 % in 1990/91 to 15.5 % in 2017/18. The asset class of shares & debentures ⁽³⁾ offers a surprising observation — its weight in the total household holdings of financial assets was increasing in the early 1990s (from 7.3 % in 1990/91 to 10.9 % in 1995/96), but then declined and remained in the range of 3–4 % in the most recent five years. This observation is contrary to the everyday reporting by the financial press and also to the claim made by RBI (2019, p. 51) that currency and deposits combined had been losing their share in the portfolios of Indian households to equities and debt securities over time. Interestingly, the share lost by the asset class shares & debentures between 1995/96 and 2017/18 was commensurate with the share gained by life insurance funds over the same period. This is not a dramatic shift as the reserves of life insurance funds comprise equities, bonds, and other financial instruments that are in effect being held on behalf of the household sector.

⁽¹⁾ The heading of life insurance funds includes central or state government employees' insurance funds and postal insurance funds. The asset position is estimated from actuarial reserves and other technical reserves for entitlements relating to individual life insurance policies (RBI (2018)).

⁽²⁾ The asset position for provident & pension funds is estimated from entitlements relating to funded retirement benefits for employees of the government and non-government sectors (RBI (2018)).

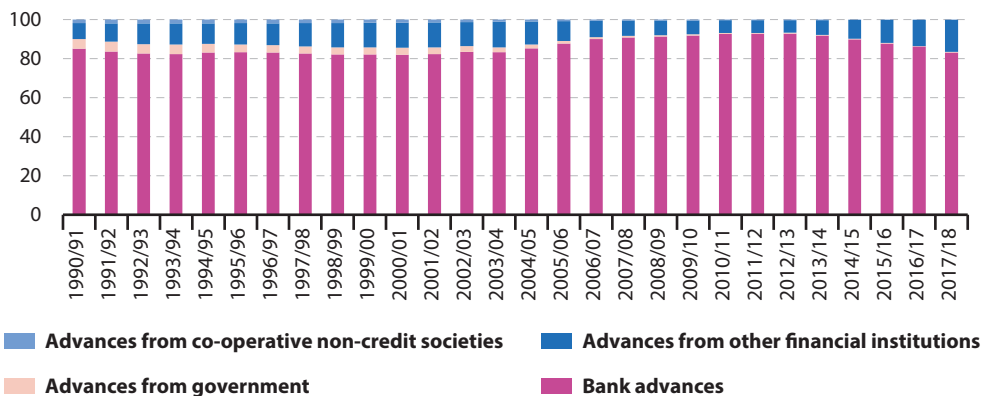
⁽³⁾ The heading of shares & debentures includes investment in shares & debentures of credit/non-credit societies and investment in mutual funds.

Figure 7: Structure of financial asset ownership for households**Figure 8: Evolution of household indebtedness**

Based on my HBS estimates in Table 1, the stock of total financial liabilities or indebtedness of the household sector rose significantly over time (in nominal terms) from INR 1.4 trillion in 1990/91 to INR 67.2 trillion in 2017/18. In per capita terms, the real indebtedness of the Indian household sector rose from INR 9 460 in 1990/91 to INR 51 082 in 2017/18. However, the rate of growth of indebtedness fluctuated wildly over time with a huge spike in the period building up to the global financial and economic crisis of 2008 and a pronounced fall thereafter as depicted in Figure 8.

The structure of household financing (from institutional sources) has changed over time as reflected in Figure 9, with loans from cooperative non-credit societies and government vanishing almost completely from the HBS and bank loans dominating even more. Commercial banks and cooperative banks and societies remained the preferred credit institution for Indian households, with the share of bank advances ⁽¹⁴⁾ varying in the range of 80-90 % of total household debt. However, in the most recent five years other financial institutions ⁽¹⁵⁾ such as NBFCs (whose share rose from 7.6 % in 2013/14 to 16.4 % in 2017/18) have taken some share of household debt from traditional banks, with their share falling from 91.7 % to 83.2 % between 2013/14 and 2017/18. In the context of Indian household finance, there exists a complementarity between the banking sector and non-banking financial companies in their financial intermediation roles, which was also reported by RBI (2019). As can be seen from Figure 9, during the periods in which the share of loans and advances from banks rose (83.3 % to 91.7 % between 2003/04 and 2012/13) the share of other financial institutions declined (13 % in 2003/04 to 7.6 % in 2012/13), while the reverse was true between 2013/14 and 2017/18.

Figure 9: Evolution of the composition of household debt from institutional sources (%)



3.1 Comparison with other estimates of India's HBS

Historically, the preferred estimate of balance sheet data for Indian households has been based on the AIDIS, which is carried out every 10 years, the most recent round being the 2012-round (NSSO (2013)). Apart from the long gaps in data release, household wealth surveys are generally not a reliable source for estimating wealth levels due to sampling and non-sampling errors which get magnified in the case of countries with high wealth inequality, such as India. The high skewness of wealth distribution makes the sampling error more pronounced. Further, the non-sampling errors, arising due to differential response rates ⁽¹⁶⁾ and under-reporting ⁽¹⁷⁾, make it difficult to extract an accurate representation of the upper tail of wealth distribution — where the bulk of wealth lies in the case of unequal societies. Consequently, household surveys usually produce lower wealth totals, especially in the case of financial assets and liabilities, when compared with the HBS data (Shorrocks et al. (2019)).

⁽¹⁴⁾ The heading of bank advances includes advances by banks and co-operative banks and societies.

⁽¹⁵⁾ The heading of loans and advances from other financial institutions includes advances by financial corporations & non-banking companies and insurance corporations.

⁽¹⁶⁾ Wealthier households are less likely to participate in the survey.

⁽¹⁷⁾ Households are likely to under-report their financial assets and liabilities.

Since 2010, the Credit Suisse Research Institute has published the *Global Wealth Databook* annually; it contains estimates of global household wealth covering all regions and countries. The most recent edition, GWD 2019, puts the total net wealth (financial plus non-financial, net of total liabilities) of Indian households at INR 875 trillion ⁽¹⁸⁾ (Shorrocks et al. (2019)). As compared with the HBS compiled in this paper, GWD 2019 underestimates household gross financial wealth by around 50 % from 2000/01 to 2009/10. The estimates provided by GWD 2019 and the computed HBS series converge after 2010/11, and in the last two years of the analysis period are similar (see Figure 10 and Table 2). However, GWD 2018 severely underestimates gross financial wealth owned by Indian households, by 72 % on average for all the years reported (see Figure 10 and Table 2). The reason behind this discrepancy between the two GWD editions is a change in methodology for estimating Indian household financial wealth by the authors of the GWD while advancing from the 2018 to the 2019 edition. The authors of GWD create an econometric model (elucidated in Davies et al. (2017)) of per capita wealth using data from countries with HBS or survey data in at least one year. They use this

Table 2: Estimates of gross financial wealth
(INR trillion, current prices)

	HBS (author's estimate)	GWD 2019 estimate (¹)	GWD 2018 estimate (²)	RBI 2019 estimate (³)
2000/01 (⁴)	19.3	9.7	1.6	
2001/02	22.1	9.8	7.6	
2002/03	25.4	12.3	9.6	
2003/04	29.3	15.2	11.5	
2004/05	33.7	16.6	12.3	
2005/06	39.6	19	12.7	
2006/07	47.2	18.5	16.3	
2007/08	54.9	29.7	19.4	
2008/09	62.2	30.8	16.2	
2009/10	72.1	29.6	14.1	
2010/11	82.9	47.9	17	
2011/12	92.2	63.4	23.9	105.7
2012/13	102.9	75.4	28.1	116.5
2013/14	114.8	93.9	29.2	129.6
2014/15	127.4	98.7	33.6	145.5
2015/16	142.3	122.5	38	161.6
2016/17	156.7	150	39	179.5
2017/18	175.4	186.1	40.7	202.6

(¹) Source: Shorrocks et al. (2019). For 2000/01 to 2016/17, the tabulated data represent the outstanding stock on 31 December in the given fiscal year. For 2017/18, the data represent the average of the stocks on 31 December 2017 and 30 June 2018.

(²) Source: Shorrocks et al. (2018). For 2000/01 to 2015/16, the tabulated data represent the outstanding stock on 31 December in the given fiscal year. For 2016/17, the data represent the average of the outstanding stocks on 31 December 2016 and 30 June 2017. For 2017/18, the data represent the stock on 30 June 2018.

(³) Source: RBI (2019)

(⁴) The relative change in stock of financial assets between 2000/01 and 2001/02 in the GWD 2018 estimate appears to be unrealistic. GWD (2018) use adjusted survey means from the last two AIDIS rounds, namely 2002 and 2012, and accordingly extrapolate the data for the remaining years in the sample. It seems that the methodology employed in GWD (2018) is producing spurious backcasts for years before 2002.

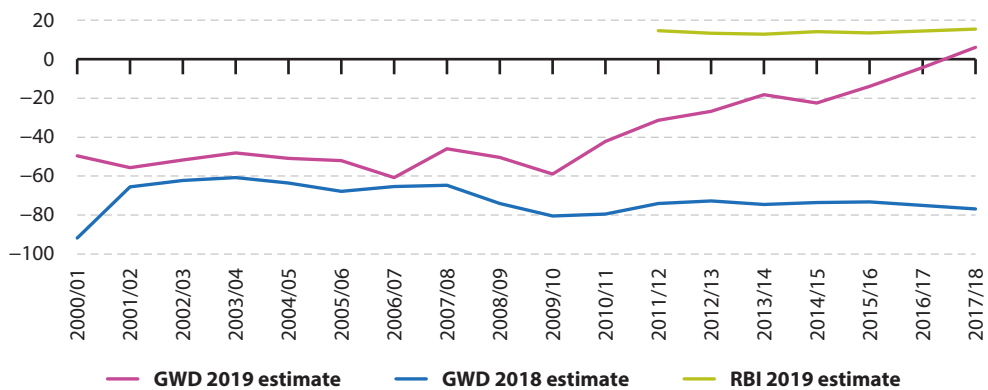
(18) Estimate as of mid-2019.

model to estimate per capita wealth levels in countries lacking direct data on household wealth; separate regressions are run for financial assets, non-financial assets and liabilities. In the model, a dummy is included for cases where the data source is a survey instead of an HBS. The coefficient of this dummy has been reported as negative and highly significant in the regression for financial assets (Shorrocks et al. (2018 and 2019)), indicating that the average levels of financial assets tend to be much lower when the data are derived from a survey rather than an HBS.

All the editions of GWD before the 2019 edition, including therefore GWD 2018, used AIDIS data on financial assets owned by Indian households. These data were adjusted upwards using the coefficient of the dummy to get an estimate of gross financial wealth in the survey year. For all years, except the survey year, the final totals are compiled by making forward projections based on estimated relationships between asset/debt totals and variables like house price indices, market capitalisation data, and GDP per capita growth in preceding years. Shorrocks et al. (2019) argue that the above methodology leads to underestimation of financial wealth, primarily because of the under-reporting of owned financial assets by Indian households in its decennial wealth survey. Therefore, GWD 2019 employs a different methodology for estimating financial wealth owned by Indian households, wherein the estimates are based on the econometric model from Davies et al. (2017), which was originally meant to be used in the case of countries lacking aggregate national data on financial assets.

Figure 10: Deviation in estimates of stock of financial assets owned by households

(%)



The RBI recently released an estimate of Indian HBS data (RBI (2019)), which consistently overestimate the stock of financial assets owned by Indian households, on average by 14 % when compared with the HBS data compiled in this paper over the period from 2011/12 to 2017/18 (see Figure 10 and Table 2). The FoF data reported within RBI (2019) incorporate changes due to transactions and revaluations; however, from the text, it is not clear whether OCVA's are included in the flow data or not. To determine this, I compare the flow data reported in RBI (2019) with data available in National Accounts Statistics (NAS) and RBI (2017a). As apparent from Table 3, the changes in financial assets reported in RBI (2019) are consistently higher than those made available in NAS; the outcome is similar when RBI (2019) flow data are compared with those of RBI (2017a). Note that the minor discrepancy between RBI (2017a) and

NAS FoF data (which have been used for the compilation of the HBS presented in this paper) is due to the fact that, when released in 2017, the best estimates available with RBI (2017a) were first and second revised estimates for 2015/16 and 2014/15 respectively. NSC (2018) reported some unavoidable statistical discrepancies between FoF accounts compiled by the RBI and the financial accounts released by the CSO. However, the report noted that these discrepancies were not significant in the case of the household sector (ibid, p. 65). If I compare the flow data for 2013/14, the year for which final estimates are available in both NAS and RBI (2017a), it becomes clear that the RBI (2019) flow data are significantly higher (by 10 %) than the official data reported in the NAS (which are, in turn, equal to the data reported in RBI (2017a)).

Table 3: Estimates of changes in financial assets
(INR billion)

	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18
RBI 2019	10 853	13 053	15 856	16 110	17 946	23 060
RBI 2017a (1)		11 908	12 826	15 142		
NAS (2)	10 640	11 908	12 572	14 962	14 384	18 696

(1) 2014/15: second revised estimate at the time of publication. 2015/16: first revised estimate at the time of publication.

(2) 2015/16: third revised estimate at the time of publication. 2016/17: second revised estimate at the time of publication. 2017/18: first revised estimate at the time of publication.

The methodology for the compilation of the FoF accounts, as listed in RBI (2015), indicates that the Indian FoF accounts published by the RBI do incorporate OCVA, apart from revaluations and transactions, and the same was confirmed after personal correspondence with the Reserve Bank of India under the *Right to Information Act, 2005*. By definition, OCVA include changes in values due to unanticipated or unintentional volume changes that are not related to transactions or revaluation (IMF (2011)). OCVA occur due to reasons related to the economic appearance and disappearance of assets, reclassifications, and external events such as debt write-offs, wars, or catastrophes (United Nations (2009)). Some common examples of OCVA related to financial assets are:

- *Liquidations or bankruptcy* — when an investor (creditor) recognises that a financial claim can no longer be collected because of bankruptcy, liquidation, or other factors, that claim is removed from his/her balance sheet (19).
- *Uncompensated seizures* — when a government decides to nationalise/seize certain industries/assets within its jurisdiction without compensation, then the equity/asset position is extinguished through a volume change.
- *Destruction of currency notes or bearer securities* — as a result of a natural catastrophe or political events.
- *Reclassification* — changes in the volume of assets due to changes in demographic assumptions in the case of insurance and pension schemes (van de Ven and Fano (2017)). Migratory flows of people can also result in the reclassification of assets, for example if a household moves from one economy to another, taking its possessions (including financial assets) with it, they are recorded as OCVA (United Nations (2009)). The reclassification of an entire NPISH (non-profit institution serving households) can occur, probably into an NPI (non-profit institution) serving some other sector of the economy and vice versa; the assets and liabilities of the unit will be transferred between the two sectors of the economy.

(19) The corresponding liability must also be removed from the balance sheet of the debtor to maintain a balance in the accounts of the total economy (United Nations (2009)).

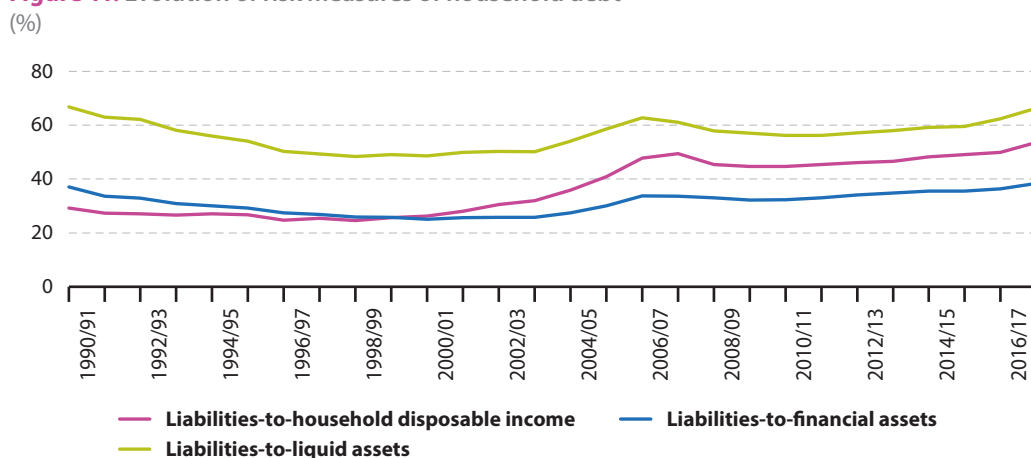
The above-mentioned examples indicate that OCVAs have mostly diminishing effects on the stock of financial assets and therefore it is likely that flows which include OCVAs will tend to be smaller than flows which omit OCVAs. If the differences observed in Table 3 are viewed in the same context, it leads me to infer that the OCVAs have not been incorporated into the flow data reported in RBI (2019) and this omission appears to be the primary reason for the deviation in financial wealth observed in Figure 10 and Table 2.

3.2 Household indebtedness in India

The risks associated with household indebtedness could be judged by comparing the stock of debt to the stock of assets or the *flow* of income. The liabilities-to-household disposable income ratio, which is a fundamental measure of the debt repayment capacity of households, has a direct application in the assessment of default risks and the solvency of the household sector (Li (2018)). The financial health of the household sector can also be assessed by the liabilities-to-financial assets ratio, which is a better measure of risk assessment than the liabilities-to-assets ratio, which also includes non-financial assets in the calculation. Non-financial assets, except gold, are plagued with liquidity risks mainly due to high trading costs, as in the case of real estate, and Akerlof's 'lemon law' of information asymmetry, as observed in markets for consumer durables such as automobiles. A still better measure of household liquidity is the ratio of outstanding household debt-to-liquid assets, wherein liquid assets comprise currency and deposits. Both currency and deposits can be liquidated in an event of financial distress without any significant loss of value ⁽²⁰⁾, unlike equities and debt securities whose value can plunge during a fire sale or market crash, or unlike pension and insurance funds which have liquidation restrictions.

By all measures, the Indian household sector has been in an upward leveraging cycle beginning from the period building up to the global financial and economic crisis (GFEC) starting in 2007 and has remained at higher levels since then (see Figure 11). The increasing debt-to-income ratio in recent years is certainly a cause of concern about the sustainability of household debt given the possibility that households might be accumulating debt for supporting consumption — which has risen at a faster pace than disposable income in every year since 2010/11. A similar picture of a worsening HBS emerges from the perspective of the liabilities-to-financial assets ratio, which has been rising in the last 15 years after a decade of balance sheet consolidation in the 1990s. Notice that this worsening of the HBS was also observed in Figure 2, which showed a significant dip in the rate of financial wealth accumulation after the crisis of 2008.

⁽²⁰⁾ Typically, Indian banks allow premature withdrawal of time deposits with a penalty in the range of 0.5-1.0 percentage points on the interest rate.

Figure 11: Evolution of risk measures of household debt

It would be interesting to compare my estimates of household indebtedness with estimates from decennial rounds of AIDIS and other subsequent estimates derived from it. As apparent from Table 4 and Figure 12, my estimates of total household indebtedness are larger than those of AIDIS (by an order of 2 or 3). This is possibly due to two main reasons. Firstly, there are differences in the definition of a 'household'. As explained in Section 2, I have followed the CSO (2012) definition of a household while constructing the balance sheet, which includes individuals, unincorporated establishments, non-profit institutions serving households, and all non-government non-corporate enterprises, whereas, the AIDIS definition of a 'household' draws from the concept of sharing a 'common kitchen' and is very similar to the United Nations (2009) definition. To be precise, AIDIS considers a 'household' to be 'a group of persons normally living together and taking food from a common kitchen' (NSSO (2013)), thus excluding NPISHs and quasi-corporations of business households. Technically, the size of the household sector according to the CSO (2012) definition would be larger than the size of the household sector derived from the AIDIS definition, and hence the tendency to produce larger estimates of household debt (or assets). Secondly, it is entirely possible that AIDIS is underestimating household indebtedness for various reasons, especially in rural areas, as has been alleged repeatedly by researchers (Gothoskar (1988); Prabhu et al. (1988); Rao and Tripathi (2001); Chavan (2012)). Rao and Tripathi (2001) blame this underestimation on the method of sampling and a reduction in the sample sizes of villages and households. Bell (1990) argues that an increase in the state sample as compared with the central sample has adversely affected the quality of AIDIS data as the state government agencies are less equipped to undertake surveys than the NSSO. Chavan (2012) shows that AIDIS underestimated rural household debt from commercial banks by about 46 % in the 1991 round and by around 35 % in the 2002 round. Rajkumar et al. (2019) compiled a supply-side estimate of institutional household debt which includes outstanding credit to individuals for different occupational activities, personal loans, professional services, and all small borrowal accounts (accounts with credit limits of up to INR 200 000) from commercial banks and cooperative banks and societies.

When compared with either Rajkumar et al. (2019) or estimates reported in this paper, AIDIS underestimates household debt owed to banks and cooperatives at the all-India level by roughly the same factor (~70 %) over the last four survey rounds (see Table 5). Rajkumar et al. (2019) also criticise the RBI supply-side data on household institutional credit, which I have used to construct the HBS in this paper, and infer that there are substantial margins of errors on both the supply-side (from RBI data) and demand-side (from AIDIS data) estimates.

Table 4: Estimates of total household debt
(INR billion, current prices)

	HBS (author's estimate)	AIDIS estimate ⁽¹⁾	GWD 2019 estimate ⁽²⁾	GWD 2018 estimate ⁽³⁾	RBI 2019 estimate ⁽⁴⁾
1980/81	288	92			
1990/91	1 401	373			
2000/01 ⁽⁵⁾	4 833		2 653	2 653	
2001/02	5 664	1 768	2 926	1 676	
2002/03	6 532		4 064	2 310	
2003/04	7 525		4 907	2 751	
2004/05	9 252		6 068	3 327	
2005/06	11 885		6 786	3 869	
2006/07	15 910		6 625	4 966	
2007/08	18 451		11 366	6 087	
2008/09	20 568		12 492	5 703	
2009/10	23 189		11 980	7 490	
2010/11	26 779		21 208	8 971	
2011/12	30 431	12 163	28 207	12 636	24 790
2012/13	35 001		33 425	15 920	28 123
2013/14	39 961		39 465	19 348	31 729
2014/15	45 172		40 611	24 666	35 521
2015/16	50 502		49 701	32 451	39 468
2016/17	56 983		56 731	38 745	43 279
2017/18	67 225		70 029	48 587	50 743

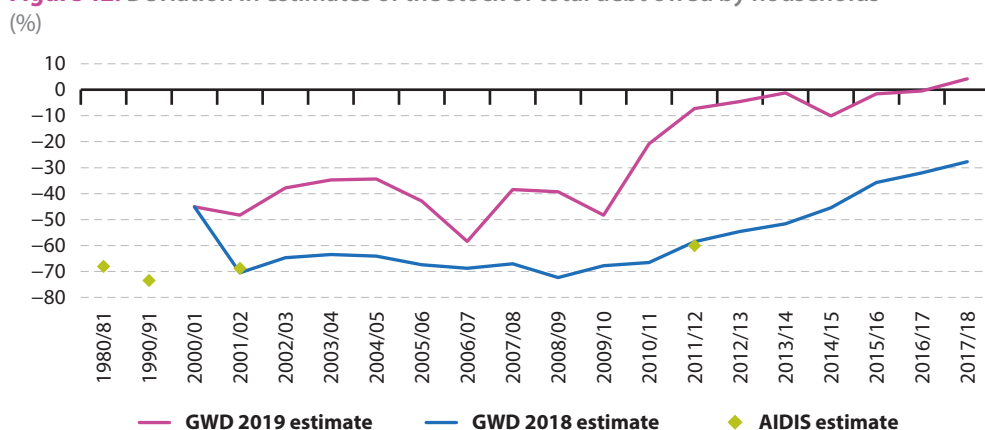
⁽¹⁾ Source: last four rounds of AIDIS, namely 1981, 1991, 2002 and 2012. The AIDIS data represent the value of stock on 30 June of the corresponding year. The stock of debt at the end of a fiscal year has been approximated by the household debt reported in AIDIS of the corresponding year. For example, the household debt reported in 1991 AIDIS is ascribed to stock at the end of 1990/91.

⁽²⁾ Source: Shorrocks et al. (2019). For 2000/01 to 2016/17, the tabulated data represent the outstanding stock on 31 December in the given fiscal year. For 2017/18, the data represent the average of the stocks on 31 December 2017 and 30 June 2018.

⁽³⁾ Source: Shorrocks et al. (2018). For 2000/01 to 2015/16, the tabulated data represent the outstanding stock on 31 December in the given fiscal year. For 2016/17, the data represent the average of the outstanding stocks on 31 December 2016 and 30 June 2017. For 2017/18, the data represent the stock on 30 June 2018.

⁽⁴⁾ Source: RBI 2019. The reported data include only institutional household debt.

⁽⁵⁾ The relative change in stock of financial liabilities between 2000/01 and 2001/02 in the GWD 2018 estimate appears to be unrealistic. GWD (2018) use adjusted survey means from the last two AIDIS rounds, namely 2002 and 2012, and accordingly extrapolate the data for the remaining years in the sample. It seems that the methodology employed in GWD (2018) is producing spurious backcasts for years before 2002.

Figure 12: Deviation in estimates of the stock of total debt owed by households**Table 5: Estimates of household debt owed to banks and cooperatives**

(INR billion)

	1981	1991	2002	2012
AIDIS estimate⁽¹⁾	47	182	905	7852
HBS (author's estimate)⁽²⁾	145	791	2 964	20 401
Rajkumar et al. 2019 estimate	135	684	3 310	19 926

⁽¹⁾ Source: AIDIS rounds. The AIDIS data represent the value of stock on 30 June of the corresponding year.⁽²⁾ Source: author's calculations. The data represent the value of stock on 31 March of the corresponding year.

However, in my opinion the supply-side household credit data collected from various banks and financial institutions are more reliable than AIDIS data collected directly from households, after correcting for the 'true' size of the household sector. It is justified to assume that debt is recorded accurately in the accounting books of banks and other financial institutions, whereas the outstanding stock of debt reported by households at the time of a survey might be affected by problems linked to memory lapses, whereby the respondent may fail to recollect the exact details (such as interest rates and maturity periods) of outstanding loans which were raised a long time in the past. Even if these details are known, the correct calculation of the principal and interest outstanding is needed from either the respondent or the surveyor to arrive at a reliable estimate of the total debt outstanding; banks are well-versed in such calculations. Moreover, it is important to keep in mind that debt is a sensitive issue in Indian society, and therefore it is highly possible that some households might be under-reporting their levels of debt owed and causing a downward bias in the AIDIS estimates (Shorrocks et al. (2019)).

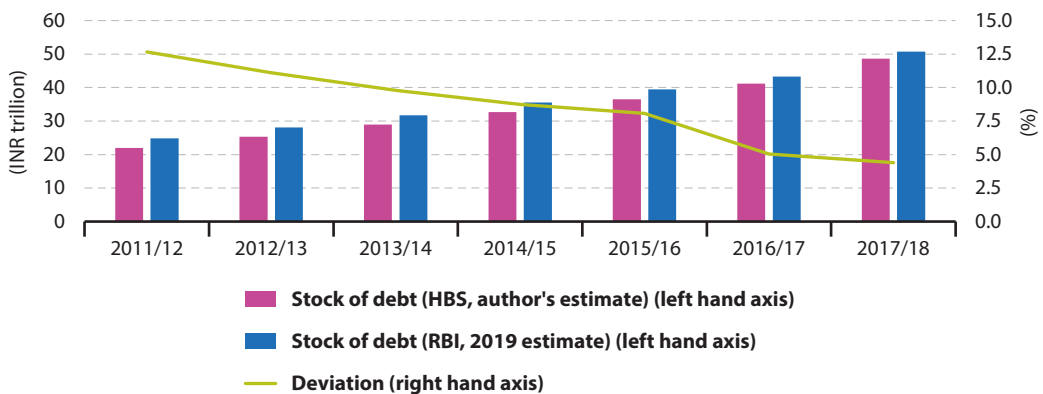
The estimates of household debt provided by Credit Suisse's *Global Wealth Databook* (GWD) include both institutional and non-institutional debt and are derived from AIDIS. Notice that GWD (2018) estimates are fairly close to the AIDIS estimate of household debt (see Figure 12 and Table 4). GWD (2019) estimates of household debt, on the other hand, diverge significantly between 2000/01 and 2010/11 (by -40 % on an average) from HBS estimates presented here. However, in recent times the estimates from GWD (2019) have come quite close to HBS estimates presented here (see Figure 12 and Table 4), with the average deviation between the two estimates falling to -3 % for the 2011/12 to 2017/18 period. The primary methodological difference between the two editions of GWD is that GWD (2018) uses raw AIDIS data on

household debt (Shorrocks et al. (2018)), whereas GWD (2019) acknowledges the under-reporting of household debt in AIDIS and corrects the survey estimate of the debt level in India using a consensus factor from the literature (Shorrocks et al. (2019)).

The stock data released along with RBI (2019) and as reported in Table 4 include only institutional household debt and therefore cannot be directly compared with HBS figures (reported in Table 4) which include both institutional and non-institutional debt. Figure 13 shows that RBI (2019) overestimates the stock of institutional financial liabilities over the period from 2011/12 to 2017/18, although to a lesser extent than was observed with financial assets in Figure 10; the deviation between the two estimates narrows over time. The reason for this overestimation of debt appears to be similar to that inferred for financial assets — the omission of OCVAs from flow data that have been used to compile stock data reported along with RBI (2019). If I compare the RBI (2019) flow data with those made available in NAS and RBI (2017a), I observe that the changes in financial liabilities reported in RBI (2019) are most of the time ⁽²¹⁾ higher than those reported in the other two sources (see Table 6).

In the context of debt, the most relevant OCVAs are debt write-offs and write-downs which involve unilateral reductions by a creditor of the amount owed to it ⁽²²⁾. This usually occurs when a creditor concludes that a debt obligation has no or a reduced value because the entire or a part of the debt shall not be paid back because of debtor's insolvency or other reasons. In such cases, the financial asset is removed from the balance sheet of the creditor and the corresponding liability is removed from the balance sheet of the debtor through other changes in the volume of assets (IMF (2014)). OCVAs corresponding to debt write-offs or write-downs will tend to diminish the debtor's gross and net debt stock (IMF (2011)). This leads me to suspect that the OCVAs have not been incorporated into the change in financial liabilities data reported with RBI (2019) and this omission appears to be the primary reason for the overestimation of institutional household debt by RBI (2019) as observed in Figure 13.

Figure 13: Estimates of household debt from institutional sources



⁽²¹⁾ A possible reason for the reversal in the sign of deviation could be the revision of estimates in national accounts at the time of publication.

⁽²²⁾ Note that these are different from debt forgiveness, which is a mutual agreement, and is therefore classified as a transaction.

Table 6: Estimates of changes in institutional financial liabilities
(INR billion)

	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18
RBI 2019	3 331	3 606	3 788	3 942	3 806	7 450
RBI 2017a ⁽¹⁾		3 587	3 634	4 318		
NAS ⁽²⁾	3 304	3 587	3 768	3 854	4 686	7 406

⁽¹⁾ 2014/15: second revised estimate at the time of publication. 2015/16: first revised estimate at the time of publication.

⁽²⁾ 2015/16: third revised estimate at the time of publication. 2016/17: second revised estimate at the time of publication. 2017/18: first revised estimate at the time of publication.

4. An international comparative analysis: broad trends in household finance

In this article, I have not attempted to estimate the stock of non-financial assets owned by households, mainly due to the paucity of reliable flow/stock data for physical assets. However, the AIDIS data contain an estimate for non-financial assets as well. Based on the latest AIDIS data, RBI (2017b) notes that Indian households have, on average, 77 % of their total assets in real estate, 11 % in gold, 7 % in other durable goods, and just 5 % in financial assets. Davies and Shorrocks (2000, p. 630) notice that wealth surveys, in general, do remarkably well for owner-occupied housing — which, as pointed out above, is the main component of non-financial assets in the case of India. Therefore, I accept the estimates of non-financial assets reported in GWD (2019) — which are based on AIDIS data — as fairly reliable and use the same source for the international comparative analyses presented below, by combining the estimates for non-financial assets reported in Shorrocks et al. (2019) with the estimates of financial assets and total liabilities compiled in this paper.

There appears to be a link between the household debt level (as a proportion of gross wealth) and the stage of economic development. As depicted in Figure 14, household liabilities amount to 8 % of gross wealth in India, 6 % in China, and 10 % in Africa, but average more than 13 % in both North America and Europe ⁽²³⁾. The world average is close to 12 %. Globally, wealth is mainly held in financial assets (55 %), rather than non-financial assets (45 %). In India, non-financial assets form the bulk of household wealth, whereas financial assets account for only 20 % (see Figure 14). This dominance of non-financial assets in India is anomalous even when I take into account global patterns based on the stage of economic development. In general, developing regions such as Latin America or Africa hold around 45 % of their wealth in financial assets. By contrast, in developed regions, such as North America, financial assets are relatively more important and constitute 72 % of gross assets. Chinese households maintain 56 % of their gross wealth in the form of financial assets. If I look at the country-level composition of wealth, it becomes clear that there exists a direct relationship between the share of financial assets in gross household wealth and the level of financial development in a country (as measured by the IMF's financial development index). Figure 15 plots the 3-year average ⁽²⁴⁾ ratio of financial to non-financial assets against the 3-year average (2015 to 2017)

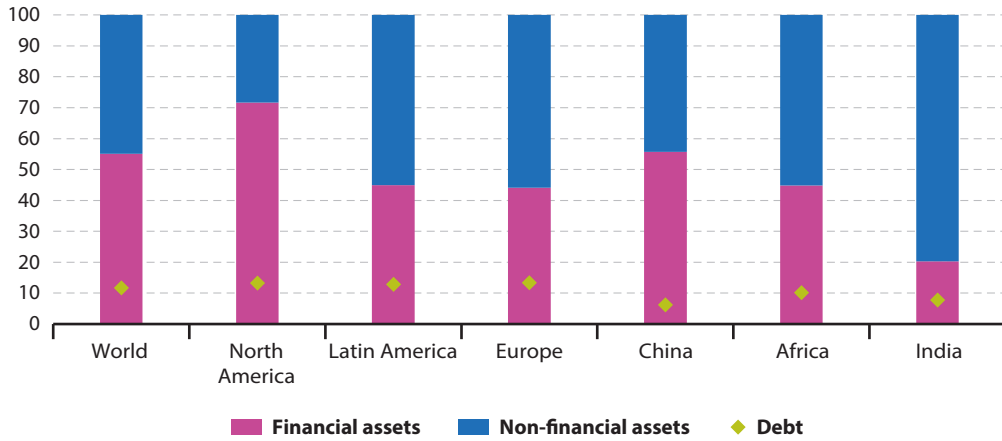
⁽²³⁾ In this paper, countries are classified into geographical regions in accordance with the geoscheme of the United Nations Statistical Division.

⁽²⁴⁾ For the financial to non-financial assets ratio, the annual average from 2015 to 2018 is taken for all countries, except for India for which the average from 2015/16 to 2017/18 is considered.

of the financial development index value for 27 selected countries ⁽²⁵⁾. When compared with other countries with similar financial development scores, such as Indonesia and Mexico, the share of non-financial assets in household wealth appears to be disproportionately high in India. However, there do exist other countries within the sample, such as Sri Lanka, Iran, Turkey, and Thailand, which exhibit similar disproportionately high shares of non-financial assets in relation to their respective levels of financial development as is the case in India.

Figure 14: Components of household wealth by region, 2017/18

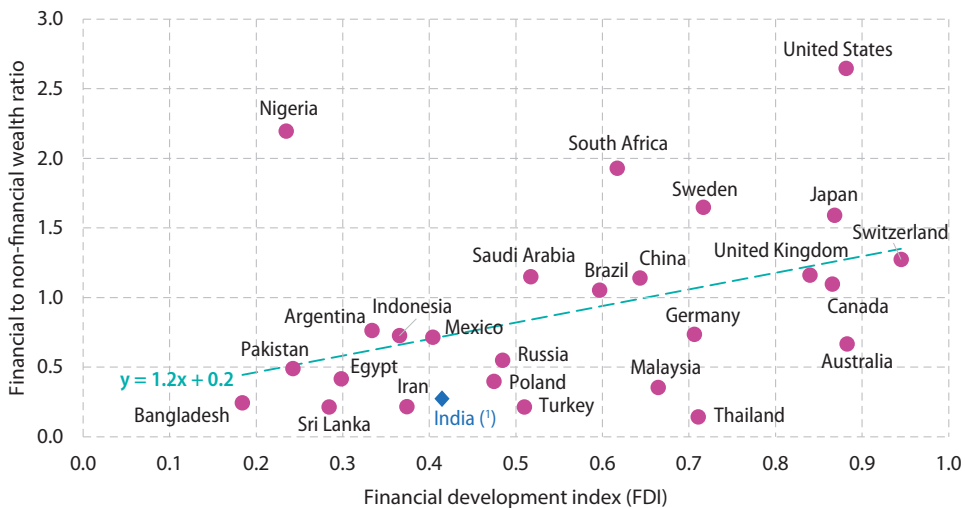
(% of gross wealth)



Source: author's calculations and Shorrocks et al. (2019)

Figure 15: Financial to non-financial assets ratio and the financial development index, average 2015-2017

(%)



(!) Average 2015/16 to 2017/18.

Source: author's calculations, IMF database and Shorrocks et al. (2019)

⁽²⁵⁾ The sample represents a mix of leading OECD economies, neighbouring states and emerging economies whose performances are often presented as benchmarks in the Indian economic discourse.

Figure 16 shows the breakdown ⁽²⁶⁾ of gross financial wealth into three categories: liquid assets (currency and deposits); equities (defined as all shares and other equities directly owned by households) ⁽²⁷⁾; and other financial assets (which mainly comprise the reserves of life insurance companies and pension funds). Unlike Chinese households — which have redistributed their portfolio away from liquid assets and towards equities over time, Indian households have maintained their preference for liquid assets. Based on 2018 data, households across the world hold an average of 37 % of their financial wealth in liquid assets, 29 % in equities, and 34 % in other financial assets (Shorrocks et al. (2019)). The share of liquid assets is higher than the world average in all of the emerging ⁽²⁸⁾ market countries in the sample (corresponding to Figure 16). The relatively low share of directly held equities and the higher share of other financial assets observed in India is partly related to a systematic negative association between equities and other financial assets observed across countries in Figure 16 and also reported by Shorrocks et al. (2019). Shorrocks et al. (2019) argue that, in countries where private pension systems are highly developed, other financial assets may tend to crowd out privately-held equities. However, in my opinion, a unique set of factors might be at play in India: the bulk of Indian household savings categorised as other financial assets go primarily into provident & pension funds (48 % of all other financial assets in 2017/18), life insurance funds (40 %) and small savings schemes. Notably, all of these investment vehicles come with income tax exemptions and the majority have either implicit or explicit sovereign guarantees. The two preferred investment instruments of the Indian middle class — the Public Provident Fund (PPF) and small savings schemes — come with income tax exemptions, a guaranteed interest rate and sovereign guarantee (net collections received in the PPF and the National Small Savings Fund are either invested in debt securities issued by central and state governments or lent to public agencies like the Food Corporation of India and the National Highways Authority of India, among others). Moreover, every salaried employee in India is obliged to contribute a defined portion of his income to an Employee Provident Fund (EPF) and a matching sum is contributed to the fund by the employer as well. These contributions are exempt from income tax just like 401(k) in the United States. However, the EPF is different from 401(k) in terms of the flexibility of choosing an investment fund that comes with 401(k) and the variability in returns that results from the performance of the chosen fund; the EPF is managed by a public agency — the Employees' Provident Fund Organisation (EPFO) — that offers a guaranteed return and is perceived as an extension of the government. In the case of insurance funds, the market is dominated by the Life Insurance Corporation of India (LIC), a state-owned corporation with more than 70 % market share. Premiums paid towards life insurance policies enjoy income tax exemption up to a certain limit but, more importantly, investing in LIC policies is considered safe due to an explicit sovereign guarantee provided to policyholders under Section 37 of the Life Insurance Corporation (LIC) Act passed in 1956 by the Indian parliament. Apparently, these sovereign guarantees, tax exemptions, and fixed rates of return offered by most of the investment instruments making up the other financial assets category might be responsible for crowding out direct ownership of shares & debentures by Indian households, alongside other factors such as a preference towards currency and deposits.

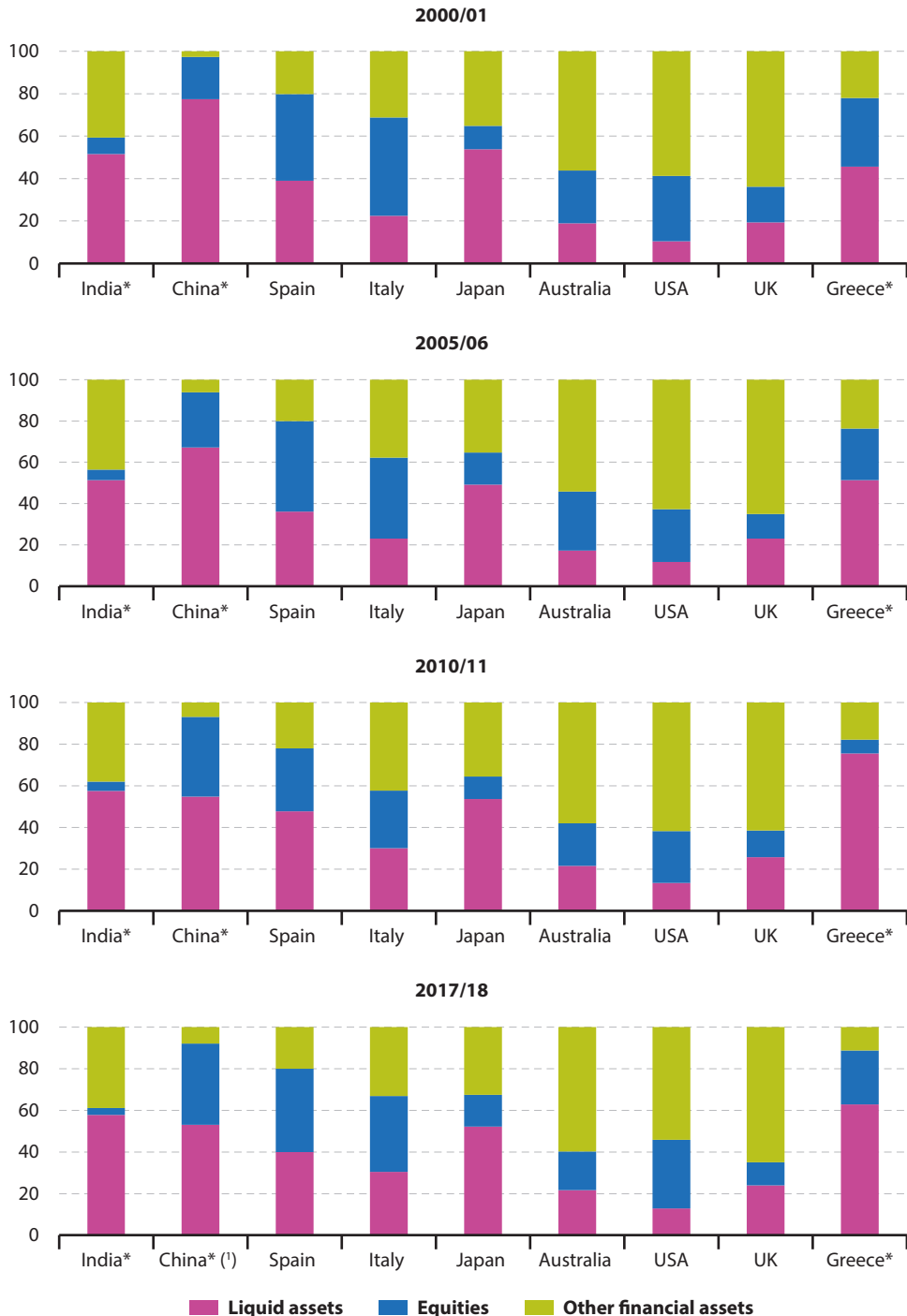
⁽²⁶⁾ For the cross-country analyses presented in this section, I have attempted to create a fairly balanced mix of leading advanced and emerging economies. However, the reader might notice that the final set of countries chosen varies from analysis to analysis (refer to Figures 16, 18 and 19) primarily due to restrictions imposed by data availability, such as, a lack of data on the share of liquid assets in the HBS of certain countries, and so on.

⁽²⁷⁾ Due to data restrictions, equities in the case of India include shares & debentures directly held by households.

⁽²⁸⁾ In this article, economies are designated as 'emerging' based on the list of countries included in the MSCI Emerging Markets Index (April 2020 composition).

Figure 16: Evolution of the composition of gross financial wealth

(%)



Note: * implies emerging economy.

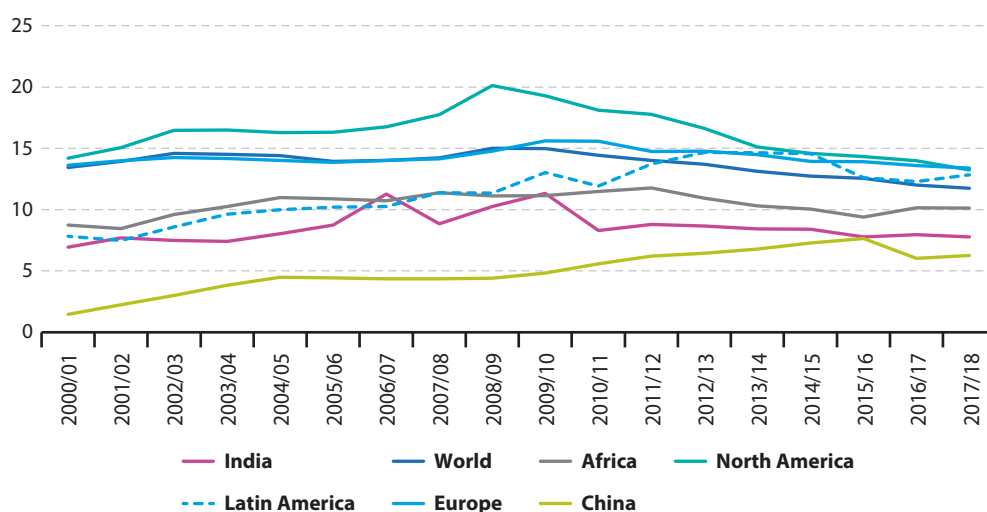
(†) 2015. Estimate.

Source: author's calculations, OECD.Stat and Shorrocks et al. (2019)

Next, I scrutinise the financial soundness of the Indian household sector compared with households in other regions of the world. The first indicator I employ is the liabilities-to-assets ratio, which is a useful metric in gauging solvency risks embedded in any sector. As depicted in Figure 17, household debt (as a percentage of gross wealth) in India remained below 10 % over the whole sample period, barring the spike observed during the build-up and peak of the global financial and economic crisis (GFEC). For the world as a whole, the household debt-to-wealth ratio rose from 13.4 % in 2000/01 to 15 % in 2008/09 and then subsequently declined to 11.8 % in 2017/18. Although the debt-to-wealth ratio for Indian households remained at modest levels compared with the world average during the sample period, household debt (as a percentage of gross wealth) had not returned to pre-crisis levels even 10 years after the GFEC, and remained at somewhat relatively elevated levels. A similar development has been observed even more clearly in other developing regions like Africa and Latin America. The household debt-to-wealth ratio in Africa rose from 8.7 % in 2000/01 to reach a peak of 11.8 % in 2011/12 and thereafter settled near 10 % in recent years. Similarly, the household debt-to-wealth ratio in Latin America rose from 8.7 % in 2000/01 to reach a peak of 11.8 % in 2011/12 and thereafter settled near 10 % in recent years. The most striking rise in household debt (as a percentage of gross wealth) has been noted in China (see Figure 17), where it more than quadrupled from 1.4 % in 2000/01 to 6.3 % in 2017/18, with a peak of 7.6 % coinciding with the 2015/16 stock market crash in China. Although households in developed regions such as North America and Europe have, in general, higher debt-to-wealth ratios than their counterparts in developing regions, households in the former seem to have undergone a course of deleveraging post-GFEC and have repaired their balance sheets to pre-crisis levels. In North America, household debt (as a percentage of gross wealth) rose from 14.2 % in 2000/01 to 20.1 % in 2008/09, but then came down to 13.2 % in 2017/18. A similar inverted V-shaped path was charted by European households over the sample period, as observed in Figure 17.

Figure 17: Evolution of household debt as a proportion of gross wealth

(%)



Source: author's calculations and Shorrocks et al. (2019)

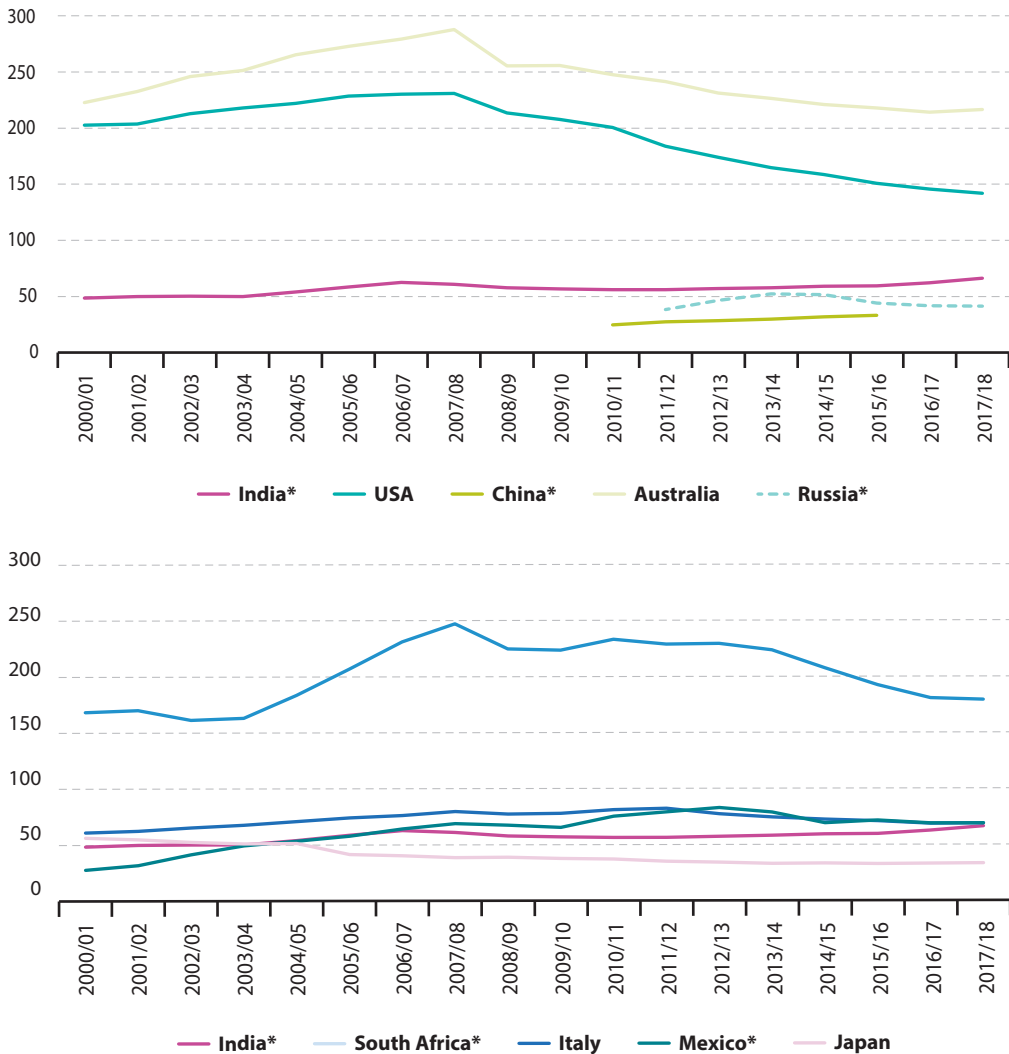
A liquidity-adjusted measure of risk associated with household debt is the liabilities-to-liquid assets ratio, which addresses the concerns over the liquidity of non-financial and some other financial assets as discussed in Section 3.2. As depicted in Figure 18, the outstanding stock of household debt has remained lower than the stock of liquid assets in all the emerging economies in the sample (corresponding to Figure 18), except in South Africa. The debt-to-liquid assets ratio for Indian households rose from 49 % in 2000/01 to 66 % in 2017/18. This increasing leverage ratio certainly poses a threat to household debt sustainability in India, but the threat is relatively tame when compared with the situation in China, where the ratio has exploded from 5 % in 2000/01 to 33 % in 2017/18. China has witnessed a phase of rapid financial development in the last two decades — its financial development index score rose from 0.43 to 0.64 between 2000 and 2017. Individuals and families have gained greater access to credit, especially to consumption credit and mortgage loans (amidst the boom in the Chinese housing market). Moreover, the share of liquid assets in gross financial wealth held by Chinese households has fallen from 77 % in 2000 to 53 % in 2015, as depicted in Figure 16. Indian households, on the other hand, have increased their allocation in liquid assets modestly from 52 % in 2000 to 58 % in 2017 and the development level of the financial sector has stagnated over the sample period — India started with a financial development index score of 0.42 in 2000 and ended up with the same score in 2017. These factors have probably prevented a China-like eruption of household debt in India. Moving to households in other emerging economies, such as South Africa and Greece ⁽²⁹⁾, I observe these countries charting an inverted V-shaped path, peaking in 2007/08 (around the time of the GFEC) and 2012/13 (during the euro area sovereign debt crisis) respectively, and ending up at a higher household debt-to-liquid assets ratio at the end of the sample period in comparison with their 2000/01 ratios. Developed economies like Australia, Italy and the United States also follow an inverted V-shaped path, peaking in 2007/08, but households in these economies consolidated their balance sheet post-GFEC and ended up with a lower debt-to-liquid assets ratio in 2017/18 as compared with the numbers they had at the beginning of this century. Over the same period, Japanese households significantly deleveraged by steadily reducing their debt-to-liquid assets ratio from 56 % in 2000/01 to 33 % in 2017/18.

Last but not the least, I look at the development of the household debt-to-income ratio — a critical measure of the debt servicing capacity of households. Unlike advanced economies such as the United States and Japan, the outstanding stock of total household debt for all the emerging economies in the sample (corresponding to Figure 19), including India, remained smaller than the flow of household disposable income (HDI) in any year of the sample period (see Figure 19). The debt-to-HDI ratio in India more than doubled from 26 % in 2000/01 to 54 % in 2017/18, indicating a dire collapse in the debt servicing capacity of Indian households. An even more frightening trend emerges in other emerging economies such as China and Mexico. The household debt-to-income ratio snowballed from 9 % to 56 % between 2000/01 and 2016/17 in China, whereas in Mexico it swelled from 7 % in 2003/04 to 38 % in 2017/18. However, this was not a phenomenon repeated across all emerging economies; rather, households in South Africa and Brazil charted a path that was similar to their counterparts in the United States, where the liabilities-to-income ratio reached a peak during the GFEC and then eventually fell back to pre-crisis levels. In my sample, Japan was the only country that systematically reduced its household debt-to-income ratio in the last two decades.

⁽²⁹⁾ It might come as a surprise to some readers that Greece is classified as an 'emerging' market, after it lost its status as a 'developed' market in 2013.

A general pattern of convergence is visible in Figure 19, which shows that households in countries where the debt-to-HDI ratio was high (close to or greater than 100 %) at the beginning of the century have improved their debt-servicing capacity. By contrast, countries that entered this century with a low household debt-to-income ratio have seen a general increase in the ratio and have come closer to the former set of countries in recent years.

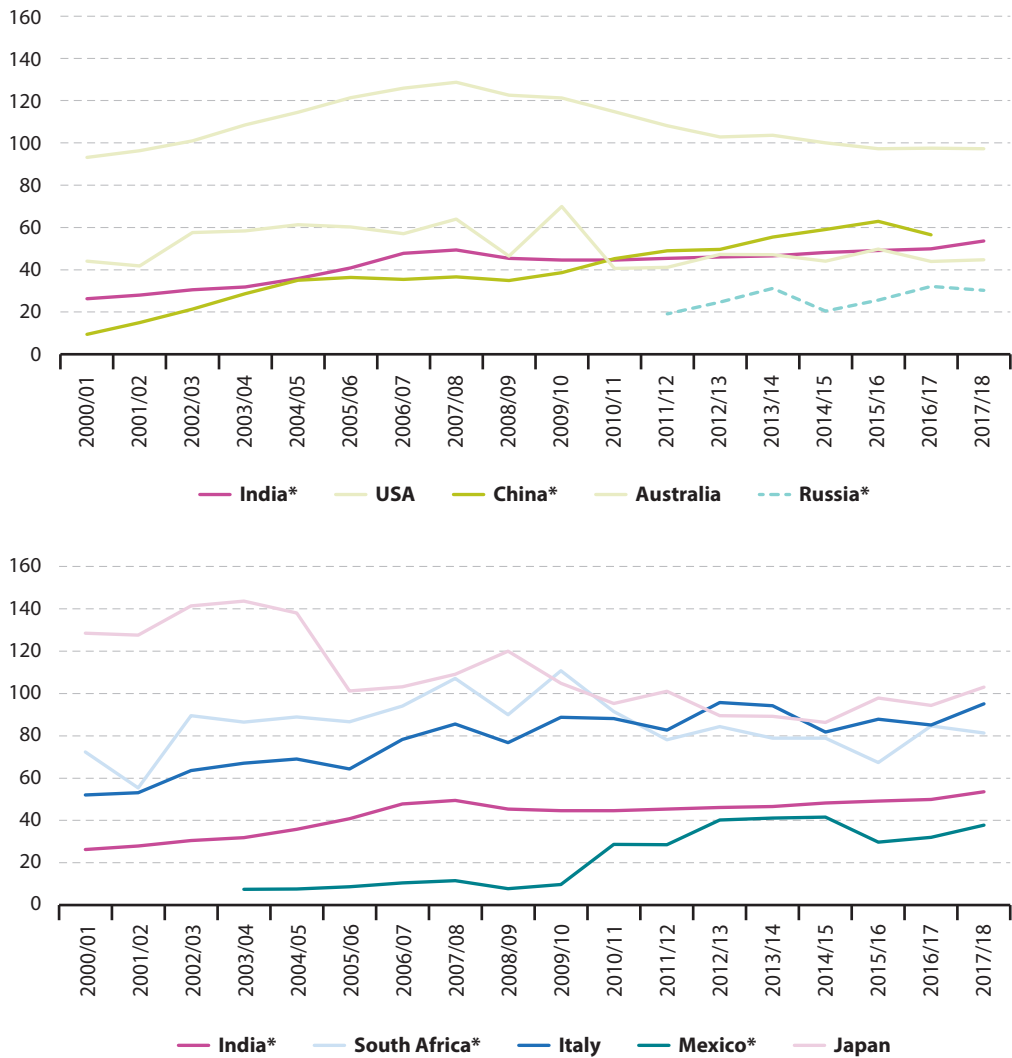
Figure 18: Evolution of household debt as a proportion of liquid assets (%)



Note: * implies emerging economy.

Source: author's calculations, OECD.Stat and Shorrocks et al. (2019)

Figure 19: Evolution of household debt as a proportion of household disposable income (%)



Note: * implies emerging economy.

Source: author's calculations, OECD.Stat and Shorrocks et al. (2019)

5. Concluding remarks

In this paper, I compile India's HBS starting from 1970/71 to 2017/18 and use the compiled HBS to study the accumulation of financial wealth by the Indian household sector. Specifically, I begin by addressing some technical issues faced during the compilation, including the definition of the 'household sector' and an assumption regarding the initial value of stocks. Next, I cumulate the FoF data to arrive at the final HBS presented in this paper. I study the development of the structure of Indian household finance over time and the constitution of asset classes making up the assets, as well as the liability side of the balance sheet. Contrary to everyday reporting by the financial press and also to the claim made by RBI (2019), I find that there has been neither a systematic decline in the share of currency and deposits (combined) in the portfolios of Indian households nor a shift towards equities and debt securities over time. Interestingly, I found that the share lost by the asset class of shares & debentures between 1990/91 and 2017/18 was commensurate with the share gained by life insurance funds over the same period. This certainly makes the calculation of total exposure of the household sector to market risk difficult, as the reserves of life insurance funds comprise equities, bonds, and other financial instruments that are in effect being held on behalf of the household sector, but at the same time households also enjoy an explicit sovereign guarantee on their rights to many insurance benefits. More specifically, I find that, in comparison with other parts of the world, households in India hold a very small share of their financial wealth in the form of directly-held equities and debt securities. I observe that this is mainly due to the general preference of Indian households for liquid assets (currency and deposits) and unique benefits, such as implicit or explicit sovereign guarantees, tax exemptions, and fixed rates of return that come with investments in provident & pension funds, life insurance funds and small savings schemes.

I also compare my estimates of the HBS with other estimates for India existing in the literature. Concurring with many other studies on Indian household finance, I too find that AIDIS grossly underestimates the level of indebtedness in the household sector. Moreover, the recently released HBS estimate (2011/12 to 2017/18) along with RBI (2019) seems to have been compiled from FoF data that do not include OCVAs and hence overestimate both the stock of financial assets owned by Indian households and the institutional household debt.

I observe that the Indian household sector, after witnessing a decade of balance sheet consolidation in the 1990s, has been in an upward leveraging cycle since the period building up to the GFEC and that critical leverage ratios have remained at higher levels since then. Strikingly, this is not a global phenomenon; the household sector in many major economies of the world repaired their balance sheets post-GFEC and ended up with either lower or similar levels of leverage ratios (debt-to-liquid assets or debt-to-income ratio) in 2017/18 as compared with the ratios they had at the beginning of this century, whereas indebtedness in the Indian household sector continued to increase, even during the economic recovery post-2010. If I just look at the household debt-to-wealth ratio, India appears to be among the least leveraged countries in the world. However, the picture becomes clearer when I look at the household debt-to-disposable income ratio — a critical measure of household debt-servicing capacity — which has more than doubled in India between 2000/01 and 2017/18. Alarmingly, this steady increase in the debt-to-income ratio might be related to the phenomenon of households

accumulating debt for supporting consumption — which has grown at a faster pace than disposable income every year since 2010/11. Subramanian and Felman (2019) highlight the *Four balance sheet* challenge being faced by the Indian economy in recent years: the stress in infrastructure companies and the real estate sector on the corporate side, and banks and NBFCs on the financial side. However, my findings reveal that there is a neglected, but equally important, fifth dimension to the country's deteriorating balance sheet — the HBS — which is the backbone of domestic consumption. More precisely, a *Five balance sheet* challenge lies before India's growth juggernaut.

On the policy front, the increasing indebtedness of Indian households requires the utmost attention. If ignored, it could catch the central bank by surprise since the data on total household debt are currently disseminated with a less-than-desirable frequency and are not monitored adequately. This holds serious implications for the formulation and efficacy of monetary policy. In view of that, synchronised efforts of the CSO, the RBI and academia are urgently needed to collect more and better data for compiling HBS, both at the national and state levels, and to implement other recommendations of NSC (2018). As a first step, an HBS should be published regularly and in a timely manner with the yearly statement of *National Account Statistics*. Data on household accounts are critical to identify and understand sectoral linkages and spillover mechanisms, and to map the transmission of potential shocks across sectors. Therefore, higher frequency information on household finance becomes crucial for a more comprehensive macroeconomic assessment. Subsequently, the release of data on household finance should be made a quarterly exercise, at least at the national level. In this regard, it is important for the reader to know that the RBI published quarterly data on changes in household financial assets and liabilities for the first time in RBI (2018) for the period 2015/16 quarter 1 to 2017/18 quarter 2, but did not continue the practice thereafter (as observed at the time of writing).

Finally, I would like to offer some caveats and suggestions which have direct implications for future avenues of research. Due to the paucity of information, NPISHs such as political parties, religious societies, trade unions, social, cultural and recreational sports clubs, and so on are included within the household sector in the national accounts of India. Consequently, estimates for HBS presented in this paper are not corrected for the 'true' size of the household sector. Also due to data constraints, I made a zero-value assumption for the initial value of the stock while cumulating the FoF data for some items in the balance sheet. Moreover, in Section 4 where I compared the financial position of Indian households with those in other countries, I used the 2019 edition of *Credit Suisse Global Wealth Databook* (Shorrocks et al. (2019)) as the source of data on household accounts for other countries. There do, however, exist other accounts of household wealth, especially for countries with questionable national accounting practices like China or Russia, which might differ from the estimates presented in GWD 2019; for example, the reader may notice that the estimates of household wealth for China presented in GWD 2019 are significantly different from those presented in Li (2018). In light of the above assumptions, the reader should focus more on the trends in the levels than the value of the level itself, especially while interpreting the findings based on international comparative analysis. The Indian HBS compiled in this article should be considered as an indicator of the order of magnitude of financial wealth allocated across different instruments and certainly not as a substitute to official balance sheet estimates⁽³⁰⁾, which, unfortunately, remain unavailable (or unreliable) at the time of writing.

⁽³⁰⁾ The estimates of household financial wealth and institutional debt presented in this paper are certainly superior to survey estimates provided by the last four rounds of AIDIS.

The HBS compiled here could be further extended to include non-financial assets, namely real estate, precious metals and household durables. At the moment, the only primary source of data for non-financial wealth owned by Indian households is the AIDIS and the subsequent estimates derived from various rounds of it, such as Shorrocks et al. (2019). These estimates are not entirely perfect and could be further improved. The biggest hurdle in this exercise is a lack of land records⁽³¹⁾ and the rampant usage of *black money* in the real estate and gold markets. It is important to note that any exclusion of non-financial assets from the HBS is a serious limitation as the Indian household sector holds at least 80 % of its wealth in non-financial assets. On the liabilities side, it is important to realise that the share of non-institutional credit is very much significant in Indian household debt as uninsured households borrow frequently from non-institutional sources (RBI (2017b)). I have tried to incorporate the non-institutional debt in the HBS estimates presented in this paper by making some simplifying assumptions.

Although I observed a steady rise in household debt-to-income ratio in India, the Indian household sector, as a whole, emerges financially sound from the perspective of the debt-to-wealth ratio (an important measure of solvency risk), and the possibility of widespread household defaults does not seem imminent or likely. However, distributive aspects of household wealth and debt (and the associated risks) across regions and income cohorts are required to get a clearer picture of the extent of debt distress in the sector. Aggregated data, as provided by the HBS, could aid in the investigation of such distributive aspects. Currently, AIDIS data are the only official source for measuring wealth distribution/inequality in India. However, there has been some debate over the reliability of wealth inequality estimates derived from AIDIS (Himanshu (2019)). Wealth surveys are generally subject to misreporting (generally under-reporting) of assets and debts by respondents, and differential response according to income or wealth level. To this end, HBS data could be used to adjust and improve AIDIS data and correct for non-sampling errors. There are different approaches to this: one, which has been applied by Wolff (2017) on United States data, is to align the survey data with the HBS data, which means adjusting all holdings of a particular kind of asset or debt in the survey data in order to make its aggregate the same as is observed in the HBS data. Notice that this approach essentially ignores the differential response problem and maintains the hypothesis that all respondents misreport by the same percentage, irrespective of their wealth level. The second approach relies on the assumption that, irrespective of the source of the shortfall (of survey aggregates below HBS totals), the error due to under-reporting mainly affects mean wealth in the top X % of the distribution; this is followed by fitting a Pareto distribution⁽³²⁾ to the upper tail that is consistent with the adjusted survey estimate of the mean wealth of the top X % of households (Blanchet et al. (2017)). It is hoped that my preliminary work would pave the way for further refinements aimed at improving the reliability of the relevant data, concepts, and estimation methods.

⁽³¹⁾ The reader may note that land administration and record-keeping is a state subject in India as per the Seventh Schedule to the Constitution of India. It is a well-known fact that the land records in India are unclear (especially the cadastral maps) and do not guarantee ownership. In India, there is a system of registered sale deeds and not land titles. Consequently, property and land ownership is established through multiple documents maintained by different departments of state government (which usually work independently of each other), making it cumbersome to collate and access them. For the aforementioned reasons, central government agencies like the CSO depend on household surveys, such as AIDIS, for estimating the contribution of land and dwellings to household wealth.

⁽³²⁾ A Pareto distribution often approximates the top tail of wealth distribution quite well (Davies (2008) p. 412).

Appendix

The household balance sheet for India covering the period from 1970/71 to 2017/18 can be accessed at the following data repository: <https://doi.org/10.7910/DVN/DPQPJY>.

References

- Bell, C. (1990), 'Interactions between Institutional and Informal Credit Agencies in Rural India', *The World Bank Economic Review*, Volume 4, No. 3, pp. 297-327.
- Blanchet, T., J. Fournier and T. Piketty (2017), *Generalized Pareto Curves: Theory and Applications*, WID.world Working Paper Series, No. 2017/3, World Inequality Database.
- Chavan, P. (2012), 'Debt of Rural Households in India: A Note on the All-India Debt and Investment Survey', *Review of Agrarian Studies*, Volume 2, No. 1, pp. 151-161.
- CSO (2012), *National Accounts Statistics: Sources and Methods 2012*, Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India, New Delhi.
- Davies, J. B. (Ed.) (2008), *Personal Wealth from a Global Perspective*, Oxford University Press, Oxford.
- Davies, J. B. and A. F. Shorrocks (2000), 'The distribution of wealth', *Handbook of Income Distribution*, Volume 1, pp. 605-675.
- Davies, J. B., R. Lluberás and A. F. Shorrocks (2017), 'Estimating the Level and Distribution of Global Wealth, 2000-2014', *Review of Income and Wealth*, Series 63, No. 4, pp. 731-759.
- Gothoskar, S. P. (1988), 'On Some Estimates of Rural Indebtedness', *Reserve Bank of India Occasional Papers*, Volume 9, Issue 4, pp. 299-325.
- Himanshu (2019), *Inequality in India: A review of levels and trends*, WIDER Working Paper 2019/42, UNU-WIDER, Helsinki.
- IMF (2011), *Public Sector Debt Statistics: Guide for Compilers and Users*, International Monetary Fund, Washington D.C.
- IMF (2014), *Government Finance Statistics Manual 2014*, International Monetary Fund, Washington D.C.
- Li, C. (2018), 'China's household balance sheet: Accounting issues, wealth accumulation, and risk diagnosis', *China Economic Review*, Volume 51, pp. 97-112.
- Moore, T. (2007), *India's emerging financial market: A flow of Funds model*, Routledge Studies in the Growth Economies of Asia, Routledge, London.
- NSC (2018), *Report of the Committee on Financial Sector Statistics*, National Statistical Commission, Government of India, New Delhi.

NSSO (2013), *Household Indebtedness in India, NSS 70th Round (January-December 2013)*, National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India, New Delhi.

Piketty, T., L. Yang and G. Zucman (2019), 'Capital Accumulation, Private Property, and Rising Inequality in China, 1978-2015', *American Economic Review*, Volume 109, No. 7, pp. 2 469-2 496.

Prabhu, K. S., A. Nadkarni and C. V. Achuthan (1988), 'Rural Credit: Mystery of the Missing Households', *Economic and Political Weekly*, Volume 23, No. 50, pp. 2 642-2 646.

Rajakumar, J. D., G. Mani, S. L. Shetty and V. M. Karmarkar (2019), 'Trends and Patterns of Household Indebtedness', *Economic & Political Weekly*, Volume 54, Issue No. 9, 54(9), pp. 41-49.

Rao, K. R. and A. K. Tripathi (2001), 'Indebtedness of Households: Changing Characteristics', *Economic & Political Weekly*, Volume 36, Issue No. 19, pp. 1 617-1 626.

RBI (2015), *Manual on the Compilation of Flow of Funds Accounts of the Indian Economy*, Reserve Bank of India, Mumbai.

RBI (2017a), *Flow of Funds Accounts of the Indian Economy: 2015-16*, Reserve Bank of India, Mumbai.

RBI (2017b), *Indian Household Finance*, Report of the Household Finance Committee, Reserve Bank of India, Mumbai.

RBI (2018), *Quarterly Estimates of Households' Financial Assets and Liabilities*, Reserve Bank of India, Mumbai.

RBI (2019), *Financial Stocks and Flows of the Indian Economy 2011-12 to 2017-18*, Reserve Bank of India, Mumbai.

Shorrocks, A. F., J. B. Davies and R. Lluberias (2018), *Global Wealth Databook 2018*, Credit Suisse Research Institute, Zurich.

Shorrocks, A. F., J. B. Davies and R. Lluberias (2019), *Global Wealth Databook 2019*, Credit Suisse Research Institute, Zurich.

Subramanian, A. and J. Felman (2019), *India's Great Slowdown: What Happened? What's the Way Out?*, CID Faculty Working Paper No. 370, Center for International Development, Harvard University.

UN (2009), *System of National Accounts 2008*, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank, United Nations, New York.

van de Ven, P. and D. Fano (eds.) (Eds.) (2017), *Understanding Financial Accounts*, OECD Publishing, Paris.

Wolff, Edward N. (2017), *A Century of Wealth in America*, The Belknap Press of Harvard University Press, Cambridge MA.

2

The concept of service(s) in recent European Union legal texts: the case of online digital content

DAMIEN BROUSSOLLE (1)

Abstract: This paper addresses adverse consequences of the European Union definition of a 'service', especially regarding the Digital Single Market Strategy. After an introduction, the second section analyses through legal texts, its somewhat outdated view, where a 'service' covers all intangible items. The third section depicts the new economic approach, based on ownership rights, almost adopted by, among others, the System of National Accounts, where intangible information products are goods. The fourth section studies typical problems, such as the tax discrimination between tangible goods and intangible information goods, mistakenly regarded as services (for example books compared with e-books). As to the provision of 'digital content', it results in a contorted contractual treatment of intangible goods. The new approach to a 'service' could solve these problems. Overall the paper advocates for more exchanges between economists and law experts.

JEL codes: L8, L86, K34, O52

Keywords: intangible good, service, online digital content, digital strategy

(1) Senior Lecturer in economics, with accreditation to supervise research (LAboratoire de Recherche en Gestion et Economie : LARGE), Université de Strasbourg, Institut d'Études Politiques, 7 Rue de l'Écarlate CS 20024, 67082 Strasbourg, damien.broussolle@unistra.fr

A tribute to T. P. Hill (1929-2017)

1. Introduction

In 2015, the European Commission (EC) launched *A Digital Single Market Strategy for Europe* (COM(2015) 192 final). It specifically concerns the production, development and sale of digital content, where digital content designates, the production, supply and processing of data (Annex B), in other words information ordered in a specialised way and supplied in digital form. Digital content refers to intangible entities, which, to paraphrase T. P. Hill, have come to play a major role in the functioning and growth of economies and need to be appropriately defined and classified (see also De Backer et al. (2015)), 'however, the concept of an intangible good has proved to be difficult and elusive, especially because of a long standing tendency to confuse intangibles with services' (Hill (2014), p. 60).

This article, starting with the case of e-books, provides an example of such confusions in the field of European Union (EU) law. It intends to demonstrate the need to clarify the EU's views on services (and services-producing activities), in accordance with the spirit of the System of National Accounts (SNA) methodology (UN (2009)). Moreover it advocates that national accountants should be aware of legal texts and debates which, to a certain degree, are addressing questions of common interest. The article shows how certain legal and economic concerns are interrelated and that each field might benefit from advances made in the other. This is all the more true since the view of services adopted since 1993 by the SNA focuses on the criterion of ownership rights, which inescapably involves legal concerns.

Obviously, information has not always been brought to users by electronic means. The coexistence of traditional and modern means of access to information that are regulated according to differing rules is nowadays inducing categorisation difficulties; the tax treatment of e-books is certainly the most visible example.

In a judgment issued on 5 March 2015, the Court of Justice of the European Union (the ECJ) reiterated the criteria defining a service and the VAT rate applicable to 'electronically supplied services' (CJEU (2015)). The Court confirmed that, under EU law digital books are indeed such a service and that, since these services may not benefit from a reduced VAT rate, they must be taxed at the standard rate. Many observers consider traditional paper and digital books as two forms of the same product, being tangible or intangible, which should not be discriminated for tax purposes (Linklater (2015))^(?). However, this common sense consideration comes up against several formal difficulties, including the prevailing definition of a service in EU legal texts.

This situation calls for an examination of the distinction currently made between goods and services by EU legal texts, the adverse consequences of this, and how they could be overcome. This topic is of major importance for more than just practical economic, law or tax

(?) Germany, France, Italy and Poland asked the EC on 19 March 2015 to bring the existing situation to an end.

reasons. It illustrates similar methodological debates to those addressed in national accounts, especially how digital entities should be regarded. Many activities representative of the new economy are producing or processing information. It is quite common to regard digitalised or information products that such activities supply as services, as endorsed by the EU. However, following Hill's (1999, 2014) illuminating views, this article, advocates that they are not, and incidentally, that knowledge-capturing products should be regarded as information goods. Moreover, using the example of EU legal texts, it also explains that avoiding to recognise that these are information goods leads to an unduly complex regulatory framework, which in turn increases legal complexity and transactions costs. In so doing, the article intends to show on the one hand, that the proper characterisation of a service is not merely a vain methodological debate (Miroudot (2019)), and on the other hand, that national accountants should more often cooperate with legal experts, as both professions can benefit from exchanging views on intangibles to clarify their methodology.

The paper is organised as follows. Section 2 examines the definition of a 'service' in key EU legal texts. It points out that the EU view owes much to the traditional economic conception, but also that the wording of the seminal rules tends to lead to some confusion. The third section recalls the revised economic approach to a 'service' promoted by Hill (1999) and shows that this could help to solve the problems relating to the categorisation of digital content between goods and services. As an example, the fourth section focuses on legal difficulties introduced into the EU's digital strategy by the present EU view of a 'service', especially regarding the VAT rate and the Digital Content Directive.

2. Services in key EU official texts

EU directives and regulations frequently employ the term 'service(s)' without defining it, as if its meaning were obvious. This is not uncommon practice; for instance, the United Nations Convention on Contracts for the International Sale of Goods (Vienna, 1980) (CISG) does not actually define what constitutes a good or, conversely, a service (UN (2010)). EU legal texts do, however, provide some elements of a characterisation, generally referring to Article 60 of the Treaty of Rome (now TFEU Article 57). This states that 'Services shall be considered to be "services" (...), in so far as they are not governed by the provisions relating to freedom of movement for goods, capital and persons' (Box 1). Although it gives no clues as to what a service actually is, this sentence defines it by default, referring to a *residual* definition as Hill (2014) puts it.

2.1. A residual notion: a service is a non-good

The Treaty of Rome defines a service by what it is not: it is an economic transaction pertaining neither to goods, nor capital, nor persons. While it is easy to understand that a service is not capital or a person, the key difference between a good and a service remains implicit and unexplained. This most probably stems from the fact that the Treaty considers a service from an international trade perspective. Consequently, it is probably influenced by the balance of payments approach, in which current accounts are traditionally split between goods and a vast conglomerate of transactions, once associated with services (*invisible trade*). In this view,

services are a complementary set to goods. For instance, the VAT Directive (Council Directive 2006/112/EC) plainly illustrates this approach: ‘Supply of “services” shall mean any transaction which does not constitute a supply of goods’ (Article 24(1)).

Since a service is defined as being opposed to a good, it is necessary to explain the characteristics of goods. Here again, there is no definition as such in the Treaty. However, from a reading of other EU legal texts, it is fairly easy to see that they are characterised by their materiality, or tangibility (see also Snell (2002)). This view is, for instance, illustrated by the VAT Directive, Article 14(1), which states that “Supply of goods” shall mean the transfer of the right to dispose of tangible property as owner’. A similar expression is employed in the proposal for a Common European Sales Law, which is detailed below (COM(2011) 635 final, Article 2(h)).

A side effect of this view is that it establishes a boundary between tangible and intangible goods, where only the former are deemed proper goods. In confirmation of this reasoning, the VAT Directive states that the supply of a service may consist in, among other things, ‘the assignment of intangible property’ (Article 25). This standpoint brings complexity and raises transactions costs.

Since a digital product is regarded as a service, an e-book is therefore treated differently from its paper counterpart. The March 2015 judgement of the ECJ (CJEU (2015)) thus duly recalls that an e-book is ‘an electronically supplied *service*’ (emphasis added). It is important to understand that this classification is not linked to the mode of supply required by digital products (commonly downloading). What determines the classification is the very fact that these products are intangible. Although the case of e-books has been much publicised, the reasoning applies to all digital products, including music, images, movies, and so on. Nevertheless, the economic literature usually regards these products as information goods (Shapiro and Varian (1998)). It should be recalled that, since digital products are encapsulating information, the System of National Accounts identifies them as *knowledge-capturing products* (UN (2009), paragraph 6.22) ⁽³⁾.

Despite its relevance being increasingly challenged, the tangibility criterion has remained established and has resisted change. A revision of the above-mentioned VAT Directive in 2009, aiming to take account of technical developments in book production, changed the expression ‘supply (...) of books’ to ‘supply (...) of books on all physical means of support’ (Council Directive 2009/47/EC amending Directive 2006/112/EC as regards reduced rates of VAT, Annex 1). Regardless of this change, the ECJ did not feel authorised, in its 2015 ruling to extend the notion of books to include digital books.

This kind of classification problem is not unique. It has appeared at least once previously, in a dispute regarding the taxation of printing activities, where the defendant argued that ‘printing is a service activity’. The ECJ stated that ‘printing works should not be characterised as services, since the direct output of a printer activity is a material entity (...)’ (CJEU (1985)). As a matter of fact, from an economic point of view only the distribution part of the publishing activity is a true services-producing activity (see Annex C).

⁽³⁾ Hill (2014) convincingly suggests that information-capturing would have been a better label, since goods can only encapsulate information, while knowledge is held by human beings. However, the main issue is that these products are goods (see later).

In focusing on the distinction between materiality and immateriality, case and EU legal texts most probably draw their inspiration from the traditional economic view. Although the standard economic approach uses a number of criteria to differentiate a service from a good, including non-storability or the necessity for the service to be provided in the customer's presence, intangibility is certainly the most prominent one. The 2004 World Investment Report illustrates this approach: 'Services are usually perceived as intangible, invisible, perishable and requiring simultaneous production and consumption, while goods are tangible, visible and storable and do not require interaction between producers and consumers' (UNCTAD (2004), p. 145). By their nature, digital products such as e-books, when considered apart from the medium on which they are supplied (in other words pure information), are obviously intangible and invisible. These observations have notable consequences for the Sale of Goods and Digital Content Directives (see Section 4).

Box 1: service(s) as defined in European Treaties

Article 60 (Treaty of Rome) / Article 50 (Treaty establishing the European Community) / Article 57 (Treaty on the Functioning of the European Union)

Services shall be considered to be 'services' within the meaning of the Treaties where they are normally provided for remuneration, in so far as they are not governed by the provisions relating to freedom of movement for goods, capital and persons.

'Services' shall in particular include:

- (a) activities of an industrial character;**
- (b) activities of a commercial character;**
- (c) activities of craftsmen;**
- (d) activities of the professions ⁽⁴⁾.**

Without prejudice to the provisions of the Chapter relating to the right of establishment, the person providing a service may, in order to do so, temporarily pursue his activity in the Member State where the service is provided, under the same conditions as are imposed by that State on its own nationals.

⁽⁴⁾ This last word is most probably a poor translation of the French '*professions libérales*', which could have been better rendered as 'professionals'.

2.2. Service products or services-producing activities?

Article 60 of the Treaty of Rome raises another concern when it states that services include, among other things: ‘activities of an industrial character; (...) activities of craftsmen (...)’ (see Box 1).

The concern stems from the assimilation between the term service(s), which in the article primarily designates an output (in other words a type of product), and the expression *services activities*, which explicitly refers to activities. There is an implicit but unwarranted shift in the rationale here. While it is true that service products are typically provided by services-producing enterprises (in other words, enterprises classified to services-producing activities), they may also be provided by enterprises classified to other economic activities, for instance enterprises providing manufacturing services (Miroudot (2019)). A service may also be embedded in or combined with a manufactured good. Nevertheless, *activities of an industrial character or of craftsmen*, which the article lists, only produce a service when they typically carry out specific tasks. For instance, repairs or installation tasks fall under services provision (see Council Implementing Regulation (EU) No 282/2011, Article 8), whereas other craftsmen’s works fall under goods production. Yet it should be obvious that craftsmen’s activities as a whole and ‘activities of an industrial character’ are not usually services-producing activities.

Including activities (economic activity approach) in a characterisation of a service (product approach), unfortunately introduces a risk of confusion into the whole meaning of Article 60.

It is worth underlining that the two perspectives are, from a methodological point of view, clearly separate. On the one hand, there are classifications of activities, such as the United Nations Statistics Division’s (UNSD) International Standard Industrial Classification (ISIC), and on the other hand, product classifications, for example the UNSD’s Central Product Classification (CPC).

The wording of the Treaty of Rome was undoubtedly a simplification. Nevertheless, its inclusion in a seminal text which establishes binding obligations has produced unforeseen consequences. The intermingling between product and economic activity tends to run through most EU legal texts. The texts sometimes regard a ‘service’ as a provided product (for example the VAT Directives), sometimes as an activity (for example the Services Directive (Directive 2006/123/EC on services in the internal market), but they all refer to Article 60, without however making explicit which aspect is addressed. In fact, an economic activity perspective should be taken whenever establishment provisions are being considered; whereas cross-border provisions call for a perspective based on a service as a product.

To avoid the risk of confusion between a service as a product and a service as an activity, it is expedient to reserve the term ‘service(s)’ for the output (the product), and ‘service activity(ies)’ or ‘services-producing activity(ies)’ for activities. This mix-up of terms for products and activities might go some way to explaining the problem relating to e-books (or even to all digital products); they might possibly be included in the service category (of products) on the grounds that the enterprises providing them are allegedly in services-producing activities. Nevertheless this might also reveal that several activities are wrongly classified (see Annex C, information and communication).

Box 2: excerpt from the VAT Directive (Council Directive 2006/112/EC on the common system of value added)

Supply of services

Article 24

1. 'Supply of services' shall mean any transaction which does not constitute a supply of goods.

(...)

Article 25

A supply of services may consist, inter alia, in one of the following transactions:

(a) the assignment of intangible property, whether or not the subject of a document establishing title;

(...)

This section has underlined some issues related to the treatment of the term 'service' in EU legal texts, raised by their growing disharmony with the information-based economy. The following section will reveal that recent economic analysis offers a way to solve most of these issues.

3. The updated economic approach to a service

New information and communication technologies, especially digitisation and internet access, have tended to render obsolete the long-established views on the definition of a service. On the one hand, since certain services can be provided remotely, the proximity requirement between consumers and providers is weakened (Mishra et al. (2012)). On the other hand, several goods, especially those called knowledge-capturing in the SNA, can be dematerialised and provided instantly to the consumer, making them appear to be services. Several authors thus consider that the distinction between a good and a service has become blurred (Hojnik (2016)), or has even lost significance (Pilat and Wölfl (2005), Jansson (2009)). However, at least in tax or contractual terms, in other words considering practical market transactions, it remains fundamental. For instance, Dreier (2013, p. 139) writes 'What is needed is (...) a coherent theory of when to treat online offerings of copyrighted works as "services", and when to still treat them as "goods", in spite of their intangible and immaterial nature'. For matters concerning national accounts and (activity and product) classifications, this issue is also of significant relevance. Fortunately, there has been a revival of the economic concept of a service, based on the work of Hill (1977, 1999, 2014), which could help to eliminate the aforementioned ambiguities ⁽⁵⁾.

⁽⁵⁾ It would also help overcoming such considerations: 'it becomes almost impossible to draw a clear line between manufacturing and service activities (...)', (Miroudot (2019), p. 22).

3.1. The ownership criterion

The updated approach does not rely on tangibility to differentiate a good from a service, but on the dissimilarity between flows and stocks. As a flow or a change (Hill (1977)), a service cannot be seized, and is thus intangible. This aspect has often been likened to immateriality, giving rise to the idea that the distinction between a good and a service relies primarily on a physical criterion. However, it is too narrow a view and too misleading a simplification. On the one hand, goods are not only tangible items; on the other hand, as regards services, immateriality is an established fact, not a defining criterion. More crucially, a service cannot be isolated from either its provider or its recipient, in other words it is not intrinsically separable. Therefore, as Hill (1999) stresses, it is not an identifiable, distinct entity over which ownership rights could be established, unlike a good, which, whether tangible or not, clearly is. This way of thinking should drive economists to pay more interest to several legal discussions.

It is important to understand the reasoning here: although a service itself cannot give rise to ownership rights, it can be applied to an object over which they can be established. For instance, repair is a process which (when applied commercially) is applied to the property of a second party (the client), the output of which is not transferable separately from the means (human or technical) employed to perform it. Ownership rights can however, be established over the repaired object. A similar perspective applies to online delivery or access, where delivery is the true services-producing activity (commonly a distributive trade), whereas what is delivered is often a good, even an intangible one (see below concerning originals and copies). Likewise, information although immaterial, fulfils the updated service definition. Consequently, activities such as data production and processing may be viewed as goods-producing activities. The information and communication section of ISIC (Rev.4) includes 'the production and distribution of information and cultural products' (UN (2008), p. 205). Several of these cultural products are, in the terms of the SNA, knowledge-capturing products, and therefore may be viewed as information goods (Annex C provides a detailed discussion of this ISIC section). Only their delivery or the provision of temporary access to them is a services-producing activity, the latter being a rental activity. It should be noted that, while the updated economic approach has been mostly endorsed by the SNA (UN (2009), paragraph 6.17) and the balance of payments (BoP) (Broussolle (2015)), the categorisation problem of knowledge-capturing-products, which are not yet fully recognised as information goods, remains.

The updated approach preserves the distinction between a good and a service and grounds it on more enlightening and effective principles. One unexpected, but decisive, consequence is that a service cannot be stolen, at least in the traditional sense that applies to a good. This is consistent with the fact that a service cannot be subject to ownership rights. For instance, theft of a transport service supposes either stealing the means of transportation (whatever it may be) or the transportation voucher, or being a stowaway. In none of these circumstances is a service itself stolen, separately, in such a way that it could be eventually

used ⁽⁶⁾). Even the illegal taking of a transportation ticket is not stealing a service, but that of a purchase instrument. Another legal consequence is that a good may be replaced by another available good, or repaired, whereas a service has to be produced anew (see Morais Carvalho (2019)). This circumstance has obviously contractual consequences in the event of a lack of conformity, as for example the Sale of Goods Directive (Directive (EU) 2019/771 on certain aspects concerning contracts for the sale of goods) shows.

To a certain extent, EU legal texts already acknowledge that ownership rights typically relate to goods. The VAT Directive (Article 14(1)) reads “Supply of goods” shall mean the *transfer of the right to dispose of tangible property as owner*’ (emphasis added); and the Consumer Rights Directive (Directive 2011/83/EU on consumer rights, Article 2(5)) states: “sales contract” means any contract under which the trader *transfers or undertakes to transfer the ownership of goods to the consumer*’ (emphasis added). In both cases, goods are associated with ownership or the transfer of rights; conversely it indicates that this is not a characteristic of a service. In that sense, Recital 29 of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society is certainly right when asserting ‘The question of exhaustion [in other words of ownership rights] does not arise in the case of services and online services in particular’. However, this critical point is not fully recognised, which has had adverse consequences.

3.2. The case for intangible information goods and originals

One adverse consequence of the EU’s conformity to the traditional view of a service is that an intangible good is treated as a service. Whereas this approach may acknowledge a traditional separation in economic activities between manufacturers (producers of vinyl records) and what seemed to be services-producing activities (providers of music to be downloaded), it is less and less workable. The updated approach in contrast promotes an integrated and more consistent treatment of both material and immaterial goods, distinct from that of services. The 2011 ECJ C-128/11 *UsedSoft* case (CJEU (2012)), actually follows a comparable reasoning, when declaring: ‘from an economic point of view, the sale of a computer program on CD-ROM or DVD, and the sale of a program by downloading from the internet are similar’ (al. 61). As such, it allows goods to be treated independently of their physical aspect (for example downloaded, or in a supporting media, or else built-in software), which is a key issue. This subsection describes and explains the updated approach to intangible goods, which are composed of two classes of goods: originals and copies.

Creative immaterial goods such as patents, software source code, architectural plans, copyrights, and so on are called ‘originals’ in the updated economic approach. An original may be embedded in a physical (material) medium, but the original itself is immaterial,

⁽⁶⁾ A service incidental to the provision of certain goods (energy provision) or information goods, such as TV shows, movies, or databases, may be indirectly stolen or hacked. Nevertheless it should be recalled that what is stolen is a bundle of a service and a good. In that bundle, the value of the delivery service is low compared with that of the good (which is delivered or to which access is granted). Above all, the delivery service is not stolen as such; it is either latent (not yet produced) or, when produced, embedded in the good that is stolen. In any case, in the particular situation of temporary access, no ownership transfer is at stake. As such, hacking a provider of temporary access to digital goods might result in the theft of a pure service (see also Section 3.2, originals and copies). This circumstance is nevertheless dubious. On the one hand, it supposes that the information goods that were accessed, are, for one reason (limited life) or another (obsolescence), ‘perishable’, or are otherwise useless to the receiver. Instead, on the other hand, the temporary access theft actually consists in the stealing of an intangible good. There is no imaginable rationale in stealing only the right to access.

because it is the result of an intellectual creation. It is important to stress that the ‘original’ is not the first physical materialisation of a plan, book, patent, piece of music or movie; the ‘original’ is the corresponding intellectual production that may be stored on a physical medium (Hill (2003)). Thus, the original is pure information, but organised in a specific way that enhances knowledge. The term ‘original’ has been chosen to highlight that it is the output of a creative process and that its value stems from the (first) result of this process. An original is a good; it possesses the features of a commodity, except for the physical aspect (Hill (1999)). Let us notice that the way intangible goods are produced, in other words by duplication, entails an important characteristic. They are non-rival goods, because if an owner wants to share or sell an information good, he/she does not need to give away his/her own original or copy (Benabou (2017)), this circumstance has contractual consequences (see below).

Originals must be distinguished from their copies for at least two reasons.

- Firstly, compared with a transaction concerning an original, a transaction concerning a copy is linked to its mode of transfer, in other words to the medium through which it is sold. When a publisher buys an original novel, they may receive a paper manuscript or a digital file; the medium does not alter its value or price. When a customer buys a copy of a work of fiction, they may choose between a paper book and an e-book (purchase of a good), or they may prefer to borrow from a library or use a temporary online access (purchase of a service); the transaction and the price are linked to the mode of provision.
- Secondly, an original is an asset, whereas a copy is an ordinary good. In other words, originals come under the category of investment goods, because the expenses their creation has incurred are expected to generate future incomes (Corrado et al. (2017)). Copies are consumption goods, whether final or intermediary. This perspective was almost adopted in the most recent SNA and BoP revisions (Broussolle (2014, 2015)).

As will be noted in the next sections, the economic and legal approaches on this topic address similar issues and can benefit mutually from each other’s perspective. An original may be duplicated as many times as necessary with no alteration of information. Copies of the original are ordinary goods, which, especially in the case of digital copies, may be viewed as equivalent to the original. However, the contract, rights and obligations relating to each of these are not alike. The type of contract determines within which category, in other words as a good or as a service, the provision of a specific good falls. When a good, whether tangible or not, is sold outright, or the right to use it sold for an unlimited period, the transaction involves a sales contract. As ECJ C-128/11 *UsedSoft* judgement recalls in paragraph 42 that ‘According to a commonly accepted definition, a “sale” is an agreement by which a person, in return for payment, transfers to another person his rights of ownership in an item of *tangible or intangible property* belonging to him’ (emphasis added). Let us here emphasise the critical link made between goods and the sales contract. When a good is made accessible by means of a contract or license (for a limited duration), or remote access (software or a file installed in the cloud), in other words when it is rented, it is a service transaction. In those circumstances the transaction falls under service or rental contracts. These considerations are essential for an accurate treatment of digital content, which an examination of the corresponding directive will illustrate (Directive (EU) 2019/770 on certain aspects concerning contracts for the supply of digital content and digital services) (see the next section).

4. Services and information goods in the EU's 2015 Digital Strategy

The traditional views regarding goods and services have unfortunate consequences on recent legal texts aiming to foster the digital economy. They create difficulties in the treatment of digital products, generating the need for special cases and legal contortions. Moreover they are of major interest for national accountants, as they provide valuable and fresh practical cases as to the underlying goods-services definition problem. Two major issues pertaining to the EU's 2015 Digital Strategy are examined below: VAT and the provision of digital content.

4.1. Adverse consequences of the current notion of a service for VAT

Under EU law, only goods and a few selected services may benefit from reduced VAT rates (Annex A). Consequently, to fulfil the aforementioned goals, a clear understanding and a precise delineation of the boundary between goods and services are needed. However, the traditional approach in the EU is becoming increasingly questionable.

A first issue relates to the categorisation of the output of utilities, such as electricity and gas. At least since the adoption of Directive 1999/44/EC on certain aspects of the sale of consumer goods and associated guarantees (Article 1(2) (b)), EU law has regarded these outputs as a service, because at first sight they appear to be intangible. It is worth observing that this apparent intangibility is not of the same nature as that of information goods, because it essentially concerns sight (invisibility), whereas the outputs of utilities are tangible by other senses, or physical dimensions, such as feeling, smell, or even hearing (see also Hill (2014)).

In any event, because several EU Member States wished to apply a reduced VAT rate to the output of utilities, an exception to EU principles was established. Hence the VAT Directive states 'Electricity, gas, heat, refrigeration and the like shall be treated as tangible property' (Article 15(1)). Similarly the Sale of Goods Directive reads 'water, gas and electricity are to be considered as goods within the meaning of this Directive where they are put up for sale in a limited volume or a set quantity' (Article 2. al. 5a.). Let us notice that the latter stipulation oddly means that the nature of those products depends on their size.

These special provisions could have been avoided. Given that these outputs can be isolated from their provider as well as their recipient, and that consequently ownership rights can be established over them, they are in fact goods. Moreover, since they are usually priced by volume, their sale should come under the sales contract (see next subsection). This would be much simpler and more effective.

The major adverse effect nevertheless stems from the dichotomous treatment of information goods, depending on whether or not information is incorporated in a medium. The fact that this view is somewhat in line with the BOP rules (IMF (2009), paragraphs 10.164 and 10.140 sq., Box 5), where information goods transactions are classified as a good or as a service depending on whether they are materially (CD, book, software ...) or immaterially (downloaded) provided, is not a helpful justification, since it appears not entirely consistent with the SNA definition of a service (Broussolle (2015)).

This split induces tax distortions (Linklater (2015)) and a legal burden for enterprises. Yet, since the medium is not the actual source of the economic value, it should not be used as a decisive characteristic. The EC is somewhat aware of this problem and has proposed two ways to overcome it. Unfortunately, both necessitate legal contortions and are unsatisfactory because they do not address the underlying goods-services definition problem. This subject also concerns the knowledge-capturing products debate.

Firstly, concerning tax distortions, in the (paper) books compared with e-books problem, the EC finally agreed in December 2018 to equalise their VAT rate. However, it was done by generating an exception specifically for e-books to the general rule. It would have been much simpler and more consistent to regard all the pseudo 'e-services', as belonging to the broader set of intangible goods. This would have prevented distortions being created against other electronically provided intangible information goods, such as music or movies. Conversely, this also suggests that, in the SNA, knowledge-capturing goods might rather be called information goods and simply classified with goods.

Secondly, the Consumer Rights Directive (Directive 2011/83/EU on consumer rights) had initially and appropriately decided to include 'tangible' digital content under the sale of goods contract, although separated from other 'intangible' information goods which were seen as services (Annex B). The consequence of this choice is what was referred to as the dichotomous approach. In 2015, the EC undertook to eliminate this fragmented situation by launching two combined directives. The first one, Directive (EU) 2019/771 on certain aspects concerning contracts for the sales of goods, otherwise known as the Sale of Goods Directive, aims at offering a unified framework for goods sold in the EU. The second, Directive (EU) 2019/770 on certain aspects concerning contracts for the supply of digital content and digital services, also known as the Digital Content and Digital Services Directive, is devoted to all digital content (see the next subsection), but mainly to the sale of services (Manko (2016)).

The effort to eliminate the dichotomous view has nonetheless generated complexities because, although the EC decided to address all digital content in the same text, unfortunately the EC chose to favour the text that was principally devoted to services to implement this approach.

Since the EU has still not changed its 'materialist' principles, the implicit reasoning is that, despite the fact that information embedded in a medium/carrier ('tangible' digital content) is a good and intangible digital content (e-publications ...) is deemed a service, they will be treated in the same directive as true e-services. This results in greater complexity because the two first entities (tangible and intangible digital content) are not actually services. Consequently not all the properties of these information goods may be covered by the Digital Content and Digital Services Directive. Several have to remain covered by the Consumer Rights Directive that is dedicated to goods, especially considering copyright matters (see Recital 20). Once more it can be observed that in the Digital Content and Digital Services Directive those provisions are not needed, which confirms that services are not subject to property rights. Indeed, when ownership rights are at stake concerning a services-producing activity (for instance a services-providing process, a license or a franchise), it is not intrinsically the service as such which is involved, but its design or a business model. The service itself has yet to be produced. In any case, the sale and use of CDs and DVDs therefore legally come under the remit of two Directives, whereas one sufficed before.

Both observed solutions add complexity. It would be simpler and more effective to regard all these products as information goods, and treat them as such.

At the beginning of the legislative process, in the very first place, to fulfil its all-encompassing goal the EC had extended the notion of digital content to cover not only intangible and tangible information *goods*, but also information *services* (Annex B — excerpts from a proposal for a Directive on certain aspects concerning contracts for the supply of digital content (COM(2015) 634 final). Thankfully the adopted Digital Content and Digital Services Directive did not follow this all-embracing path: rather, digital content is instead more tightly delimited (Annex B — excerpts from Directive (EU) 2019/770). Nonetheless several legal and categorisation problems related to contractual matters remain, which are addressed in the next subsection.

4.2. A non-sale, non-service contract for information goods?

One of the three main objectives of the EU's Digital Single Market Strategy is to ensure 'better access for consumers and businesses to online goods and services across Europe', through the adaptation of the legal framework established by the Consumer Rights Directive to rapid economic developments in online markets. It is based on the two 2019 directives — the Sale of Goods Directive and the Digital Content and Digital Services Directive — mentioned in the previous subsection. This new strategy aims at creating a harmonised EU legal framework for all types of digital content supply.

However, in doing so it comes up against the conception of the boundary between a service and a good. This makes it impracticable to give a clear answer to the key question: to what type(s) of contract(s) should the supply of digital content belong? This also hinders the understanding that, if the true nature of the goods-services distinction were recognised, it would be much easier to relate each type of digital content to the appropriate type of contract and thus the strategy would become more effective.

4.2.1. THE PARTICULAR APPROACH OF THE CONSUMER RIGHTS DIRECTIVE

Usually, depending on the economic nature of the supplied entity, the supplier–customer relationship comes under a sale, a service, or a rental contract. However, the Consumer Rights Directive decided that, following a debatable decision on supply by utilities, intangible information goods would not come under these established types of contracts (Annex B — excerpts from Directive 2011/83/EU).

Unsurprisingly, the dissimilarities in the supply of a good and a service result in significant contractual divergences (see Mak (2016); Wendehorst (2016) and following paragraphs). The Consumer Rights Directive depicts one example, in Recital 40: 'In the case of service contracts, the withdrawal period should expire after 14 days *from the conclusion of the contract*. In the case of sales contracts, the withdrawal period should expire after 14 days from the day on which the consumer or a third party other than the carrier and indicated by the consumer, *acquires physical possession of the goods*' (emphasis added). In any event, EU law could not, on the one hand, because of its principles, recognise supply by utilities or of 'immaterial'

information goods as a supply of goods; nor could it, on the other hand, because of practical incompatibilities, associate them with service contracts. Consequently, it created a particular (*sui generis*) contract, which relates neither to that of a sale, nor to that of a service. Incidentally, it can be suggested that this choice is similar in essence to that of the 2008 SNA when it was decided to label information goods with the idiosyncratic name of 'knowledge-capturing products'. Nevertheless, Recital 19 of the Consumer Rights Directive, states '(...) Similarly to contracts for the supply of water, gas or electricity (...), contracts for digital content which is not supplied on a tangible medium should be classified, for the purpose of this Directive, neither as sales contracts nor as service contracts'.

This is somewhat unexpected and rather unjustifiable. Why does the sale of the outputs of utilities, usually done by volume and which has the key characteristics of goods, not come under the sale contract? Likewise, why are intangible information goods, which are sold outright, not associated with the sale contract? This would have met the suggestion of Wendehorst (2016, p. 11) calling for: '(...) a cautious **"digitalisation" of general sales law**, i.e. in making the law on sale of goods fit for the challenges of the digital age' (emphasis already in the original text). On the other hand, why are genuine information services not linked to the appropriate rental or service contract?

Unfortunately, the EU's Digital Strategy perpetuates the same approach.

4.2.2. THE EU'S DIGITAL STRATEGY PERPETUATES THE CONTRACTUAL SINGULARITY

The Digital Content and Digital Services Directive deliberately adopts a very broad approach encompassing all information goods and e-service products⁽⁷⁾, and leaves it to EU Member States to decide whether the supply of digital content should come under sale, service, rental or particular contracts (see Recital 12). Nevertheless, according to Mak (2016) the consistent and desired solution appears to be the particular regime. While this solution formally prevents legal fragmentation among Member States, it does not overcome the doctrinal problem.

As Sénéchal (2015) has remarked about the discarded proposal for a Regulation on a Common European Sales Law (COM(2011) 635 final), the growth of digital supply does not eliminate the 'logic of possession'. Thus, it is crucial to remember that — for good reasons — digital supply may still fall within the scope of the sale framework (see also CJEU (2012)). She therefore stresses that it would be more consistent to separate the contractual treatment of the supply of 'digital content', into two different categories, depending on this 'logic of possession'. On the one hand, the sale contract would be appropriate for possession-compliant digital content (in other words, according to this paper, information goods). On the other hand, service or rental contracts would apply to digital content which lacks the 'logic of possession' (in other words, according to this paper, e-services).

This weakness is also illustrated by the points made by Mak (2016, p. 13) (see also Wenderhost (2016)): 'Whereas certain types of digital content could be treated as goods – e.g. software, e-books, content delivered on a tangible medium – other types do not fit well in this framework. Services through which digital content provided by the consumer is stored or

(7) Excepting digital elements included in goods, such as smart TVs.

processed (e.g. cloud services or social media platforms) are not so much similar to contracts for the sale or supply of goods, as they are to services'. According to this paper, the latter refers to e-service products or e-services-producing activities.

The source of the legal difficulties concerning the supply of digital content does not primarily stem from the digital nature of this content. If all information goods, whether of 'physical' appearance or otherwise, were regarded as goods, contractual exceptions would no longer be needed, and the categorisation of the supply of digital content between sales or service contracts would be simplified.

5. Conclusions

Undeniably, the progresses of the information-based economy have brought into question customary economic views and the content of legal texts. However, contrary to some opinions, the ongoing dematerialisation of goods does not primarily obliterate the dissimilarities between a good and a service, but encourages efforts to provide a better delineation of the boundary between them. It has been explained in this article that EU legal texts still rely on a traditional conception of this boundary, based on tangibility. This view, widespread in the economic literature of the 1960s and 1970s, was successfully challenged by an updated perspective, increasingly endorsed by the SNA since 1993 and the BoP in 2008. This revised approach is mainly based on the capacity to establish (or not) ownership rights on products. It is better suited to the information-based economy where intangibles and information are crucial for economic growth.

In 2015, the EC launched *A Digital Single Market Strategy for Europe* aimed at fostering economic growth in the digital and internet fields. However, this strategy faced several difficulties regarding the VAT rate and the legal framework to be set for the supply of digital content, which is basically concerned with the production and processing of information. In these areas, the present definition of a 'service' creates tax distortions, leads to legal contortions (unnecessary exceptions), and results in the invention of an idiosyncratic type of contract (a 'neither sale, nor service' contract). Many of these problems could be solved by adopting the updated approach to a service and by recognising that all digital products, even intangible ones, are information goods. Several legal exceptions would disappear as a result and the contractual relationship between sellers and customers, applicable to digital content, would fit more easily into the usual categories of sale, service or rent.

This article uses EU legal texts as examples of the goods-services border issue. Arguably, it quite often discusses legal considerations, which may be disturbing to economists or statisticians who are unfamiliar with this literature or approach. Firstly, this comes from the fact that the updated view of a service focuses on ownership rights. As such, it necessarily leads to reflections on principles of law. This also means that a convenient way to classify complex situations, regarding a good or a service, might be to look at the contractual links that bind suppliers and customers. Secondly, the idea behind the paper stems from the opinion that the perspective of EU legal texts on goods and services had not previously been studied, at least from an economic point of view. It appears that their principles, set out in 1958, are less and less relevant for the information-based economy and begin to produce adverse

consequences. Thirdly, another goal of the article is to illustrate the need for economists and legal experts to talk together more often. This would be beneficial for both fields. The circumstances under scrutiny provide valuable practical situations, which might help adapting the views of the SNA or that of ISIC (see Annex C).

Incidentally and finally, the updated perspective of a service would shed a more positive light on the deindustrialisation debate (Miroudot (2019), De Backer et al. (2015)). Indeed, this might reduce the measure of the services-producing part of modern economies, since ‘the rapid growth of intangibles is often viewed as contributing to the growth in service production, whereas it actually contributes to the growth of goods production’ (Hill (2014), p. 62) (see Annex C).

Acknowledgements

I am greatly indebted to the reviewers and the editor whose comments were invaluable to the final version of this paper. Obviously any remaining obscurities would be mine.

Annex A: excerpts from Council Directives 2006/112/EC and 2009/47/EC on the common system of value added tax and as regards reduced rates of value added tax

Article 14

1. 'Supply of goods' shall mean the transfer of the right to dispose of tangible property as owner.

Article 15

1. Electricity, gas, heat, refrigeration and the like shall be treated as tangible property.

(...)

Article 24

1. 'Supply of services' shall mean any transaction which does not constitute a supply of goods.

(...)

Article 25

A supply of services may consist, inter alia, in one of the following transactions:

(a) the assignment of intangible property, whether or not the subject of a document establishing title;

(...)

SUBSECTION 5

Supply of miscellaneous supply services

Article 56

(...)

(a) transfers and assignments of copyrights, patents, licences, trade marks and similar rights;

(...)

(k) electronically supplied services, such as those referred to in Annex II;

SECTION 2

Reduced rates

Article 98

(...)

2. The reduced rates shall apply only to supplies of goods or services in the categories set out in Annex III.

of which (as modified by the Council Directive 2009/47/EC)

(6) supply, including on loan by libraries, of books on all physical means of support (including brochures, leaflets and similar printed matter, children's picture, drawing or colouring books, music printed or in manuscript form, maps and hydrographic or similar charts), newspapers and periodicals, other than material wholly or predominantly devoted to advertising; (emphasis added)

Annex B: definitions of services, goods and digital content — excerpts from Directive 2011/83/EU, a proposal for a Directive (COM(2015) 634 final) and Directive (EU) 2019/770

A) FROM THE CONSUMER RIGHTS DIRECTIVE (DIRECTIVE 2011/83/EU ON CONSUMER RIGHTS)

(5) (...) in particular in the services sector, for instance utilities, (...)

(19) Digital content means data which are produced and supplied in digital form, such as computer programs, applications, games, music, videos or texts, irrespective of whether they are accessed through downloading or streaming, from a tangible medium or through any other means. Contracts for the supply of digital content should fall within the scope of this Directive. If digital content is supplied on a tangible medium, such as a CD or a DVD, it should be considered as goods within the meaning of this Directive. Similarly to contracts for the supply of water, gas or electricity, where they are not put up for sale in a limited volume or set quantity, or of district heating, contracts for digital content which is not supplied on a tangible medium should be classified, for the purpose of this Directive, neither as sales contracts nor as service contracts. (...)

Article 2

(3) 'goods' means any tangible movable items (...)

(5) 'sales contract' means any contract under which the trader transfers or undertakes to transfer the ownership of goods to the consumer and the consumer pays or undertakes to pay the price thereof, including any contract having as its object both goods and services;

(6) 'service contract' means any contract other than a sales contract under which the trader supplies or undertakes to supply a service to the consumer and the consumer pays or undertakes to pay the price thereof;

(...)

(11) 'digital content' means data which are produced and supplied in digital form;

B) FROM A PROPOSAL FOR A DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL ON CERTAIN ASPECTS CONCERNING CONTRACTS FOR THE SUPPLY OF DIGITAL SERVICES (COM(2015) 634 FINAL): THE BROAD VIEW OF DIGITAL CONTENT

Article 2

1. 'digital content' means

- (a) data which is produced and supplied in digital form, for example video, audio, applications, digital games and any other software,
 - (b) a service allowing the creation, processing or storage of data in digital form, where such data is provided by the consumer, and
 - (c) a service allowing sharing of and any other interaction with data in digital form provided by other users of the service;
- (...)

11. 'durable medium' means any instrument which enables the consumer or the supplier to store information addressed personally to that person in a way accessible for future reference for a period of time adequate for the purposes of the information and which allows the unchanged reproduction of the information stored.

C) FROM THE DIGITAL CONTENT AND DIGITAL SERVICES DIRECTIVE (DIRECTIVE (EU) 2019/770 ON CERTAIN ASPECTS CONCERNING CONTRACTS FOR THE SUPPLY OF DIGITAL CONTENT AND DIGITAL SERVICES): THE NARROW VIEW OF DIGITAL CONTENT

(19) The Directive should address problems across different categories of digital content, digital services, and their supply. (...) this Directive should cover, inter alia, computer programmes, applications, video files, audio files, music files, digital games, e-books or other e-publications, and also digital services which allow the creation of, processing of, accessing or storage of data in digital form, including software-as-a-service, such as video and audio sharing and other file hosting, word processing or games offered in the cloud computing environment and social media. As there are numerous ways for digital content or digital services to be supplied, such as transmission on a tangible medium, downloading by consumers on their devices, web-streaming, allowing access to storage capabilities of digital content or access to the use of social media, this Directive should apply independently of the medium used for the transmission of, or for giving access to, the digital content or digital service. However, this Directive should not apply to internet access services.

Article 2

(1) 'digital content' means data which are produced and supplied in digital form;

(2) 'digital service' means:

- (a) a service that allows the consumer to create, process, store or access data in digital form; or
- (b) a service that allows the sharing of or any other interaction with data in digital form uploaded or created by the consumer or other users of that service;

Annex C: information and communication (ISIC Section J) — an updated view of service activities and their outputs

In accordance with the views presented in this paper, any activity, the output of which is, on a main and usual basis, neither separable from the provider nor from the recipient, is classified within the services-producing economic sector. This especially concerns manufacturing type activities, which are carried out on a subcontracted basis for third parties on not owned (or more commonly described as not self-produced) materials. Conversely, any activity which has output over which, on a main and usual basis, ownership rights may be established (in other words output which is separable either from the producer or the consumer (or both)), is classified within the goods-producing economic sector (as for example standard manufacturing), whether traditionally regarded as providing services or not. This especially concerns information activities, which are usually classified within the services-producing economic sector.

The information and communication section, which was created with ISIC Rev.4, is the principal section to be impacted by this updated definition of services and service activities. Its foreword stresses that it includes ‘the production and distribution of information and cultural products’ (UN (2009), p. 205). Several of these cultural products are, in SNA’s terms, knowledge-capturing products, over which ownership rights may be established; therefore they are considered to be information goods, and it is only their delivery that is a service activity.

The cases of *printing and reproduction of recorded media* (ISIC Division 18) and *publishing* (ISIC Division 58) are especially interesting. Although the former activity is traditionally viewed as belonging to manufacturing, with a few exceptions ⁽⁸⁾, and the latter to service activities, the updated perspective tends to revisit this disconnection.

Indeed printing results in a separable entity and thus belongs to goods-producing, but publishing may also be categorised this way. As the ISIC Rev.4 manual states ‘Publishing includes the acquisition of copyrights to content (information products) and making this content available to the general public by engaging in (or arranging for) the reproduction and distribution of this content in various forms’ (UN (2008), p. 206). While the distributive part of the activity definitely belongs to services-producing, it is shared with various stores and retailers (included in the distributive trades section) and amounts to a small fraction of the turnover. The other foremost components of the activity therefore concern production and reproduction. The product (asset if an original) under exchange is prone to ownership rights, and thus the activity falls within the goods-producing activities. Let us emphasise that this rationale is applicable to all activities which deal with the production of information goods, such as software, directory or mailing lists, newspapers, motion pictures (see Table 1).

⁽⁸⁾ Division 18 printing and reproduction of recorded media, is included in manufacturing in ISIC Rev.4, of which Class 1812 is service activities related to printing.

These principles applied to the two divisions *motion picture, video and television programme production, sound recording and music publishing activities and programming and broadcasting activities* (ISIC Divisions 59 and 60), lead to separate programming or production (in other words the activity of making a motion picture, a video or a TV programme, or else making a master recording ⁽⁹⁾) and broadcasting or distribution. Programming and production fall within goods-producing activities, since they result in information goods, whereas broadcasting and distribution fall within services-producing activities. However, these two differing economic functions (programming and production on one hand and broadcasting and distribution on the other) are not always distinguished. Whenever mixed, their ensuing subheadings may generally be considered as goods-producing activities because the main share of their value added is external to the broadcasting activity. Nonetheless three classes (5912, 5913 and 5914) are indeed exclusively services-producing activities (see Table 1).

Table 1: Information and communication (ISIC Divisions 58-63)

Division	Group	Class	Description	Output
Division 58			Publishing activities	
	581		Publishing of books, periodicals and other publishing activities	goods
	582		Software publishing	
Division 59			Motion picture, video and television programme production, sound recording and music publishing activities	
	591		Motion picture, video and television programme activities	goods
		5911	Motion picture, video and television programme production activities	goods
		5912	Motion picture, video and television programme post-production activities	services
		5913	Motion picture, video and television programme distribution activities	services
		5914	Motion picture projection activities	services
	592		Sound recording and music publishing activities	mainly goods
Division 60			Programming and broadcasting activities	
	601		Radio broadcasting	mainly services
	602		Television programming and broadcasting activities	mainly goods
Division 61			Telecommunications	services
Division 62			Computer programming, consultancy and related activities	
		6201	Computer programming activities	goods
		6202	Computer consultancy and computer facilities management activities	services
		6209	Other information technology and computer service activities	services
Division 63			Information service activities	
	631		Data processing, hosting and related activities; web portals	
		6311	Data processing, hosting and related activities	services: not owned or self-produced data
		6312	Web portals	goods
	639		Other information service activities	
		6391	News agency activities	see text
		6399	Other information service activities n.e.c.	see text

⁽⁹⁾ Music and sound recording may either result in a service or in an information good depending on the commercial contract. However since it is mixed with publishing it is categorised in goods-producing.

Computer programming, consultancy and related activities (Division 62) includes programming activities that result in information goods, such as software.

Information service activities (Division 63) includes data processing and web portals activities, of which several headings may be considered to be goods-producing activities, when they produce and sell information. Most frequently, data processing operates on data belonging to a client (or a third party) (see ISIC Rev. 3.1 accompanying comments), whereas web portals activities provide data or information they have mostly produced or gathered. When the provision of information is organised through a long-term access license, it is worth recalling that the SNA considers that transaction similar to an outright sale.

Other information service activities (Group 639) is quite ambiguously labelled. This class includes activities such as news agencies, which mainly sell information, which is obviously separable from its provider or receiver, and therefore belongs to information goods.

References

Benabou, V-L. (2017), «*Entrée par effraction d'une notion juridique nouvelle et polymorphe : le contenu numérique* », *Dalloz IP/IT : droit de la propriété intellectuelle et du numérique*, IPIT no 1. 2017, pp. 7-14.

Broussolle, D. (2014), '*Service, Trade in Services and Trade of Services Industries*', *Journal of World Trade*, Volume 48, Issue 1, pp. 31-58.

Broussolle, D. (2015), '*The Emerging New Concept of Services in the System of National Accounts and the Balance of Payments*', *Review of Income and Wealth*, Volume 61, Issue 3, pp. 573-586.

Corrado, C., J. Haskel, M. Iommi, C. Jona-Lasinio, M. Mas and M. O'Mahony (2017), '*Advancements in measuring intangibles for European economies*', *EURONA — Eurostat Review on National Accounts and Macroeconomic Indicators*, Issue 2/2017, pp. 89-106, Publications Office of the European Union, Luxembourg.

Council of the European Union (2006), *Council Directive 2006/112/EC on the common system of value added tax*, Official Journal of the European Union, L 347, pp. 1-118.

Council of the European Union (2009), *Council Directive 2009/47/EC of 5 May 2009 amending Directive 2006/112/EC as regards reduced rates of value added tax*, Official Journal of the European Union, L 116, pp. 18-20.

Council of the European Union (2011), *Council Implementing Regulation (EU) No 282/2011 of 15 March 2011 laying down implementing measures for Directive 2006/112/EC on the common system of value added tax*, Official Journal of the European Union, L 77, pp. 1-22.

Court of Justice of the European Union (1985), *Tax advantages for newspaper publishers*, Case C-18/84, Judgement of the Court of 7 May 1985.

Court of Justice of the European Union (2012), *Legal protection of computer programs — Marketing of used licences for computer programs downloaded from the internet*, Case C-128/11, Judgement of the Court (Grand Chamber) of 3 July 2012.

Court of Justice of the European Union (2015), *Failure of a Member State to fulfil obligations — Taxation — VAT — Application of a reduced rate — Supply of digital books or electronic books*, Case C-479/13, Judgement of the Court (Fourth Chamber) of 5 March 2015.

De Backer, K., I. Desnoyers-James and L. Moussiégt (2015), '*“Manufacturing or Services - That is (not) the Question”: The Role of Manufacturing and Services in OECD Economies*', *OECD Science, Technology and Industry Policy Papers*, No. 19, OECD Publishing, Paris.

Dreier, T. (2013), '*Online and Its Effect on the “Goods” Versus “Services” Distinction*', *International Review of Intellectual Property and Competition Law*, Volume 44, Issue 2, pp. 137-139.

European Commission (2011), *Proposal for a regulation of the European Parliament and of the Council on a Common European Sales Law*, COM(2011) 635 final.

European Commission (2015a), *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Digital Single Market Strategy for Europe*, COM(2015) 192 final.

European Commission (2015b), *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee: Digital contracts for Europe — Unleashing the potential of e-commerce*, COM(2015) 633 final.

European Commission (2015c), *Proposal for a directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content*, COM(2015) 634 final.

European Commission (2015d), *Proposal for a directive of the European Parliament and of the Council on certain aspects concerning contracts for the online and other distance sales of goods*, COM(2015) 635 final.

European Commission (2017a), *Communication from the Commission to the European Parliament and the Council: A Fair and Efficient Tax System in the European Union for the Digital Single Market*, COM(2017) 547 final.

European Commission (2017b), *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the follow-up to the Action Plan on VAT: Towards a single EU VAT area — Time to act*, COM(2017) 566 final.

European institutions (2001), *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society*, Official Journal of the European Union, L 167, pp. 10-19.

European institutions (2006), *Directive 2006/123/EC of the European Parliament and of the Council of 12 December 2006 on services in the internal market*, Official Journal of the European Union, L 376, pp. 36-68.

European institutions (2011), *Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council*, Official Journal of the European Union, L 304, pp. 64-88.

European institutions (2019a), *Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services*, Official Journal of the European Union, L 136, pp. 1-27.

European institutions (2019b), *Directive (EU) 2019/771 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the sale of goods, amending Regulation (EU) 2017/2394 and Directive 2009/22/EC, and repealing Directive 1999/44/EC*, Official Journal of the European Union, L 136, pp. 28-50.

Hill, T. P. (1977), 'On Goods and Services', *The Review of Income and Wealth*, Volume 23, Issue 4, pp. 315-338.

- Hill, T. P. (1999), 'Tangibles, intangibles and services: a new taxonomy for the classification of output', *Canadian Journal of Economics*, Volume 32, Issue 2, pp. 426-446.
- Hill, T. P. (2003), 'Intangible assets in the SNA: originals, innovation and human capital formation', Working Papers, CASS Business School, London.
- Hill, T. P. (2014), 'Intangibles and services in economic accounts', *EURONA — Eurostat Review on National Accounts and Macroeconomic Indicators*, Issue 1/2014, pp. 59-72, Publications Office of the European Union, Luxembourg.
- Hojnik, J. (2016), 'The servitization of industry: EU law implications and challenges', *Common Market Law Review*, Volume 53, Issue 6, pp. 1 575-1 623.
- IMF (2009), *Balance of Payments and International Investment Position Manual, Sixth Edition (BPM6)*, International Monetary Fund, Washington D.C.
- Jansson, J. O. (2009), 'The myth of the service economy — An update', *Futures*, Volume 41, Issue 3, pp. 182-189.
- Linklater, E. (2015), 'The end of the story for reduced VAT rates for E-books? Commission v. France, Commission v. Luxembourg', *Common Market Law Review*, Volume 52, Issue 6, pp. 1 679-1 690.
- Mak, V. (2016), *The new proposal for harmonised rules on certain aspects concerning contracts for the supply of digital content: In-depth analysis*, Workshop for the JURI Committee, Policy Department C: Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, European Parliament.
- Manko, R. (2016), *Contracts for supply of digital content: A legal analysis of the Commission's proposal for a new directive*, EPRS — European Parliamentary Research Service, PE 582.048, Directorate-General for Parliamentary Research Services, European Parliament.
- Miroudot, S. (2019), *Services and Manufacturing in Global Value Chains: Is the Distinction Obsolete?*, ADBI Working Paper Series, No. 927, Asian Development Bank Institute, Tokyo.
- Mishra, S., S. Lundström Gable and R. Arnand (2012), 'Service export sophistication and economic growth', *VoxEU*, Centre for Economic Policy Research.
- Morais Carvalho, J. (2019), 'Sale of goods and supply of digital content and digital services – overview of directives 2019/770 and 2019/771', *Journal of European Consumer and Market Law*, Volume 8, Issue 5, pp. 194-201.
- OECD (2000), *The service economy*, Business and Industry Policy Forum Series, Organisation for Economic Co-operation and Development, OECD Publications, Paris.
- Pilat, D. and A. Wölfl (2005), *Measuring the Interaction between Manufacturing and Services*, OECD STI Working Paper 2005/5, OECD Directorate for Science, Technology and Innovation (STI), Paris.
- Sénéchal, J. (2015), « Le contrat de fourniture de contenu numérique en droit européen et français : une notion unitaire ou duale », *Revue de l'Union européenne*, Numéro 590, pp. 442-454.

Shapiro, C. and R. Varian Hal (1998), *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston, Massachusetts.

Snell, J. (2002), *Goods and Services in EC Law: A Study of the Relationship Between the Freedoms*, Oxford University Press, Oxford.

UN (2008), *International Standard Industrial Classification of All Economic Activities, (ISIC), Rev.4.*, Statistical Papers, Series M, No. 4/Rev.4, Department of Economic and Social Affairs, Statistics Division, United Nations, New York.

UN (2009), *System of National Accounts 2008*, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank, United Nations, New York.

UN (2010), *United Nations Convention on Contracts for the International Sale of Goods*, United Nations Commission on International Trade Law, United Nations, New York.

UNCTAD (2004), *World Investment Report 2004: The Shift Towards Services*, United Nations Conference on Trade and Development, United Nations, New York and Geneva.

Wendehorst, C. (2016), *Sale of goods and supply of digital content — two worlds apart? Why the law on sale of goods needs to respond better to the challenges of the digital age*, Study for the JURI Committee, Policy Department C: Citizens' Rights and Constitutional Affairs, Directorate-General for Internal Policies, European Parliament.

3

A systematic approach for data validation using data-driven visualisations and interactive reporting

JULES ZACCARDI (*) AND ENRICO INFANTE (*)

Abstract: In this article we present the development of a tool that can serve most official statistics. This tool is fully interactive and based mostly on visualisations, which makes it a powerful ally to validate or analyse data. To make our tool a pragmatic solution in an era facing ever larger volumes of data, we integrated a data-driven analysis to identify the most unusual data and suggest them to the user. By defining the right metrics and smartly designing our report, we managed to develop an easily-accessible tool that can be used to validate or analyse data. We tested this tool for data validation in annual financial accounts, and it demonstrated its ability to identify the main issues in the dataset very quickly, therefore improving i) the productivity of the user of the tool ii) the long-term viability of the validation and iii) ultimately the quality of the data.

JEL codes: C02, C61, G20, P43

Keywords: data literacy, data validation, automation, visualisation, design, outlier detection

(*) Eurostat, Unit C2 — National accounts production.

1. Introduction

1.1. Research problem

Eurostat supplies citizens, governments and institutions with high quality statistics and data on the European Union (EU). One of its most important missions is ensuring the finest statistical quality. As stated in the *European statistics code of practice* (Eurostat (2018)), 'We see quality as the basis of our [Eurostat's] competitive advantage in a world experiencing a growing trend of instant information which often lacks the necessary proof of quality'. 'Statistical excellence is essential because most economic speculation, social analysis and political decisions are based on statistical foundations' (Cohen (1938)).

The purpose of data validation is to ensure a certain level of quality of the final data. Eurostat commits to a continuous improvement framework, by always challenging its processes and looking for possible weaknesses. The modernisation of data validation is thus at the centre of our mission. Besides performing data validation, official statisticians also use data for analysis, exploration and the monitoring of performance indicators. The scope of our tool therefore also includes the possibility to present and analyse data produced from official statistics.

The core principles of quality as defined in Eurostat (2018) are: relevance, accuracy, timeliness and punctuality, accessibility and clarity, as well as comparability and coherence. **In this article we propose innovative methods aiming at controlling the accuracy and the comparability of data.** These methods apply a systematic approach based on data-driven visualisations and interactive reporting. With the final goal of ensuring the finest data quality, the tool described in this article aims at integrating innovative statistical techniques to control the plausibility of data and using modern technologies to produce simple but straightforward error messages.

Within the European Statistical System (ESS), Eurostat and the EU Member States have already invested a lot in data validation. The objective of this article is to suggest further developments for data validation.

1.2. State of the art

We conducted a bibliographic review to establish what is the state-of-the-art regarding data-driven visualisations and outlier detection, with a special interest in financial accounts.

As outlined in Bay et al. (2006), there is a need for an automatic analysis of irregularities in large datasets (for example, fraud or errors). Thus, a variety of models currently exist to examine financial accounts. While Yang and Cogill (2013) considered a graph similarity algorithm using the string edit (or Levenshtein) distance to identify an unconventional financial statement, Bueno and Infante (2013) considered using SARIMA forecast intervals for the predictability of financial time series. We decided to follow an approach based on the ideas presented in Lenderink (2019) in which Lenderink explains that a fraud or an error 'has to deviate from "normal" behaviour and from regular financial transaction patterns'. Following this idea, the most interesting approach seemed to us the one of Zhu (2006) which lies in estimating the deviation from a tested data point to a reference dataset.

Furthermore, given the increasing development of visualisation-based analytical tools and big data, a lot of work has been carried out to develop the design, the speed and the relevance of such tools. Mackinlay (1986) proposed a way of *Automating the design of graphical presentations of relational information* which we considered in the way we designed our tool overall. Hao et al. (2005) presented a framework for visualising large sets of time series. Although they faced the same problem as us, their approach was to find a way to show all the series while introducing a metric of importance that made it possible to prioritise the most useful ones. We considered this to be too heavy and therefore we preferred the approach of Vartak et al. (2015) and Kandel et al. (2012). Vartak et al. (2015) developed *SeeDB*, a visualisation recommendation engine which — given a dataset — intelligently explores the space of visualisations, evaluates promising ones and recommends those it deems more useful to the user. Similarly, Kandel et al. (2012) developed *Profiler*, a visual analysis tool for assessing quality in tabular data. The key to these approaches lies in the suggestion of interesting views and scalable visual summaries that support real-time interaction with the user across millions of data points.

2. About validation in official statistics

2.1. The generic procedure

The ESS has carried out a considerable amount of work to produce ‘a generic framework for data validation in order to have a reference context, and to provide tools for setting an efficient and effective data validation procedure’ (Di Zio et al. (2018)). According to the ESS definition, ‘data validation is an activity verifying whether or not a combination of values is a member of a set of acceptable combinations’ (Di Zio et al. (2018)), and its process includes:

- establishment of checking rules;
- detection of outliers or potential errors;
- communication of the detailed problems to the actors in best position to investigate them.

The purpose of data validation is to ensure a certain level of quality of the data. It focuses on specific dimensions of data quality, such as accuracy, comparability and coherence. Accuracy refers to the measurement of the difference between the ‘target parameter’ and the ‘estimated parameter’, while coherence and comparability refer to statistics being consistent internally (between related data points), over time and space (for example, comparable between regions and countries). Data validation can be divided in two categories.

- **Structural validation**, aimed at verifying the technical integrity of the file (in other words, consistency with the expected IT structural requirements).
- **Content validation**, aimed at verifying the logical and statistical consistency of the data.

This includes:

1. consistency within the dataset;
2. consistency with other datasets within the same domain and within the same data source;
3. consistency within the same domain between different data sources;
4. consistency between separate domains for the same data provider;
5. consistency with data from other data providers.

In this article we propose an improvement for points 1 and 4 of content validation. Following the definition of data validation (as provided at the beginning of this section), our approach has two main goals. The first is an improvement in detecting outliers or potential errors in a big set of time series by introducing advanced statistical methodologies. The second is to ease the communication process with statistical reporting agencies to facilitate the investigation and the correction of errors, by automating validation procedures as much as possible.

2.2. The case of financial accounts

Financial flows and stocks are the data described in Tables 6 and 7 of the *European system of accounts 2010 — ESA 2010 — Transmission programme of data* (Eurostat (2014)). The data consist of several thousand time series (up to 25 000) with annual observations since 1995. As presented in ESA 2010, financial flows and stocks are structured in the following ways.

- They can be stocks, transactions or other flows. Let us denote $F = \{\text{stocks, transactions, other changes in volume, revaluation account}\}^{(?)}$.
- They are split between institutional sectors (most of which are split into several subsectors). Let us denote S as the set of sectors, described by:
 - non-financial corporations (S.11),
 - financial corporations (S.12),
 - general government (S.13),
 - households (S.14),
 - non-profit institutions serving households (S.15).
- They are described by financial instruments, which constitute the total economy overall. Let us denote I as the set of financial instruments, described by:
 - monetary gold and special drawing rights (F.1),
 - currency and deposits (F.2),
 - debt security (F.3),
 - loans (F.4),
 - equity and investment fund shares or units (F.5),
 - insurance, pension and standardised guarantee schemes (F.6),
 - financial derivatives and employee stock options (F.7),
 - other accounts receivable/payable (F.8).
- They can be assets or liabilities. Let us denote $A = \{\text{assets, liabilities}\}$.
- Finally, they can be consolidated or not consolidated. Let us denote $C = \{\text{consolidated, non-consolidated}\}$.

^(?) As the other changes in volume and revaluation accounts are specific types of flows, in this article we will only mention either 'flows' or 'transactions', always referring simultaneously to transactions, other changes in volume and revaluation accounts.

Most of the institutional sectors and financial instruments mentioned above consist of different layers of subsectors (such as subsectors S.121, S.122 and so on) and different layers of instruments (such as F.51, F.511, F.512 and so on). Multiplied by the $|F| \times |A| \times |C| = 16$ possible categories, we obtain several thousand time series in the data. Financial accounts are by nature very volatile data, therefore it is hard to model trends in the data and predict future values.

The current validation system produces an automatic report based on implemented rules. Although this report includes all the inconsistencies, it is hard to read and to make use of. Weaknesses appear especially regarding the analysis of revisions and the detection of outliers (see next section for more details). The current system is a great basis to build on, but it demonstrates the need for a systematic approach using data-driven visualisations and interactive reporting, which is the focus of this article.

Our goal is to develop a reporting framework which is user-friendly, makes it possible for validators to spot the biggest inconsistencies easily, and reduces the burden for statistical reporting agencies to investigate and correct errors.

3. A framework for interactive and automatic validation reporting

3.1. General framework

Validation procedures usually include a reporting step during which the inconsistencies are listed in a document which is called the **validation report**. The latter generally includes the following sections.

- **Completeness.** This section shows the rate of missing values overall.
- **Confidentiality.** This section shows the rate of confidential data overall.
- **Zero and negative values.** This section shows counts of zeros and negative values. In many domains the latter should not occur, for example for statistics on the level of population. In the case of financial accounts, stocks should always be non-negative.
- **Internal consistency.** This section shows whether breakdowns add up to their main series. Most domains deal with series that are breakdowns and are faced with verifying that those add up to their relevant main series. In financial accounts, we check institutional sectors that are broken down into subsectors ^(?), to see if the sum of the values for the subsectors adds up to the value of the main sector.
- **Revisions.** This section is based on a comparison of the data to be validated with the data validated during the previous production round. It shows changes (called revisions) in the data.

In this article, we propose an updated framework for validation reporting that is more user-friendly, more straightforward and based on innovative statistical methodologies.

^(?) Formally, denoting by $n \in \mathbb{N}$ the maximal number of layers in a sector we can write that $S_n \subset S_{n-1} \subset \dots \subset S$.

The main aspect that we will leverage to this purpose will be the visualisation of time series, which makes it easier for the validator to spot errors and understand how to act. However, it is impossible to visualise thousands of time series manually. Therefore, we propose in this article a method based on data-driven visualisation suggestions. To integrate this new approach, we propose the following new framework:

- **completeness;**
- **confidentiality;**
- **zero and negative values;**
- **internal consistency** ⁽⁴⁾;
- **macroeconomic indicators** — here we show key indicators used by institutions to have a first overview of the data ⁽⁵⁾;
- **generic visualisations** — here we propose a systematic visualisation of the most important mandatory series;
- **revisions visualisations** — here we propose visualisations of series, based on computations to spot the most unusual revisions;
- **detection of outliers** — here we propose visualisations of series based on computations to spot outliers.

Such a framework ensures that the report covers all the needs for official statistics. Indeed, not only does it improve data validation, but it also makes it possible to explore the data and to monitor macroeconomic indicators.

3.2. Metadata

3.2.1. DESCRIPTION OF THE PROCEDURE

As stated in Section 2, the process of data validation includes the communication of detailed problems to the actors in the best position to investigate and correct them. Once we have spotted errors (or inconsistencies) and contacted statistical reporting agencies, either the errors are corrected or the agencies provide us with a reason(s) for the inconsistency(ies). These explanations then become part of the **metadata**.

Given that the exchange of information described in the previous paragraph happens during every production round, we introduce in this article the use of a **metadata template** ⁽⁶⁾. This template is a formatted document that includes a list of known (by the statistical reporting agency) inconsistencies and detailed explanations for them. The inconsistencies are clearly identified in terms of series and year, and when possible categorised. Statistical reporting agencies would fill in this template and transmit it together with the data.

In the case of national accounts, the template is structured in such a way to clearly identify the series, the nature of the problem and the reason for it. It therefore contains fields to indicate the elements of **F**, **S**, **I**, **A** and **C** necessary to identify a series, one field to describe the nature

⁽⁴⁾ These elements are not refined further in the article because we consider that the current solution is sufficiently developed.

⁽⁵⁾ In the case of financial accounts we use the [macroeconomic imbalance procedure indicators](#).

⁽⁶⁾ Such a template is not an invention of the authors. Indeed, this solution exists and is used in several domains across official statistics. Therefore, this proposal specifically concerns financial accounts, where such templates are not yet in use.

of the issue (it can be a negative value, a major event or an outlier, an important revision, or another reason) and an explanation for it.

3.2.2. ADDED VALUE OF THIS CONCEPT

The use of such a template has three main advantages.

The first one is to develop an easy communication between the different actors involved (in the case of financial accounts in the EU, these would mainly be the validator in Eurostat and the statistical reporting agencies in the EU Member States), therefore formalising, structuring and simplifying communication, making it more precise and unambiguous, ensuring overall a smooth validation. By using such a template, we give the Member States a chance to confirm and explain the known unusual values that are present in their data (⁷). We then know that any other unusual values that we identify will need to be investigated and possibly corrected.

The second one is that this template makes it possible for us to store the inconsistencies and reasons over time, therefore opening a world of possibilities, from improving the methodology of the data collection to ensuring a bilateral follow-up on data quality with the EU Member States. Importantly, it should also avoid data validators repeatedly asking statistical reporting agencies to confirm values that have already been confirmed, thereby reducing the burden on both actors.

Finally, this template makes it possible to use the metadata during the validation stage and include the information provided by countries in the validation report. This is a huge improvement in the sense that it takes the validation one step further. Whereas we used to only spot inconsistencies, we now have the possibility to investigate them immediately (at the time of the validation). Indeed, we will specify in the validation report whether an inconsistency has been mentioned in the metadata template or not.

The idea of introducing a metadata template is very generic and can easily be adapted to any domain of official statistics. In fact, several domains already make use of one, and in this article we propose the implementation of this concept for financial accounts. As metadata play a key role in the validation of official statistics, easing the way in which they are collected is a valuable innovation.

3.3. Cross-domain checks: the case of comparing annual and quarterly financial accounts

Cross-domain checks consist of ‘checks between the data available in the data provider (institution) and the data available outside the data provider (institution). This implies no control over the methodology on the basis of which the external data are collected, and sometimes a limited knowledge of it.’ In our case, annual financial accounts and quarterly financial accounts are based on different legal bases (regulations), and are often compiled by different institutions, although they are theoretically the same thing, just with different periodicity.

(⁷) For instance, EU Member States may be aware of a big revision they made, or of an unusual value in their data, and explain it at the time of transmission.

3.3.1. DESCRIPTION OF THE PROCEDURE

Eurostat receives annual financial accounts (AFA) from each EU Member State, while the European Central Bank receives quarterly financial accounts (QFA) from them. To ensure the quality of the data, the consistency between AFA and QFA is checked. To do this, we implemented an automatic consistency check which generates individual reports for each Member State. Although this validation might be specific to financial accounts, it can be generalised to other domains considering any simple cross-domain check. Indeed, the added value lies not in the check itself, but in the automation of it.

The elements to be checked are the following:

- for transactions, $\sum_{i=1}^4 Q_i = A$;
- for stocks, $Q_4 = A$;
- for all series, that the observation flags ⁽⁸⁾ and confidentiality flags ⁽⁹⁾ are identical for AFA and QFA.

In this case, the checks to be made are very simple, but there is a need for an automatic and elegant procedure. For this article we developed a tool that uses the raw data received through the SDMX procedure ⁽¹⁰⁾ and generates reports outlining the inconsistencies in a clear way. Both of the following rules need to be satisfied to consider a discrepancy as a proper inconsistency:

- $A - Q > 10$ million of national currency;
- $\frac{A - Q}{GDP} > 0.3\%$.

The code is structured into three main phases. Firstly, it reads and cleans the data, removing any information that would not be useful in this situation. Then it merges the AFA and QFA data into one single table and computes the gap between the values. Finally, it uses this information to generate individual reports for each EU Member State.

3.3.2. ADDED VALUE OF THIS CONCEPT

The implementation of an automated procedure is of high added value for three reasons.

Firstly, it is effortless for the validation team, as it only requires uploading the data and pressing a button. Everything from reading the data to computing the errors is done with R (R Core Team (2020)). It strengthens the **productivity** of the validation team.

⁽⁸⁾ Observation status flags (OBS STATUS): these provide information on the quality of a value or an unusual or missing value.

⁽⁹⁾ Confidential status flags (CONF STATUS): these provide information about the confidentiality status of the object to which this attribute is attached.

⁽¹⁰⁾ Standard for data and metadata exchange.

Secondly, it is an elegant document to send the statistical reporting agencies. We found a solution that can serve to easily read and interpret a long set of inconsistencies. By using a reactive document with different tabs, and listing separately the series presenting inconsistencies and the complete set of errors we made the report easy to read whatever the number of errors is ⁽¹⁾. We started by an introductory paragraph which presents the situation, stating the purpose of the report, specifying when the AFA and QFA data were each received, and indicating the number of inconsistencies (all of this, as mentioned, is done automatically). We continued with a tab showing the number of errors grouped by institutional sector and by financial instrument, which helps identifying from where errors come. Then we added a tab listing the series for which an inconsistency has been observed. Finally, we listed all the inconsistencies, including the values and the gap (in absolute terms and relative to GDP). With such simple reports, this procedure encourages **reactivity** from the statistical reporting agencies and provides the basis for a simple communication between the various actors involved in data validation. In the long term, it should improve the quality of the data.

Finally and most importantly, it is a very reliable solution. Indeed, the procedure makes it possible to control automatically the consistency of each single time series, while a manual process might focus (only) on the most aggregated time series or on some indicators. Furthermore, this procedure also makes it possible to include an analysis of consistency, checking the observation and confidentiality flags. Therefore, it strengthens the **sustainability** of the validation.

4. An innovative way to visualise large datasets of time series

4.1. An approach based on interactivity

The main innovation of our approach is the interactivity of our report. Rather than generating a static list of problems, we give the opportunity for the user to explore the data. This is made possible by carefully considering the structure of the report, and giving the possibility to the user to experiment with it.

The user can click on, expand and extract graphics. By dragging the mouse over the graphics, the user can observe closely the values and additional information such as the percentage rate of change. The user can also zoom in or out and download graphics as images.

Furthermore, the sections of the report were carefully chosen to give the user the most relevant overview of the data. Starting from key indicators and the most aggregated data, it then offers a closer look at the revisions made on the data and possible outliers. The graphics of the outliers' section include a smoothed line and a confidence interval.

⁽¹⁾ See Illustration 1 and Table 1 for more details.

Finally and most importantly, the user can choose what to see by way of a reactive drop-down menu, which gives the possibility to explore the data more fully. More specifically in the case of financial accounts, the user can choose by way of drop-down menus each element of *F*, *I*, *C* and *A* in order to isolate one specific series. The strength of this model is that it leaves the opportunity for the user to explore the data very closely, isolating each chosen series, but also making possible an exploration of the full dataset. Indeed, the report automatically recommends series to select, based on data-driven visualisations.

4.2. Data-driven visualisations recommendations

In this section we explain how we managed to identify the most relevant data in order to suggest its visualisation to the user of the report. The question we wanted to answer is how to find the most relevant visualisation among a large amount of data. Our idea to solve this problem is based on *SeeDB: efficient data-driven visualisation recommendations to support visual analytics* (Vartak et al. (2015)). Their approach to suggest relevant visualisations is to have one metric to **filter** out what is not relevant and one other metric to **order** series in terms of relevance. They adopted deviation-based metrics for a visualisation utility.

4.2.1. REVISIONS

The approach of this section is to examine by how much data have been revised. This can be done by comparing data that are currently disseminated and data that are to be validated during a production round. The main goal is to evaluate the changes made to a set of data *X*, that we will call the *reference*, to a set of data *Y* that we will call the *target*. Let us introduce the following notation:

1. *n* the number of observations in the reference ⁽¹²⁾;
2. *m* the number of observations in the target ⁽¹³⁾;
3. $X = \{x_1, \dots, x_{n-1}\}$ the *reference* data;
4. $Y = \{y_1, \dots, y_{n-1}, y_n\}$ the *target* data.

We will standardise both *X* and *Y*, in other words we will apply the transformation

$$f : x \rightarrow \frac{x - \min(x)}{\max(x) - \min(x)}.$$

We will refer to the standardised data as X_s and Y_s . We will then consider three different distances:

- The Euclidean or L_2 norm distance between the standardised distributions.

$$D_e(X_s, Y_s) = \|X_s - Y_s\|^2$$

- The Chebyshev distance or L_∞ norm.

$$D_c(X, Y) = \max |X_i - Y_i|$$

⁽¹²⁾ In the case of financial accounts, time series usually start from 1995, so we have around 25 annual observations per series.

⁽¹³⁾ Most of the time, $m = n$ or $m = n + 1$.

- The Hamming distance which corresponds to the rate of revisions made.

$$H(X, Y, \delta) = \frac{1}{n} \sum_{i=0}^n \omega_i \text{ where } \omega_i = \begin{cases} 0 & \text{if } |X_i - Y_i| < \delta \\ 1 & \text{otherwise} \end{cases}$$

The last of these is very relevant for our purpose given that it indicates the number of observations that have been revised, which is a key indicator to understand if the revision should be investigated or not. Often, data are revised following a 3-year cycle, meaning that revisions of a small amount for the last 3 years are not suspect. Thus, we want to focus on large revisions over the whole time span. Therefore, we computed the Euclidean distance to order the series based in terms of the magnitude of their revisions, and we weighted this by the Hamming distance to build the following score:

$$S(X, Y, \delta) = D_e(X, Y) \times H(X, Y, \delta)$$

With these metrics, we applied the technique mentioned above, in other words ordering and filtering:

1. **order** considering $S(X, Y, \delta)$;
2. **filter** considering $D_e(X, Y) > \delta$;
3. **filter** considering $\frac{D_e(X, Y)}{\text{Ref}} > \theta$.

Practically speaking for financial accounts, we considered $\delta = 100$ million euro, and we used as reference value Ref the total financial assets or liabilities of the sector ⁽¹⁴⁾. Given that we receive the data not in euro but in national currency, we convert the 100 million euro into national currencies for each EU Member State using exchange rates from Eurostat’s database.

The reference value is challenging to find. We first thought about using GDP of each EU Member State — which is what was done previously — but it implied necessarily cutting off some series that are very small compared with GDP. Finally, we decided to compute the total assets and liabilities for each institutional sector of the economy, and consider the value of the revisions in these sectors relative to the total economy. We used $\theta = 1\%$.

4.2.2. DETECTION OF OUTLIERS

Let us consider similar notations to those in the previous section. However, this time our interest lies only in identifying y_n as an outlier or not, compared with $\{y_1, \dots, y_{n-1}\}$. We will therefore consider $\mathbb{E}(y_n | \{y_1, \dots, y_{n-1}\})$. We considered a locally estimated scatterplot smoothing (LOESS) to model the time series. With a similar approach to before, let us introduce:

- \bar{Y} the set $\{Y_1, \dots, Y_{n-1}\}$
- the empirical average of \bar{Y} , $\hat{\mu}_Y = \frac{1}{n-1} \sum_{i=1}^{n-1} Y_i$
- the empirical standard deviation of \bar{Y} , $\hat{\sigma}_Y = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n-1} (Y_i - \hat{\mu}_Y)^2}$

⁽¹⁴⁾ The value of δ and θ have been chosen based on empirical knowledge of the domain. The two thresholds are complementary, as δ is the same for all EU Member States, while θ considers the size of the financial sector of each Member State.

With these notations, we applied the same technique as for the revisions coming from Vartak et al. (2015), in other words:

1. **order** considering the z-score of Y_n relatively to the distribution \bar{Y} ,

$$z = \frac{y_n - \hat{\mu}_Y}{\hat{\sigma}_Y};$$

2. **filter** considering $\hat{\sigma}_Y^2 > \varepsilon$;
3. **filter** considering $y_n > \delta$;
4. **filter** considering $\mathbb{E}(y_n | \{y_1, \dots, y_{n-1}\})$.

The first rule ensures that we order based on the distance between the latest observation and the distribution of previous observations. Given that z is conversely proportional to $\hat{\sigma}_Y$, the risk is to select mostly constant series that have a standard deviation close to zero. To avoid this, we consider rule 2 to ensure that we select series with a variance larger than a certain ε . Furthermore, we filtered only the values larger than δ (the same as before), to avoid working with minor series. Finally, we filter out all the observations that belong to a confidence interval built using a regression on \bar{Y} , considering that those are not outliers. We are left with the most interesting series to observe, ordered by z .

5. Results

In this section, we present the output that we developed to answer our research question. As explained previously in Section 3, we developed a report structured in several parts. Here we will make a case study considering real data from the annual financial accounts sent by Belgium. We are comparing data sent in March 2021 for validation with data previously disseminated.

5.1. Overview of the report

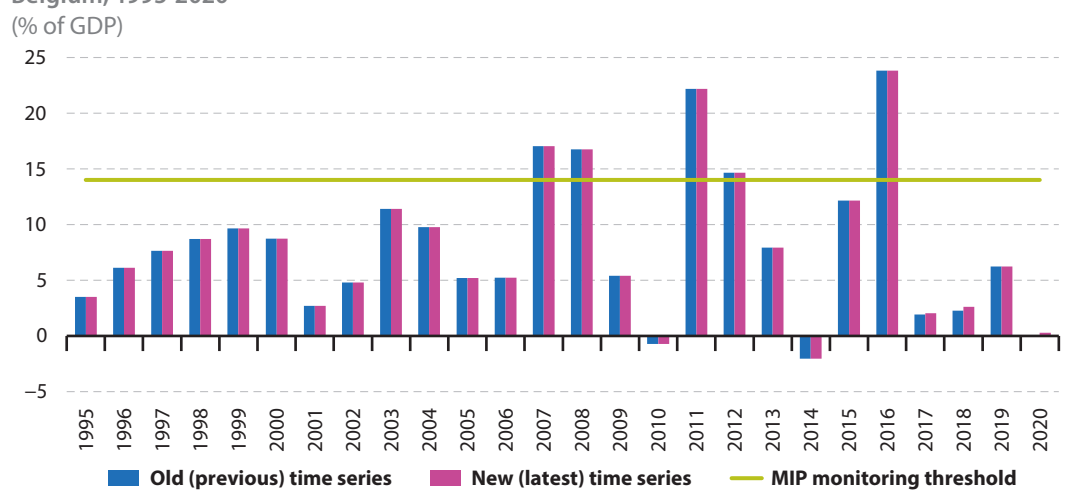
5.1.1. MACROECONOMIC INDICATORS

In this case study, we are dealing with three macroeconomic indicators used in annual financial accounts to monitor the data which are part of the *macroeconomic imbalance procedure* (MIP) indicators. Financial accounts MIP indicators are used for policy purposes. As they are the most used indicators in this domain, they have been selected to be in the report⁽¹⁵⁾. Clearly, other indicators should be used for other domains.

Figures 1 and 2 show the visualisations we chose. We can easily compare the data transmitted with the data previously disseminated. Furthermore, we can simultaneously monitor the data, considering the green-line threshold on the figure. As for all the other figures in our report, these are interactive, meaning that we can zoom in and out, select a specific year or value, and export it as a *.png file if necessary.

⁽¹⁵⁾ See the following Eurostat webpage for details:
<https://ec.europa.eu/eurostat/web/macro-economic-imbalance-procedure/indicators>.

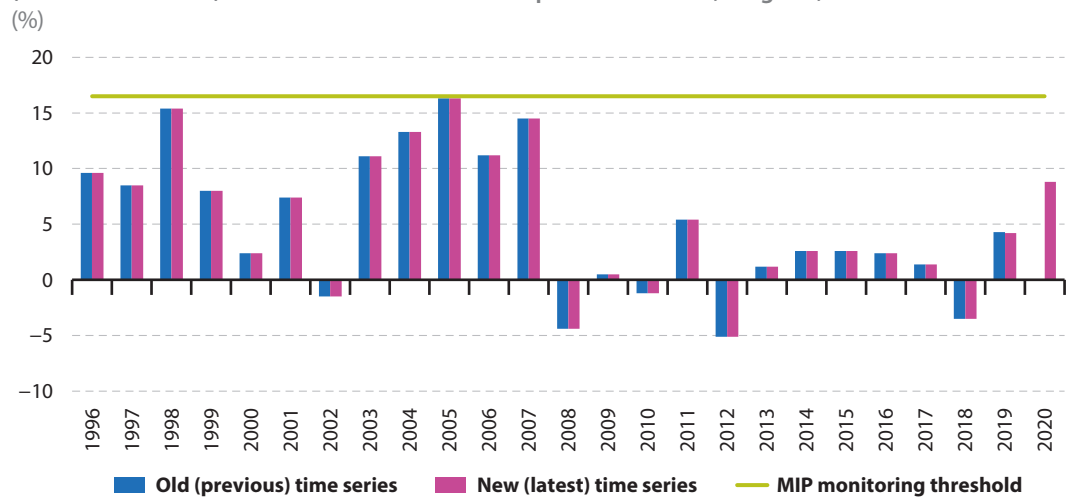
Figure 1: Macroeconomic indicators for annual financial accounts — private sector credit flow, Belgium, 1995-2020



Note: the private sector credit flow represents the net amount of liabilities that have been incurred by non-financial corporations (S.11), households (S.14) and non-profit institutions serving households (S.15) during the year. Old (previous) time series, not available for 2020.

Source: Eurostat (online data code: nama_10_gdp)

Figure 2: Macroeconomic indicators for annual financial accounts — annual change in total (non-consolidated) liabilities of the financial corporations sector, Belgium, 1996-2020



Note: old (previous) time series, not available for 2020.

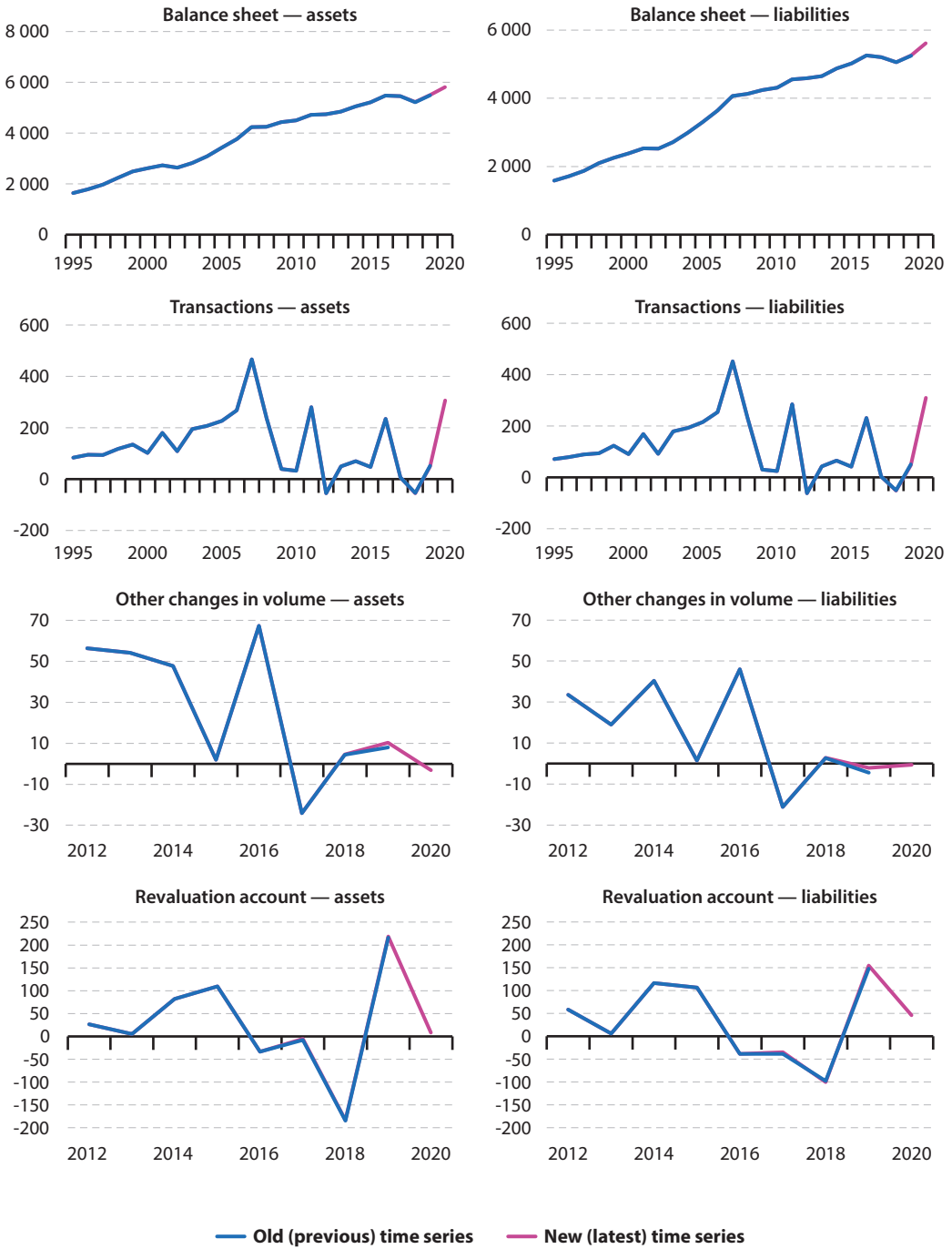
Source: Eurostat

5.1.2. GENERIC VISUALISATIONS

The second part of the report consists of generic visualisations that provide a quick overview of the data, considering only the most aggregated series. Figure 3 shows the eight elements of $F \times A$ in AFA, for the total economy. This section makes it possible to observe very quickly the tendencies of the data in the time series, and to identify if anything particularly unusual happened.

By having a quick overview of the most aggregated series, the user manages to have information on the trend for the most important series within the dataset (which are not excluded from the automatic selection). Trivial errors (for example scalar errors through the use of the wrong unit) are also easily spotted here.

Figure 3: Main series for annual financial accounts (non-consolidated), total financial assets/liabilities, total economy, Belgium, 1995-2020 or 2012-2020 (billion national currency)



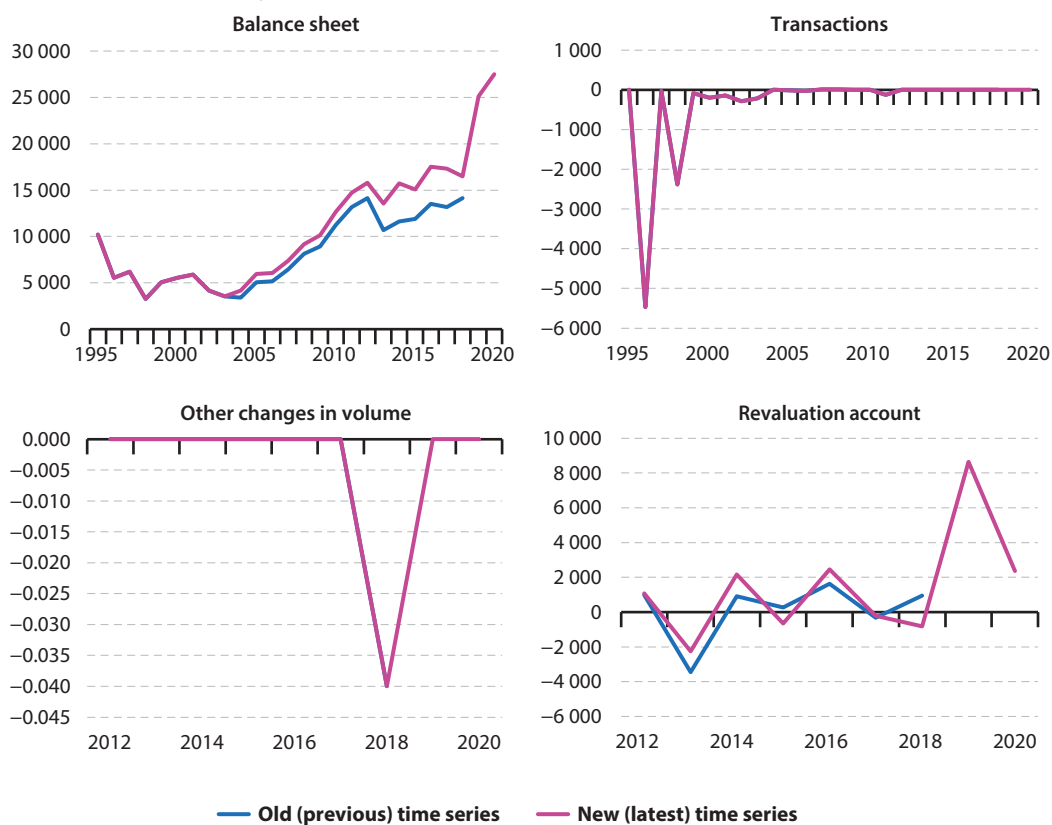
Note: old (previous) time series, not available for 2020.

Source: Eurostat

5.1.3. REVISIONS VISUALISATIONS

In this section, we show the most unusual revisions. In the example in Figure 4, it is possible to see revisions in the balance sheet of the consolidated other equity (F.519) liabilities of the Central Bank. The most recent revisions are triggered by the revaluation accounts. From the example in Figure 5, which shows the revisions of the consolidated listed shares (F.511) assets of the total economy, it is possible to note a feature of the tool which shows more details of a single data point when the cursor is moved over it.

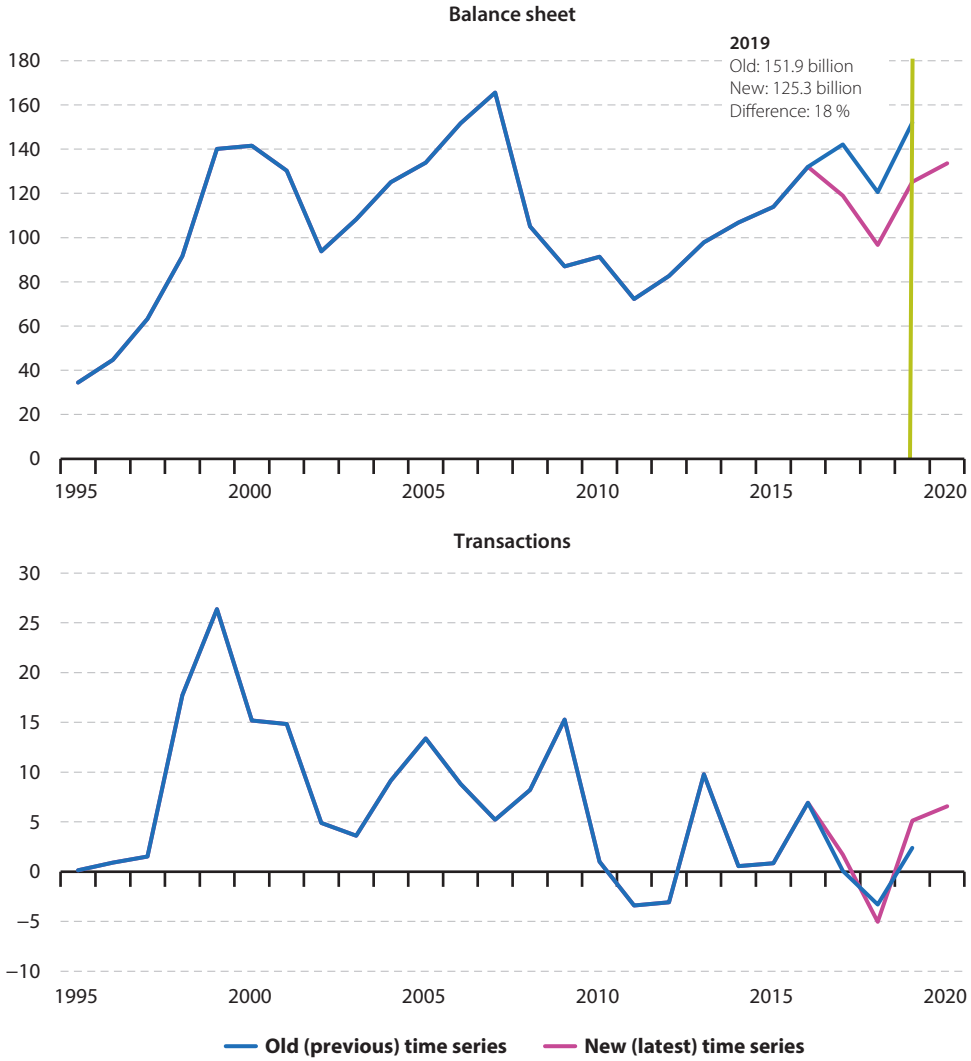
Figure 4: Revisions for annual financial accounts (consolidated), other equity liabilities, central bank, Belgium, 1995-2020 or 2012-2020
(million national currency)



Note: old (previous) time series, not available for 2019 and 2020.

Source: Eurostat

Figure 5: Revisions for annual financial accounts (consolidated), listed shares assets, total economy, Belgium, 1995-2020
(billion national currency)



Note: old (previous) time series, not available for 2020.

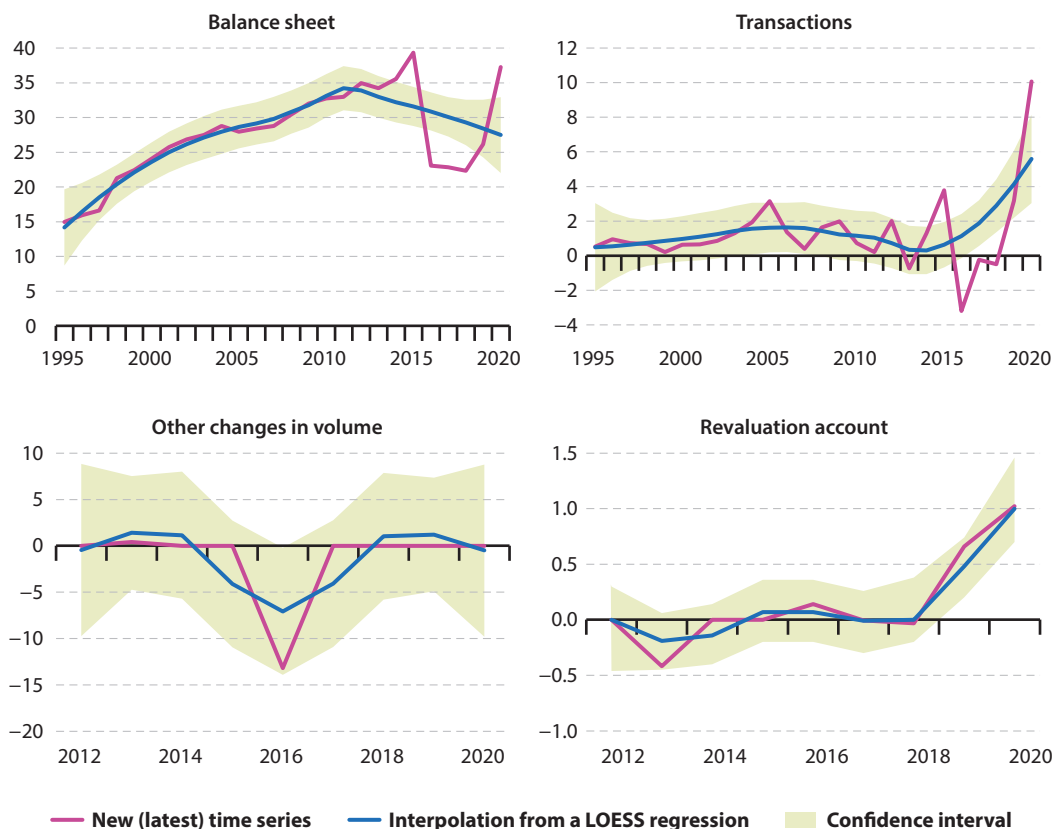
Source: Eurostat

5.1.4. DETECTION OF OUTLIERS

This section shows the bigger outliers detected, such as those presented in Figures 6 and 7.

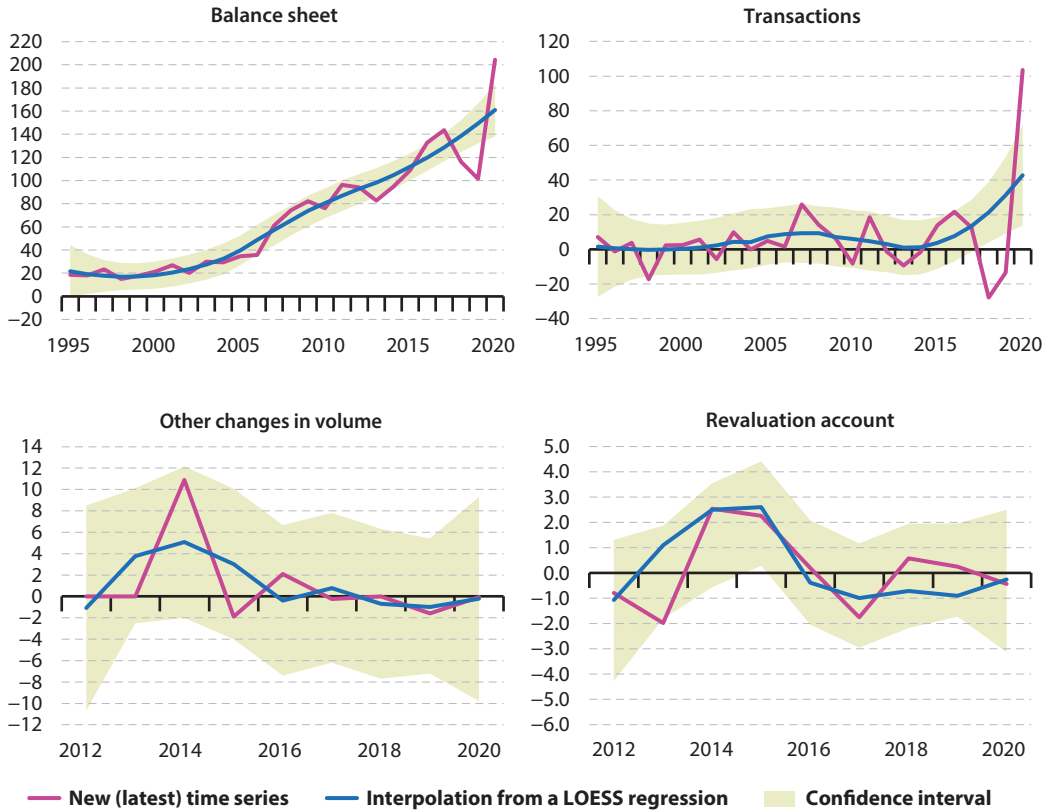
In the example of Figure 6, we can see that the latest value of the non-consolidated transaction non-life insurance technical reserves (F.61) assets of the Central Bank is clearly outside the interval and needs to be investigated. The balance sheet correspondent series is given alongside the sum of the transactions, the other changes in volume and the revaluation accounts; given that the other changes in volume and the revaluation accounts are in line with recent developments, it is (probably) enough to investigate the transactions figure. A similar conclusion can be made for the example shown in Figure 7, concerning the non-consolidated transactions of transferable deposits (F.22) as assets of financial corporations, which shows erratic values for the last three points of the series.

Figure 6: Detection of outliers for annual financial accounts (non-consolidated), non-life insurance technical reserves assets, total economy, Belgium, 1995-2020 or 2012-2020 (billion national currency)



Source: Eurostat

Figure 7: Detection of outliers for annual financial accounts (non-consolidated), transferable deposits assets, financial corporations, Belgium, 1995-2020 or 2012-2020 (billion national currency)



Source: Eurostat

5.2. Cross-domain checks: the case of annual and quarterly financial accounts

In this section we illustrate the cross-domain check presented in Section 3.3, using the case of AFA and QFA data from Germany which were transmitted in October 2020 and April 2021 respectively. As explained, we started by an introductory paragraph which presents the situation, stating the purpose of the report, specifying when the AFA and QFA data were received and indicating the number of errors/inconsistencies found, as shown in Illustration 1. We continued with a tab showing the number of errors grouped by institutional sector and by financial instrument, which helps to identify where errors come from, as shown in Table 1. This tab also included the values found in AFA and in QFA in order to be able to assess immediately the importance of the error and understand the nature of the problem.

Illustration 1: Outline of the report

AFA VS QFA REPORT

Dear colleagues from Germany,

We have received your annual financial accounts transmission on 2020-10-08 13:11:46 and the quarterly data was transmitted to us by the ECB on 2021-04-29. There are internal inconsistencies between the annual and quarterly financial accounts which will need your attention.

While reviewing both datasets we have identified **204** series which show inconsistencies and **0** series which show flag inconsistencies. Please find below details of methodology and the list of series and observations failing. We kindly ask you to review these inconsistencies and transmit the corrected annual dataset to us.

Please address any questions you might have to ESTAT-FINA@ec.europa.eu

Table 1: Cross domain checks — details concerning the inconsistencies for transactions

Value inconsistencies breakdown by sector List of series with value inconsistencies Value inconsistencies by Transactions

Value inconsistencies by Stocks Inconsistencies in observation flags Inconsistencies in confidentiality flags

Here, we checked that $\sum_{i=1}^4 Q_i = A$. The rules we used to flag errors are the following ones:

1. $A - Q > 10$ Millions of National Currency.
2. $\frac{A-Q}{Q} > 1\%$.
3. $\frac{A-Q}{GDP} > 0.3\%$.

Series <chr>	Year <int>	A value <int>	Q value <int>	diff <int>	%diff (AFA-QFA)/QFA <dbl>	diff as % GDP <dbl>
A_S11_F	2013	178191	263973	85782	32.50	3.05
A_S11_F	2018	135578	204028	68450	33.55	2.04
A_S11_F	2016	173084	133176	39908	29.97	1.27
A_S11_F	2017	320864	284663	36201	12.72	1.11
A_S11_F	2012	188718	158695	30023	18.92	1.09
A_S11_F	2014	-11080	13586	24666	181.55	0.84
A_S11_F	2004	6430	18907	12477	65.99	0.55
A_S11_F	2015	187993	173135	14858	8.58	0.49
A_S11_F	2010	165119	174730	9611	5.50	0.37
A_S11_F	2001	133043	125612	7431	5.92	0.34
A_S11_F	2000	325867	332520	6653	2.00	0.32
A_S11_F	2011	115680	124382	8702	7.00	0.32
A_S11_F2	2017	46392	28417	17975	63.25	0.55
A_S11_F2	2013	13196	26382	13186	49.98	0.47
A_S11_F29	2017	21748	2625	19123	728.50	0.59
A_S11_F29	2013	-1859	11326	13185	116.41	0.47
A_S11_F4	2013	-19275	44739	64014	143.08	2.28
A_S11_F4	2013	-35407	11928	47335	396.84	1.68
A_S11_F4	2016	30196	-8952	39148	437.31	1.25
A_S11_F4	2016	35915	2120	33795	1594.10	1.08

Note: A = assets;

S11 = non-financial corporations;

F = total financial assets/liabilities;

F2 = currency and deposits;

F29 = other deposits;

F4 = loans.

AFA = annual financial accounts. QFA = quarterly financial accounts. GDP = gross domestic product.

Source: Eurostat

6. Conclusions

In this article we described the usual practices for data validation and analysis in official statistics, typically performed by statistical offices and international organisations. Building on a general procedure, we worked within a schematic framework where subsequent validation checks are performed on a very big set of time series, where exploration of single time series is impossible due to the limited time in which the validation should be done. We outlined the need for a systematic approach making it possible to process a large volume of time series.

Within such a framework, we developed an innovative data-driven solution for visualising the most relevant time series in terms of revisions and outliers, using a highly interactive application. Computationally speaking, this solution is based on three phases, namely defining a metric to order the data according to what we are looking for, establish rules to filter out inconsequential elements and finally rendering these results in an interactive dashboard. As validation and analysis might need domain specificities, this approach offers users the possibility to change some parameters according to their needs, such as a minimum threshold or the number of periods for which more revisions or outliers can be accepted. A reference time series should still be selected in order to quantify values relatively to this reference, according to the specific domain. While we have seen that GDP can be used as a reference value for many national accounts domains, this is not necessarily true for financial accounts. More generally, this work could be improved by quantifying the impact of the choice of the reference value, and testing the robustness of it.

Finally, we have shown a practical application of the methodology that we developed, based on annual financial accounts, showing that it made possible a very quick identification of the main issues in the dataset, ultimately leading to improvements in data quality. Furthermore, we suggest the introduction of a template for metadata — similar to practices in other domains — and we developed another reporting tool focused on cross-domain checks, which also has the potential to improve data quality.

This work could be taken one step further by fully integrating the metadata exchange during the generation of the dashboard, and most importantly by giving the opportunity to the user to extract content to be sent directly to the statistical reporting agency. Finally, the concepts developed in this article regarding data analysis could serve as the basis for new developments in official statistics with special interests in visualisations, such as [Statistics coded](#).

Acknowledgements

We thank our colleagues Orestis Tsigkas (for his work in non-financial sector accounts from which our work is inspired), Luis Biedma (for his occasional support on development and coding matters) and Martin Turcan (for his valuable ideas). We thank John Verrinder for his advice. Finally, we thank Ani Todorova and Malgorzata Szczęśna for their guidance.

References

- Bay S., K. Kumaraswamy, M. G. Anderle, R. Kumar and D. M. Steier (2006), 'Large scale detection of irregularities in accounting data' in *Sixth International Conference on Data Mining (ICDM'06)*, Institute of Electrical and Electronics Engineers, pp. 75-86.
- Buono D. and E. Infante (2013), 'New technique for predictability, uncertainty, implied volatility and statistical analysis of market risk using SARIMA forecasts intervals' in *New Techniques and Technologies for Statistics 2013*, Eurostat.
- Cohen J. B. (1938), 'The misuse of statistics', *Journal of the American Statistical Association*, Vol. 33, No. 204, pp. 657-674.
- Di Zio M., N. Fursova, T. Gelsema, S. Gießing, U. Guarnera, J. Petrauskienė, L. Quensel-von Kalben, M. Scanu, K. O. ten Bosch, M. van der Loo and K. Walsdorfer (2018), *Methodology for data validation 1.1*, ESS Handbook.
- Eurostat (2014), *European system of accounts — ESA 2010 — Transmission programme of data*, Publications Office of the European Union, Luxembourg.
- Eurostat (2018), *European Statistics Code of Practice*, Publications Office of the European Union, Luxembourg.
- Hao M. C., U. Dayal, D. A. Keim, and T. Schreck (2005), 'Importance-driven visualization layouts for large time series data' in *IEEE Symposium on Information Visualisation, 2005 (InfoVis 05)*, Institute of Electrical and Electronics Engineers, pp. 203-210.
- Kandel S., R. Parikh, A. Paepcke, J. M. Hellerstein and J. Heer (2012), 'Profiler: integrated statistical analysis and visualization for data quality assessment' in *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI'12)*, Association for Computing Machinery, pp. 547-554.
- Lenderink R.J. (2019), *Unsupervised outlier detection in financial statement audits*, Master's thesis, Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente.
- Mackinlay J. (1986), 'Automating the design of graphical presentations of relational information', *ACM Transactions on Graphics*, Volume 5, Issue 2, pp. 110-141.
- R Core Team (2020), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.
- Vartak M., S. Rahman, S. Madden, A. Parameswaran and N. Polyzotis (2015), 'SeeDB: efficient data-driven visualisation recommendations to support visual analytics' in *Proceedings of the VLDB (Very Large Data Base) Endowment*, Volume 8, Number 13, pp. 2 182-2 193.
- Yang S. and R. Cogill (2013), 'Balance sheet outlier detection using a graph similarity algorithm' in *2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER)*, Institute of Electrical and Electronics Engineers, pp. 135-142.
- Zhu T. (2006), 'An outlier detection model based on cross datasets comparison for financial surveillance' in *2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06)*, Institute of Electrical and Electronics Engineers, pp. 601-604.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: <https://europa.eu/european-union/contact>

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: <https://europa.eu/european-union/contact>

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: <https://europa.eu>

EU publications

You can download or order free and priced EU publications from EU Bookshop at: <https://publications.europa.eu/eubookshop>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see <https://europa.eu/european-union/contact>).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: <https://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<https://data.europa.eu/euodp>) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

