EURONA

EUROSTAT REVIEW ON NATIONAL ACCOUNTS AND MACROECONOMIC INDICATORS

1/2017

eurostat

EUROSTAT REVIEW ON NATIONAL ACCOUNTS AND MACROECONOMIC INDICATORS

1/2017

Printed by Imprimerie Centrale in Luxembourg
Manuscript completed in June 2017

Luxembourg: Publications Office of the European Union, 2017

Print	ISSN 2443-7832	Cat.: KS-GS-17-001-EN-C
PDF	ISSN 1977-978X	Cat.: KS-GS-17-001-EN-N

Contents

Editorial 5

Output growth and inflation across space and time
W. Erwin Diewert and Kevin J. Fox 7

Uses of national accounts from the 17th century till present and three suggestions for the future
Frits Bos 41

Processing scanner data by an augmented GUV index
Ludwig von Auer 73

Big data types for macroeconomic nowcasting
Dario Buono, Gian Luigi Mazzi, George Kapetanios, Massimiliano Marcellino and Fotis Papailias 93

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: http://ec.europa.eu/eurostat/web/national-accounts/publications/eurona

Contact: ESTAT-EURONA@ec.europa.eu

Editors

Editorial board

(1) Paulus.Konijn@ec.europa.eu
(2) sartore@unive.it

Editorial

This first EURONA issue of 2017 includes four important articles that take us through history, present and future of macro-economic statistics.

Erwin Diewert and Kevin Fox kick-off this issue with a thorough discussion of methods to aggregate real output and inflation for a group of countries, as well as the long-standing issue of creating real GDP data that are consistent across time and across countries. The results underline the importance of regular cross-country price comparisons, as the authors, using actual OECD data, show that reliance on infrequent comparisons can create significant uncertainty. Fortunately, the International Comparison Program, responsible for the international price and volume comparisons, is now aiming for a future of regular and frequent comparisons.

In the second article of this issue, Frits Bos takes the reader back in history in order to learn for the future. He describes how user needs have shaped national accounts since the 17th century until the present. He shows that fiscal and monetary policy needs have been important drivers of the development of national accounts and their respective international standards, as well as the need to measure and compare (growth of) material living standards. However, policy issues such as welfare, inequality and sustainable development were less reflected in the standards. The author ends with a number of suggestions to ensure that national accounts remain relevant in the future.

Many people believe that the future belongs to big data. In the third article, Ludwig von Auer continues the discussion on the use of multilateral methods for processing scanner data, one type of big data, for consumer price indices that was started by Antonio Chessa in EURONA 1/2016. He argues that the method proposed by Chessa can be simplified in two ways: by using a simpler index formula and a simpler aggregation method. It appears that the last word on multilateral methods for consumer price indices has not yet been written.

Regarding big data in general, Dario Buono, Gian Luigi Mazzi, George Kapetanios, Massimiliano Marcellino and Fotis Papailias present, in the last paper of this issue, a categorisation of big data sources and analyse their potential for macro-economic nowcasting, for example for employment, GDP and inflation. They also include a discussion of the pros and cons of big data for nowcasting and note that the cons should not be ignored.

I think that these articles together give a lot of food for thought and, of course, for continued discussion. In its role as a platform for such discussions, EURONA welcomes any reactions to the articles on ESTAT-EURONA@ec.europa.eu.

Paul Konijn

Editor of EURONA

1 Output growth and inflation across space and time

W. ERWIN DIEWERT [1] AND KEVIN J. FOX [2]

Abstract: It is common for comparisons to be made of output growth and inflation across groups of countries, yet such comparisons can result in inconsistencies. We address two issues: (i) how to measure aggregate real output and inflation for groups of countries and (ii) how to construct measures of real GDP for a group of countries where the country measures of real GDP are consistent across time and space. A method is proposed for harmonizing conflicting estimates of OECD member-country real GDP, ensuring consistency over space and overall group consistency over time. A new measure of OECD inflation is also proposed.

JEL codes: C43, C82, E01

Keywords: Purchasing power parities, ICP, inflation, price and volume indexes, Fisher indexes

[1] Vancouver School of Economics, University of British Columbia and UNSW Sydney, Australia.
[2] School of Economics, UNSW Sydney, Australia.

'Econometricians have an ambivalent attitude towards economic data. At one level, the 'data' are the world that we want to explain, the basic facts that economists purport to elucidate. At the other level, they are the source of all our troubles.'

Zvi Griliches (1985, p. 196)

1. Introduction

Providing consistent estimates of real GDP across countries and time is important for many policy-relevant purposes, such as assessing convergence of living standards; see Eurostat (2012) and the World Bank (2013). The OECD publishes estimates of Purchasing Power Parities (PPPs) on an annual basis and these PPPs can be used to generate estimates of real GDP for member countries that are comparable across countries for the given year. However, the resulting estimates of relative real GDP are typically inconsistent with national estimates of real GDP growth for the member countries.

We use OECD data for the years 2000-2012 in order to study two issues. First, what is the optimal approach to constructing estimates of OECD aggregate real GDP and associated measures of aggregate OECD inflation? Index number theory is used to decompose national nominal GDP into price and quantity (or volume) components, but constructing estimates of inflation and real GDP growth for a group of countries that use different currencies is a more complicated operation ([3]).

Second, how can PPP information be optimally used in conjunction with country data on real GDP growth to construct estimates of OECD member country real GDP that are in principle simultaneously comparable across space and time? Using our proposed solution as a benchmark, we show that if PPP data are available only infrequently, as is the case for the World Bank provided PPP data used in the Penn World Tables, then estimates can differ considerably as new PPP information becomes available. This is rather inconvenient: studies of competitiveness and living standards convergence across countries will want to use real GDP series that are not subject to violent revision.

In Section 2, we study the first measurement issue using just national data and exchange rate information. Sections 3-6 use the OECD PPP data and study the second measurement issue (and revisit the first issue). In Section 3, we propose a harmonized method for constructing estimates of member country GDP volumes that are comparable across time and space. In Sections 4 and 5, we compare our harmonized estimates with estimates of member country real GDP that are generated by using PPP data for only one year. These base year estimates are then projected to all other years using each country's own growth rates of GDP. Section 4 uses the PPPs for 2000 and Section 5 uses the PPPs for 2012. We find that the resulting two panel sets of real GDP estimates are very different from each other and they also are very different from our harmonized estimates developed in Section 3. Section 6 considers the context where PPPs

([3]) 'At the national level, current price (value) data can typically be decomposed into a volume (or quantity) series and price series. At the international level, a second 'price' component enters the picture in the form of a conversion rate from the domestic to a common currency. The implication is that values can be expressed at current market exchange rates (or current international prices, if purchasing power parities — PPPs — are used); and at constant exchange rates (or constant international prices).' OECD (2001, p. 6).

are only available infrequently, as is the case with World Bank provided PPPs ([4]). We use PPP information for 2000 and 2012 to provide interpolated estimates of volumes for each country, finding that the current interpolation method implemented in the widely used Penn World Tables did not work well with our OECD data base. In contrast, we find that our proposed method produces estimates that are much closer to our preferred harmonized estimates of Section 3. Section 7 concludes this paper.

2. OECD growth and inflation using market exchange rates

Our first measure of aggregate GDP growth over the member countries in the OECD during the years 2000-2012 uses national growth rates of GDP and domestic prices converted into US dollar at market exchange rates. The aggregation principle used to form OECD aggregate GDP volumes and prices in this section is the same that is used to aggregate prices and quantities across different regions in a country: each commodity in each region is regarded as a separate commodity in the index number formula. In what follows, we use the OECD ordering of countries, which is as in Table 1.

Table 1: OECD country codes

1 = Australia	10 = France	19 = Korea	28 = Slovenia
2 = Austria	11 = Germany	20 = Luxembourg	29 = Spain
3 = Belgium	12 = Greece	21 = Mexico	30 = Sweden
4 = Canada	13 = Hungary	22 = Netherlands	31 = Switzerland
5 = Chile	14 = Iceland	23 = New Zealand	32 = Turkey
6 = Czech Republic	15 = Ireland	24 = Norway	33 =U.K.
7 = Denmark	16 = Israel	25 = Poland	34 = United States
8 = Estonia	17 = Italy	26 = Portugal	
9 = Finland	18 = Japan	27 =Slovak Republic	

The country values for nominal GDP in the national currencies for the years 2000-2012 can be obtained from the OECD online data base, OECD.Stat ([5]). We convert these estimates into billions and denote the estimate for country n in year t by V_n^t. The corresponding volume estimates can be obtained from the same source ([6]), and we similarly convert these estimates into billions and denote these volumes (or quantities) by Q_n^t for $n = 1,\ldots,34$ and $t = 2000, \ldots,2012$. The corresponding country price level for country n in year t is defined as $P_n^t \equiv V_n^t/Q_n^t$ for $n = 1,\ldots,34$ and $t = 2000,\ldots,2012$. These national price levels and volumes are listed in the Annex; see Tables A1 and A2.

Since the country volumes Q_n^t are measured in domestic currency units (which are not comparable across countries), we need to convert the annual domestic nominal values of GDP

([4]) For example, the International Comparison Program (ICP) of the World Bank provided PPPs for 155 countries for the year 2005 and has more recently provided a new set of PPPs for 2011 covering 177 countries. The ICP methodology is explained in World Bank (2013). How can these two benchmark sets of PPPs be used in conjunction with national data in order to provide estimates of country real GDP that are comparable across all years from 2005 to 2012? The interpolation method explained in Section 6 could be used in this context.

([5]) OECD.Stat Table B1_GE: Gross domestic product (expenditure approach); National currency, current prices, millions, annual data.

([6]) OECD.Stat Table B1_GE: Gross domestic product (GDP); National currency, constant prices, national base year, millions, annual data.

into common currency units using the average exchange rates for the corresponding year. In principle, the numeraire country could be any of the 34 OECD countries but it seems reasonable to choose the largest country. The OECD has conveniently done this for us, converting each country's nominal GDP into US dollar at the average market exchange rates for the given year ([7]). We convert these estimates into billions and denote the US dollar estimate for nominal (current price) GDP for country n in year t by v_n^t.

The year t, country n US dollar price level for GDP, p_n^t, is initially defined as v_n^t/Q_n^t where the country volumes (or real outputs) Q_n^t have already been defined using national data. The resulting p_n^t were normalized so that $p_n^{2000} = 1$ for $n = 1,\ldots,34$. The Q_n^t were then normalized in the opposite direction so that US dollar values were preserved. Denote the resulting normalized Q_n^t as q_n^t for $n = 1,\ldots,34$ and $t = 2000,\ldots,2012$ ([8]). These US dollar price levels p_n^t and the corresponding volumes q_n^t are listed in Tables A3 and A4 in the Annex.

We are now in a position to calculate annual aggregate OECD real output and the corresponding price level for each of the years 2000-2012 using the price and volume data, p_n^t and q_n^t as inputs into an index number formula. It will be useful to consider the choice of index number formula in the context of providing index levels for two periods, say periods 0 and 1 ([9]). Suppose there are n commodities to be aggregated. Denote the price and quantity vectors for period t by $p^t \equiv [p_1^t,\ldots, p_N^t]$ and $q^t \equiv [q_1^t,\ldots,q_N^t]$ for $t = 0,1$. The value of transactions in the N commodities during period t is defined as $v^t \equiv \sum_{n=1}^{N} p_n^t q_n^t \equiv p^t \bullet q^t$ ([10]). The problem of choosing functional forms for the price and quantity indexes is usually phrased as follows: find two suitable functions of $4N$ variables (i.e. prices and quantities for each good in each period), a price index function $P(p^0, p^1,q^0,q^1)$ and a quantity index function $Q(p^0, p^1,q^0,q^1)$, such that the product of these two functions is equal to the value ratio, v^1/v^0. Thus the functions P and Q are to satisfy the following equation:

(1) $p^1 \bullet q^1/p^0 \bullet q^0 = P(p^0, p^1,q^0,q^1)Q(p^0, p^1,q^0,q^1)$.

It can be seen that if the functional form for either the price or quantity index is determined then the functional form of the corresponding quantity or price index is also determined using equation (1) ([11]).

Two natural choices for the functional form for the price index are the well-known Laspeyres and Paasche price indexes, P_L and P_P, defined as follows ([12]):

(2) $P_L(p^0, p^1,q^0,q^1) \equiv p^1 \bullet q^0/p^0 \bullet q^0$;

(3) $P_P(p^0, p^1,q^0,q^1) \equiv p^1 \bullet q^1/p^0 \bullet q^1$.

Using (1), it can be seen that quantity indexes that correspond to P_L and P_P are Q_P and Q_L, respectively, defined as follows:

([7]) OECD.Stat Table B1_GE: Gross domestic product (expenditure approach); US dollar, current prices, current exchange rates, millions, annual data.

([8]) Note that $v_n^t = p_n^t q_n^t$ for $n = 1,\ldots,34$ and $t = 2000,\ldots,2012$. For each n, the US dollar volumes q_n^t are proportional to the national volumes Q_n^t; i.e., we have $q_n^t = \lambda_n Q_n^t$ for $t = 2000,\ldots,2012$ for each country n where λ_n is the factor of proportionality for country n.

([9]) For materials on the historical development of index number theory, see Diewert (1993) and Balk (2008).

([10]) The inner product of two vectors $x \equiv [x_1,\ldots,x_N]$ and $y \equiv [y_1,\ldots,y_N]$ of the same dimension N is defined as $x \bullet y \equiv \sum_{n=1}^{N} x_n y_n$.

([11]) Once P and Q satisfying (1) have been chosen, the corresponding price levels for periods 0 and 1, say P^0 and P^1, and the corresponding quantity (or volume) levels for periods 0 and 1, say Q^0 and Q^1, are generally determined as follows: $P^0 \equiv 1$; $P^1 \equiv P(p^0, p^1,q^0,q^1)$; $Q^0 \equiv v^0 = p^0 \bullet q^0$ and $Q^1 \equiv v^0 Q(p^0, p^1,q^0,q^1) = v^1/ P(p^0, p^1,q^0,q^1)$. Note that the price and quantity indexes can be interpreted as ratios of aggregate price and quantity levels; i.e., we have $P(p^0, p^1,q^0,q^1) = P^1/P^0$ and $Q(p^0, p^1,q^0,q^1) = Q^1/Q^0$.

([12]) It can be seen that the Laspeyres price index uses the 'basket' of period 0 quantities, q^0, and prices out this basket at the prices of period 0 (in the denominator) and prices out the same basket at the prices of period 1 (in the numerator) and takes the ratio of these costs as the price index. The Paasche index is similar but uses the 'basket' of period 1 quantities, q^1, as the common quantity vector in the numerator and denominator.

(4) $Q_P(p^0, p^1, q^0, q^1) \equiv p^1 \cdot q^1 / p^1 \cdot q^0$;

(5) $Q_L(p^0, p^1, q^0, q^1) \equiv p^0 \cdot q^1 / p^0 \cdot q^0$.

The Paasche and Laspeyres price and quantity indexes are equally plausible. The problem is that they can generate quite different estimates of inflation and growth. A natural solution to this problem is to take a symmetric average of these two estimates; taking the geometric mean of these two price indexes (and of the two corresponding quantity indexes) leads to indexes that have desireable axiomatic properties ([13]). These are Fisher (1922) ideal price and quantity indexes, P_F and Q_F, defined as follows ([14]):

(6) $P_F(p^0, p^1, q^0, q^1) \equiv [P_L(p^0, p^1, q^0, q^1) P_P(p^0, p^1, q^0, q^1)]^{1/2}$;

(7) $Q_F(p^0, p^1, q^0, q^1) \equiv [Q_L(p^0, p^1, q^0, q^1) Q_P(p^0, p^1, q^0, q^1)]^{1/2}$.

There is one more choice that needs some discussion: namely, should fixed base or chained Fisher indexes be used when aggregating over many periods? The chain system measures the change in prices going from one period to a subsequent period using a bilateral index number formula involving the prices and quantities pertaining to the two adjacent periods ([15]). These one period rates of change (the links in the chain) are then cumulated to yield the relative levels of prices over the entire period under consideration. If the bilateral price index is P, the chain system generates the following sequence of price levels for the first three periods:

(8) $1, P(p^0, p^1, q^0, q^1), P(p^0, p^1, q^0, q^1)\, P(p^1, p^2, q^1, q^2)$.

The fixed base system of price levels using the same bilateral index number formula P simply computes the level of prices in period t relative to the base period 0 as $P(p^0, p^t, q^0, q^t)$. The fixed base sequence of price levels for periods 0, 1 and 2 is:

(9) $1, P(p^0, p^1, q^0, q^1), P(p^0, p^2, q^0, q^2)$.

There are two major problems associated with the use of fixed base indexes in the context of annual time series data: (i) over longer periods of time, it becomes more difficult to match up products in the current period with the corresponding products in a distant base period, leading to less accurate index numbers; and (ii) fixed base indexes are subject to revisions (that can be substantial) when the base period is finally changed. When using fixed base Paasche or Laspeyres indexes, the revision problem can become massive ([16]). Thus a major advantage of the chain system in the context of aggregating annual data is that chaining will reduce the spread between the Paasche and Laspeyres indexes ([17]). These two indexes each provide an asymmetric perspective on the amount of price change that has occurred between the two periods under consideration and it could be expected that a single point estimate of the

([13]) See Fisher (1922) and Diewert (1992) (1997).
([14]) It can be verified that $P_F Q_F = v^1/v^0$; i.e., the Fisher price and quantity indexes satisfy equation (1).
([15]) The chain principle was introduced independently into the economics literature by Lehr (1885, pp. 45–46) and Marshall (1887, p. 373). Both authors observed that the chain system would mitigate the difficulties due to the introduction of new commodities into the economy, a point also mentioned by T.P. Hill (1993, p. 388). Fisher (1911, p. 203) introduced the term 'chain system'.
([16]) The US Bureau of Economic Analysis used to provide long term estimates of US GDP back to 1926 using fixed base Laspeyres volume indexes. When the base year was changed, the resulting Laspeyres estimates of real GDP growth changed massively, prompting the BEA to switch to chained Fisher indexes in the early 1990s.
([17]) See Diewert (1978, p. 895) and T.P. Hill (1988) (1993, pp. 387–388).

aggregate price change should lie between these two estimates. Chaining will usually lead to a smaller difference between the two and hence lead to estimates that are closer to the 'truth' ([18]).

For year $t = 2001,\ldots,2012$, denote the chained Fisher aggregate OECD volume level for by Q^t and the corresponding US dollar price level by P^t, and define the *OECD volume growth rate* γ^t and the corresponding OECD *US dollar inflation rate* ρ^t in percentage points as follows:

(10) $\gamma^t \equiv 100[(Q^t/Q^{t-1}) - 1]$;

(11) $\rho^t \equiv 100[(P^t/P^{t-1}) - 1]$.

The chained Fisher OECD aggregate price and volume levels, P^t and Q^t, for the years 2000-2012 are listed in Table 2 along with the corresponding percentage point annual growth rates, ρ^t and γ^t, for the years 2001-2012. For comparison purposes, we also provide the aggregate OECD chained Laspeyres and Paasche indexes, $P_L^{\,t}$ and $P_P^{\,t}$, over the same period. It can be seen that the chained Fisher, Laspeyres and Paasche price levels are all very close to each other so that for this particular application, the choice of index number formula does not matter very much.

Table 2: OECD annual aggregate volumes Q^t and price levels in US dollar P^t, $P_L^{\,t}$ and $P_P^{\,t}$, price levels in euro $P_{EU}^{\,t}$, PPP price levels $P_{PPP}^{\,t}$ and percentage point changes, 2000-2012

Year t	Q^t	P^t	$P_L^{\,t}$	$P_P^{\,t}$	γ^t	ρ^t	$\rho_{EU}^{\,t}$	$P_{EU}^{\,t}$	$P_{PPP}^{\,t}$
2000	26 694.3	1.000	1.000	1.000				1.000	1.000
2001	27 022.9	0.979	0.979	0.980	1.23	– 2.06	0.84	1.008	1.030
2002	27 432.9	1.008	1.008	1.008	1.52	2.92	– 2.14	0.987	1.055
2003	28 007.3	1.108	1.108	1.108	2.09	9.90	– 8.35	0.904	1.080
2004	28 896.6	1.193	1.193	1.193	3.18	7.71	– 2.10	0.885	1.107
2005	29 670.9	1.227	1.227	1.227	2.68	2.84	2.68	0.909	1.133
2006	30 566.7	1.256	1.256	1.256	3.02	2.35	1.46	0.922	1.161
2007	31 374.2	1.339	1.339	1.340	2.64	6.63	– 2.27	0.902	1.189
2008	31 410.0	1.411	1.410	1.411	0.11	5.36	– 1.56	0.887	1.217
2009	30 267.1	1.373	1.373	1.373	– 3.64	– 2.69	2.60	0.911	1.231
2010	31 138.6	1.401	1.401	1.402	2.88	2.07	7.06	0.975	1.248
2011	31 688.5	1.478	1.477	1.478	1.77	5.44	0.46	0.979	1.270
2012	32 162.6	1.453	1.453	1.454	1.50	– 1.63	6.42	1.042	1.289

Source: Author's calculations

The sample average of the year to year growth rates for OECD real GDP using US dollar weights, the γ^t, was 1.58 % per year. It can be seen that there was only one year where OECD real growth was negative: 2009 (– 3.64 %). The sample average of the OECD inflation rates ρ^t (measured in US dollar at market exchange rates) was 3.24 % per year. However, what is striking is the variability of the US dollar inflation rates.

([18]) There is a more elaborate justification for chaining annual data that is based on aggregating over observations that have the most 'similar' price structures; see R.J. Hill (2001), (2004), (2009), Diewert (2009) and Diewert and Fox (2017). Typically, adjacent annual observations will have more similar price structures than a pair of observations chosen from two distant periods. However, it is not always best to use chained indexes. T.P. Hill (1993, p. 388), drawing on the earlier research of Szulc (1983) and T.P. Hill (1988, pp. 136–137), noted that it is not appropriate to use the chain system when prices oscillate or 'bounce' to use Szulc's (1983, p. 548) terminology. This bouncing phenomenon can occur when aggregating subannual data when there are seasonal fluctuations or periodic sales (with deeply discounted prices). However, in the context of more or less smoothly trending prices and quantities as is the usual case using annual data, T.P. Hill (1993, p. 389) recommended the use of chained symmetrically weighted indexes such as the Fisher ideal index. Thus in this paper, we will use chained Fisher indexes when aggregating over countries.

The principles used to construct our OECD aggregate measures of real GDP, Q^t, are the same principles typically used to construct country wide estimates of real GDP within a country. Estimates of real GDP aggregate output growth over regions within the country use regional price levels as weights for the regional volumes. In constructing national estimates of real output, the national statistician does not assume that the quantities (or volumes) in each region are comparable across regions; all that is assumed is that whatever is being measured at the regional level *is comparable over time*. This is the same principle that is being used to construct the above OECD real output measures Q^t; there is no assumption that the country units of measurement are comparable across countries.

The one difference in our suggested method for constructing OECD real GDP as opposed to methods used to construct national estimates of real GDP is that we needed to convert national values of GDP into a common currency using annual average market exchange rates. We chose to make this conversion using US dollar as the numeraire currency. If we chose another currency to be the numeraire currency, the unit of measurement would change, but the overall OECD *growth rates* for real GDP would remain the same; i.e., the γ^t listed in Table 2 *do not change* if we converted all country nominal GDP estimates into a different common currency at annual average market exchange rates and then applied the same methodology to construct the overall OECD volume estimates ([19]). However switching to a different numeraire currency dramatically affects the inflation rates ρ^t; the OECD aggregate price level estimates P^t and the resulting inflation rates ρ^t defined by (11) change depending on the chosen numeraire currency.

In order to illustrate the dependence of the above OECD GDP inflation rates on the choice of the numeraire country, we computed the aggregate OECD price and volume levels, $P_{EU}{}^t$ and $Q_{EU}{}^t$, using Germany as the numeraire country. Thus instead of using the US dollar estimates for nominal GDP for country n in year t defined earlier by $v_n{}^t$, for euro zone countries we used the reported national value estimates of GDP. For non-euro zone countries, we converted the $v_n{}^t$ into euro using the implied OECD exchange rate that can be obtained by dividing the national value estimate of GDP for Germany (or any other euro zone country) by the corresponding US dollar measure. The same Fisher index number methodology was then used to construct $P_{EU}{}^t$ and $Q_{EU}{}^t$. The resulting euro based price index $P_{EU}{}^t$ and inflation growth rates $\rho_{EU}{}^t \equiv 100[(P_{EU}{}^t / P_{EU}{}^{t-1}) - 1]$ are listed in Table 2. Comparing the inflation measures using the US and then Germany as the numeraire countries shows that the resulting price levels, P^t and $P_{EU}{}^t$, and inflation rates, ρ^t and $\rho_{EU}{}^t$, are *completely different*. The price levels are plotted in Figure 1 to clearly show the extent of their divergence; P^t trends upward from 1.00 in 2000 to end up at 1.45 in 2012 whereas the euro based OECD price level trends downward to 0.89 in 2008 and then trend upward to end up at 1.04 in 2012. The explanation for these diverging results is simple: they are driven by large exchange rate movements over the sample period ([20]).

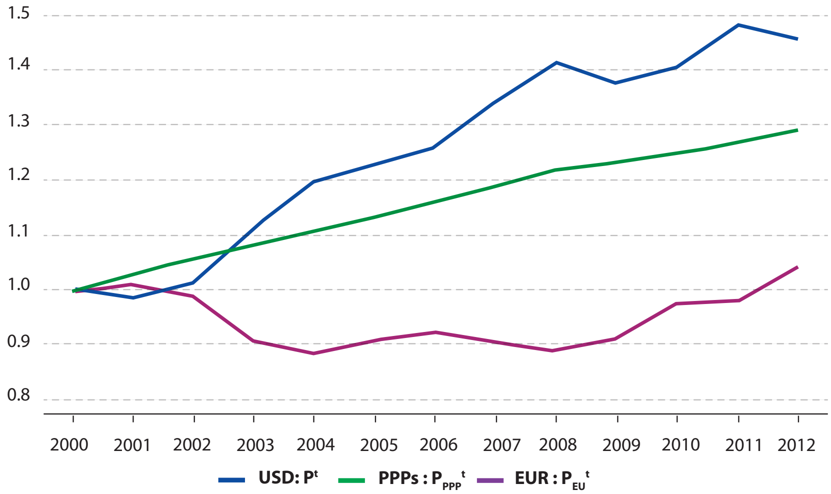
Our conclusion at this point is that our *first approach* to measuring OECD real output and inflation using national GDP data and market exchange rates is (perhaps) satisfactory for measuring real output but that it is *not* satisfactory for measuring inflation. A satisfactory inflation measure will be introduced in the following section when we introduce our *second approach* to measuring aggregate OECD inflation. The preferred price levels from this approach ($P_{PPP}{}^t$) are also listed in Table 2 and plotted in Figure 1 for ease of comparison with the corresponding exchange rate based series ([21]).

([19]) In order for this statement to be true, we need our chosen bilateral index number formula to satisfy the following two tests: $Q(\lambda p^0, p^1,q^0,q^1) = Q(p^0, p^1,q^0,q^1)$ for all scalars $\lambda > 0$ and $Q(p^0,\lambda p^1,q^0,q^1) = Q(p^0, p^1,q^0,q =)$ for all scalars $\lambda > 0$. The Fisher, Laspeyres and Paasche bilateral quantity indexes all satisfy these homogeneity-in-prices properties.

([20]) US prices in terms of euro declined markedly from 2000 to 2008 and this explains the large number of negative $\rho_{EU}{}^t$ over this period; the number of euro it took to buy one US dollar in 2000, 2008 and 2012 was 1.085, 0.683 and 0.778, respectively.

([21]) These will be discussed in the following section, where the corresponding inflation rate series are plotted in Figure 2.

Figure 1: Alternative OECD price levels in US dollar, euro and PPPs
(Index)



Source: Author's calculations

The analysis presented in this section made no assumption that the goods and services produced in any country were comparable to the goods and services produced in any other country. In the following section, it will be assumed that the goods and services produced in each country *are comparable across countries* and different measures of OECD growth and inflation will be derived.

3. OECD growth and inflation measurement using annual PPP information

The OECD (in close cooperation with Eurostat) produces an annual series of Purchasing Power Parities (PPPs) that enable the comparison of real GDP of member countries with each other ([22]). For each OECD country n and each year t, PPP_n^t is an estimate of the number of units of the national currency of country n that is required to purchase one dollar of US GDP in year t ([23]). We divide the country n nominal value of GDP in year t in domestic currency, V_n^t, by the corresponding PPP_n^t in order to obtain an estimate, r_n^t, of country n's real GDP in year t in units that are comparable across countries for year t ([24]):

(12) $r_n^t \equiv V_n^t/PPP_n^t\,;\ n = 1,\ldots,34;\ t = 2000,\ldots,2012.$

([22]) The construction of these PPPs is explained in the Eurostat and OECD PPP Manual; see Eurostat (2012). The International Comparison Program (ICP) of the World Bank constructed PPPs for over 150 countries for 2005 and 2011. The ICP methodology is explained in World Bank (2013).
([23]) OECD.Stat, Table 4: PPPs and Exchange Rates; PPPGDP; Purchasing Power Parities for GDP; National currency per US dollar; Annual; 2000-2012. This Table is reproduced in the Annex as Table A5.
([24]) These relative GDP volume measures for year t are not comparable across years.

Once the r_n^t have been calculated, they can be summed so that $r^t \equiv \sum_{n=1}^{34} r_n^t$ and then the year t country n share of OECD real output can be defined as follows [25]:

(13) $s_n^t \equiv r_n^t / r^t$; $n = 1,\ldots,34$; $t = 2000,\ldots,2012$.

These country shares of OECD real GDP are listed in Table 3, which enables the comparison of GDP volumes across all OECD countries for each year [26]. Note that country 34, the US, has the largest share (around 35-37 %), followed by country 18, Japan, (10-11 %) and country 11, Germany (7 %).

We can use this information to construct estimates of overall real GDP growth and inflation across OECD countries. A natural method is to use the country shares in Table 3 as weights for national rates of growth of real GDP. The year t growth factor for country n can be defined as Q_n^t / Q_n^{t-1} where Q_n^t is country n's GDP volume in year t, and the OECD Laspeyres type growth factor (or chain link) for year t, Γ_L^t, as the following weighted average of the national growth factors:

(14) $\Gamma_L^t \equiv \sum_{n=1}^{34} s_n^{t-1} (Q_n^t / Q_n^{t-1})$; $t = 2001,\ldots,2012$.

The measure of OECD GDP volume growth defined by (14) is the method used by the OECD to calculate their official measure of OECD volume growth. It certainly is a sensible measure, using country (one plus) growth rates going from year $t-1$ to year t, Q_n^t / Q_n^{t-1}, weighted by the country real volume shares s_n^{t-1} for year $t-1$, which were derived using PPPs. However, the above formula suffers from the same problem that the standard Laspeyres formula has: namely, it does not treat the periods in a symmetric fashion. The counterpart to the Laspeyres-type formula defined by (14) is the following *Paasche-type formula* [27]:

(15) $\Gamma_P^t \equiv [\sum_{n=1}^{34} s_n^t (Q_n^t / Q_n^{t-1})^{-1}]^{-1}$; $t = 2001,\ldots,2012$.

The corresponding Fisher-type formula for OECD volume growth for year t is defined as follows [28]:

(16) $\Gamma_F^t \equiv [\Gamma_L^t \, \Gamma_P^t]^{1/2}$; $t = 2001,\ldots,2012$.

[25] Note that the country shares s_n^t can be constructed without using country exchange rates (in principle). Using definitions (12) and (13), it can be seen that the s_n^t can be written in the following form: $s_n^t = [V_n^t / PPP_n^t] / [\sum_{i=1}^{N} (V_i^t / PPP_i^t)]$ for all n and t. Compare these 'real' shares s_n^t with the corresponding country US dollar shares $S_n^t = [V_n^t / e_n^t] / [\sum_{i=1}^{N} (V_i^t / e_i^t)]$ defined in the Annex. All of the measures derived in this section are independent of country exchange rates.

[26] Row *n+1* in the Table gives the shares for country n where we use the standard ordering of OECD countries listed in the previous section. Since the PPPs used by the OECD are invariant to the choice of the numeraire country (up to a scalar factor), it can be verified that the country shares listed in Table 3 are also invariant to the choice of numeraire country.

[27] Suppose that there is only one homogeneous commodity in each country's GDP. Then the volume for country n in year t, Q_n^t, should be equal to the number of units of this homogeneous commodity. Under these conditions, it can be seen that both Γ_L^t and Γ_P^t equal $\sum_{n=1}^{34} Q_n^t / \sum_{n=1}^{34} Q_n^{t-1}$.

[28] If the PPPs are independent of the choice of the numeraire country (up to a scalar factor), then the growth factors, $\Gamma_L^t \, \Gamma_P^t$ and Γ_F^t will not depend on the choice of the numeraire country.

The growth factors (or chain link indexes) defined by (14)-(16) can be multiplied together to generate OECD aggregate volume levels. The growth factors can also be transformed into growth rates, γ_L^t, γ_P^t and γ_F^t (in percentage points), by using the following definitions for $t = 2001,\ldots,2012$:

(17) $\gamma_L^t \equiv 100[\Gamma_L^t - 1]\,;\, \gamma_P^t \equiv 100[\Gamma_P^t - 1]\,;\, \gamma_F^t \equiv 100[\Gamma_F^t - 1]$.

Table 3: Country shares of OECD real GDP 2000-2012

n	s_n^{2000}	s_n^{2001}	s_n^{2002}	s_n^{2003}	s_n^{2004}	s_n^{2005}	s_n^{2006}	s_n^{2007}	s_n^{2008}	s_n^{2009}	s_n^{2010}	s_n^{2011}	s_n^{2012}
1	0.019	0.019	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.022	0.022	0.022	0.023
2	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
3	0.010	0.010	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.010	0.010	0.010	0.010
4	0.031	0.031	0.031	0.031	0.031	0.032	0.031	0.031	0.031	0.031	0.031	0.032	0.031
5	0.005	0.005	0.005	0.005	0.006	0.006	0.007	0.007	0.007	0.007	0.008	0.008	0.008
6	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.007	0.006	0.006	0.006
7	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
8	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
9	0.005	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005
10	0.054	0.055	0.055	0.053	0.052	0.052	0.052	0.052	0.052	0.053	0.052	0.052	0.052
11	0.074	0.074	0.074	0.073	0.072	0.072	0.072	0.072	0.073	0.071	0.072	0.073	0.073
12	0.007	0.007	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.006
13	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.005	0.005	0.004	0.004	0.004	0.004
16	0.005	0.005	0.005	0.005	0.005	0.005	0.004	0.005	0.004	0.005	0.005	0.005	0.005
17	0.051	0.052	0.050	0.049	0.047	0.046	0.047	0.047	0.048	0.047	0.046	0.045	0.044
18	0.115	0.114	0.113	0.112	0.111	0.109	0.106	0.105	0.103	0.099	0.101	0.098	0.097
19	0.028	0.029	0.030	0.030	0.031	0.031	0.031	0.031	0.031	0.032	0.033	0.034	0.033
20	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
21	0.035	0.034	0.034	0.035	0.035	0.036	0.037	0.038	0.039	0.039	0.040	0.043	0.044
22	0.016	0.017	0.017	0.016	0.016	0.016	0.016	0.016	0.017	0.017	0.016	0.016	0.016
23	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
24	0.006	0.006	0.005	0.005	0.006	0.006	0.007	0.006	0.007	0.006	0.007	0.007	0.007
25	0.014	0.014	0.014	0.014	0.015	0.015	0.015	0.016	0.016	0.018	0.018	0.018	0.019
26	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
27	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003
28	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
29	0.030	0.031	0.032	0.033	0.033	0.033	0.035	0.036	0.036	0.036	0.034	0.033	0.032
30	0.009	0.008	0.009	0.009	0.009	0.008	0.008	0.009	0.009	0.008	0.009	0.009	0.009
31	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.009	0.009	0.009	0.009	0.009
32	0.021	0.019	0.019	0.018	0.020	0.022	0.023	0.024	0.026	0.025	0.027	0.028	0.029
33	0.054	0.056	0.056	0.056	0.057	0.056	0.056	0.054	0.054	0.053	0.052	0.051	0.051
34	0.361	0.359	0.357	0.361	0.363	0.365	0.361	0.357	0.352	0.353	0.352	0.351	0.353

Note: *n* denotes the country code, given in Table 1.

Source: Author's calculations

The annual OECD volume growth measures defined by (17) as well as our earlier US dollar weighted measures γ^t are listed in Table 4. It can be seen that the Laspeyres, Paasche and Fisher measures of OECD growth explained in this section are virtually identical. Therefore, moving from the OECD Laspeyres-type measure of overall volume growth to the Fisher measure did not make much difference for this data set [29]. It can also be seen that our preferred Fisher measure of OECD growth in comparable units across countries, $\gamma_F^{\,t}$, grew on average about 1/10 of a percentage point more rapidly per year than our preferred measure of OECD GDP growth using US dollar weights, γ^t. Although this is not a large difference in growth rates, it is significant and so users need to decide which measure, $\gamma_F^{\,t}$ or γ^t, best suits their needs.

The measure γ^t can be defined using just national information on domestic price and quantity (or volume) indexes and exchange rates while the measure $\gamma_F^{\,t}$ requires information on domestic values, domestic volume indexes and PPPs. PPPs are not likely to be nearly as accurate as national measures of price and volume change due to the difficulties in matching products across countries. There are additional difficulties with the treatment of international trade in the construction of PPPs.

The γ^t measure has the problem that large fluctuations in exchange rates can lead to fluctuations in the γ^t while the PPP based $\gamma_F^{\,t}$ measures are theoretically independent from exchange rate movements [30]. Thus one has to weigh the disadvantage of possibly less reliable PPPs against the advantage of having aggregate growth measures that are independent from exchange rate movements [31].

The OECD real output shares, $s_n^{\,t}$ defined by (13), can also be used as weights for national GDP inflation rates. We define the *OECD Laspeyres, Paasche and Fisher PPP based chain link price indexes*, $\Pi_L^{\,t}$, $\Pi_P^{\,t}$ and $\Pi_F^{\,t}$ for $t = 2001,\ldots,2012$, as follows [32]:

(18) $\Pi_L^{\,t} \equiv \sum_{n=1}^{34} s_n^{\,t-1}(P_n^{\,t}/P_n^{\,t-1})\,;\ \Pi_P^{\,t} \equiv [\sum_{n=1}^{34} s_n^{\,t}(P_n^{\,t}/P_n^{\,t-1})^{-1}]^{-1}\,;\ \Pi_F^{\,t} \equiv [\Pi_L^{\,t}\,\Pi_P^{\,t}]^{1/2}.$

These chain link indexes can be multiplied together to generate the corresponding OECD aggregate price levels. The resulting Fisher OECD price level index for year t is denoted by $P_{PPP}^{\,t}$ and it is listed in the last column of Table 2 [33]. The inflation growth factors can also be transformed into *growth rates*, $\rho_L^{\,t}$, $\rho_P^{\,t}$ and $\rho_F^{\,t}$ in percentage points, by using the following definitions for $t = 2001,\ldots,2012$:

(19) $\rho_L^{\,t} \equiv 100[\Pi_L^{\,t} - 1]\,;\ \rho_P^{\,t} \equiv 100[\Pi_P^{\,t} - 1]\,;\ \rho_F^{\,t} \equiv 100[\Pi_F^{\,t} - 1]\,.$

These PPP based inflation rates (in percentage points) are listed in the last 3 columns of Table 4. The sample averages of the $\rho_L^{\,t}$, $\rho_P^{\,t}$ and $\rho_F^{\,t}$ are 2.17, 2.10 and 2.14 percentage points. Viewing Table 4, it can be seen that there are some significant differences between the three measures of OECD inflation that are PPP based [34]. Comparing the inflation numbers in Tables 2 and 4,

(29) Recall that the official OECD measure of real GDP growth is the Laspeyres measure, $\gamma_L^{\,t}$. Our estimates differ slightly from the official measures due to rounding. The exchange-rate-weighted growth rates γ^t should be somewhat lower than the PPP-weighted growth rates $\gamma_F^{\,t}$ due to the Balassa-Samuelson effect and this expectation is realized for the OECD data. We would expect the divergence to grow as less rich countries are added to the list of countries.

(30) Exchange rate movements do not directly affect the domestic rates of growth (the $Q_n^{\,t}/Q_n^{\,t-1}$) but as we have seen, they do affect the weights used to aggregate the country real growth rates into the overall OECD Laspeyres, Paasche and Fisher growth rates. Exchange rate fluctuations are large enough to materially affect the weights, which in turn lead to material fluctuations in the overall OECD volume growth rates.

(31) It will generally be the case that the $s_n^{\,t}$ will be greater than the corresponding $S_n^{\,t}$ for countries n that are relatively poor and thus the index of OECD aggregate real GDP growth defined in Section 2 will tend to be a more plutocratic index (since rich countries get larger share weights in this index) compared to the more democratic index of OECD aggregate real GDP growth defined in Section 3. Thus one could choose between the two indexes based on one's preferences over weights. We owe this point to Marshall Reinsdorf.

(32) The official OECD measure of household inflation over member countries is the Laspeyres measure defined in (18) where household consumption replaces GDP; see the OECD (2014).

(33) This price index satisfies the time reversal test whereas its Laspeyres and Paasche counterparts do not satisfy this important test. Hence the Fisher PPP based inflation index $P_{PPP}^{\,t}$ is our preferred measure of OECD aggregate inflation.

(34) In view of these differences in the three indexes of OECD GDP inflation, it may be preferable for the OECD to replace their Laspeyres type indexes of OECD household inflation by their Fisher counterparts.

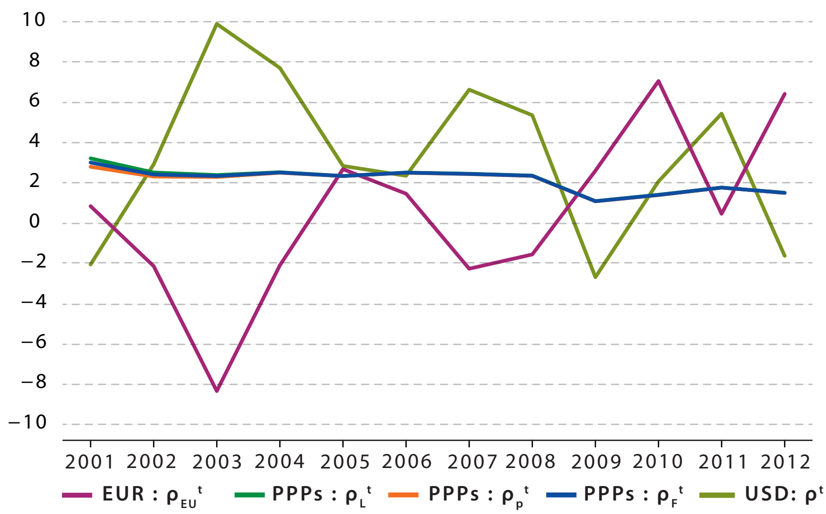
as plotted in Figure 2, it can be seen that the PPP based estimates of OECD inflation are much more reasonable than the estimates that were based on country exchange rates that were listed in Table 2, the ρ^t and $\rho_{EU}{}^t$. Our conclusion is that the OECD Fisher price index $P_{PPP}{}^t$ is a much better measure of OECD inflation than the indexes that used exchange rates instead of PPPs.

Table 4: Annual percentage point changes in OECD PPP based Laspeyres, Paasche and Fisher volume measures $\gamma_L{}^t$, $\gamma_P{}^t$ and $\gamma_F{}^t$, US dollar weighted volume measures γ^t and Laspeyres, Paasche and Fisher PPP based inflation measures, $\rho_L{}^t$, $\rho_P{}^t$ and $\rho_F{}^t$: 2001-2012

Year *t*	$\gamma_L{}^t$	$\gamma_P{}^t$	$\gamma_F{}^t$	γ^t	$\rho_L{}^t$	$\rho_P{}^t$	$\rho_F{}^t$
2001	1.291	1.296	1.294	1.231	3.216	2.798	3.007
2002	1.683	1.677	1.680	1.517	2.514	2.312	2.413
2003	2.167	2.161	2.164	2.094	2.386	2.290	2.338
2004	3.327	3.333	3.330	3.175	2.526	2.501	2.513
2005	2.832	2.831	2.831	2.680	2.341	2.336	2.338
2006	3.153	3.159	3.156	3.019	2.512	2.500	2.506
2007	2.707	2.707	2.707	2.642	2.452	2.435	2.443
2008	0.191	0.190	0.191	0.114	2.359	2.343	2.351
2009	– 3.571	– 3.574	– 3.573	– 3.639	1.092	1.081	1.086
2010	2.995	3.001	2.998	2.879	1.408	1.389	1.399
2011	1.956	1.963	1.960	1.766	1.763	1.762	1.762
2012	1.543	1.530	1.537	1.496	1.507	1.498	1.502
Average	**1.689**	**1.690**	**1.690**	**1.581**	**2.173**	**2.104**	**2.138**

Source: Author's calculations

Figure 2: Alternative OECD inflation rates using US dollar, euro and PPPs
(Percentage point changes)



Source: Author's calculations

Now we come to the most difficult problem: how can we use PPP information and national growth rates to obtain estimates of member country GDP volumes that are comparable across time and space? The Eurostat (2012) *Manual* offers the following advice:

> *'To trace the evolution of relative GDP volume levels between countries over time, it is necessary to select one of the reference years as a base year and to extrapolate its relative GDP volume levels over the other years. Extrapolation is done by applying the relative rates of GDP volume growth observed in the different countries. This provides a time series of volume indices at a constant uniform price level that replicates exactly the relative movements of GDP volume growth of each country.' Eurostat (2012, p. 18).*

We implement this strategy in Sections 4 and 5 below, where we choose the relative country GDP volumes given by the country shares of OECD aggregate GDP for 2000 (Section 4) and for 2012 (Section 5) and we use national growth rates for country GDP volumes to extrapolate these base shares to all time periods. However, it will be seen that the resulting comparable country volumes over time and space differ considerably, depending upon which base year is chosen. This is rather inconvenient: studies of competitiveness of OECD countries and living standards convergence across countries will want to use country volume series that are not subject to violent revision [35].

Our suggested solution to the problem of harmonizing national growth rates of GDP with the country shares of OECD aggregate real GDP rests on two principles. First, the resulting harmonized estimates of country volumes must be consistent with the real annual cross country volume shares $s_n^{\,t}$ listed in Table 3 above. Second, OECD aggregate real GDP growth must be equal to the rates of aggregate growth generated by our recommended Fisher indexes $\Gamma_F^{\,t}$ defined by (16).

Using these principles, the country GDP volumes are uniquely determined (up to a scalar units-of-measurement factor). To see this, first define the OECD volume index that chains the $\Gamma_F^{\,t}$ defined by (16) into a time series index, $Q_H^{\,t}$. Define $Q_H^{\,t}$ as follows:

(20) $Q_H^{\,2000} \equiv 1\,;\; Q_H^{\,t} \equiv Q_H^{\,t-1}\,\Gamma_F^{\,t}\,;\; t = 2001,\ldots,2012.$

Now use the country shares of OECD real GDP $s_n^{\,t}$ listed in Table 3 and the aggregate index $Q_H^{\,t}$ to define the following preliminary *harmonized country volumes* for country *n* in year *t*, $q_{Hn}^{\,t}$, as follows:

(21) $q_{Hn}^{\,t} \equiv Q_H^{\,t}\,s_n^{\,t}\,;\; n = 1,\ldots,34;\; t = 2000,\ldots,2012.$

Note that for each year t, $\sum_{n=1}^{34} q_{Hn}^{\,t} = \sum_{n=1}^{34} Q_H^{\,t}\,s_n^{\,t} = Q_H^{\,t}\,(\sum_{n=1}^{34} s_n^{\,t}) = Q_H^{\,t}$ and so the harmonized volumes satisfy the two principles listed above. In principle, the country volumes defined by (21) are independent of country prices and exchange rates [36].

(35) McCarthy (2013, pp. 484–486) explains in some detail why estimates of real GDP based on national growth information do not match up exactly with relative GDP estimates based on PPP benchmark information. The PPP information is generally not as accurate as national price index information due to the difficulty of matching representative products across countries. However, country methodology for constructing national price indexes (such as the Consumer Price Index) differs considerably across countries; e.g., some countries may use out of date reference expenditure baskets, some countries use Carli indexes at the elementary level while others use the Jevons or Dutot indexes which generate lower estimates of inflation at the elementary level and some countries may use quality adjustment methods more extensively than others. All of these methodological differences lead to inconsistencies between the time series and cross sectional estimates. Finally, the index number formulae used at the national levels and in the construction of the benchmark PPPs are in general not transitive and so it is impossible to achieve perfect consistency.

(36) However, in practice, the PPPs do not do a perfect job in eliminating exchange rate effects (since adjusted exchange rates are used in place of true PPPs to deflate international trade flows). If the relative PPPs are independent of the choice of the numeraire country, then the relative volumes defined by (20) will also be independent of this choice.

It is of interest to define US dollar prices for real GDP for each country. Recall that the value of country n's nominal GDP converted into US dollar at market exchange rates for year *t* was defined as v_n^t. The corresponding harmonized US dollar price of a unit of (comparable across countries) real GDP for country *n* in year *t* is defined as follows:

(22) $p_{Hn}^t \equiv v_n^t / q_{Hn}^t$; $n = 1,\dots,34$; $t = 2000,\dots,2012$.

Table 5: Harmonized OECD country GDP volumes in comparable US dollar units of measurement q_{Hn}^t

n	q_{Hn}^{2000}	q_{Hn}^{2001}	q_{Hn}^{2002}	q_{Hn}^{2003}	q_{Hn}^{2004}	q_{Hn}^{2005}	q_{Hn}^{2006}	q_{Hn}^{2007}	q_{Hn}^{2008}	q_{Hn}^{2009}	q_{Hn}^{2010}	q_{Hn}^{2011}	q_{Hn}^{2012}
1	537.3	553.2	572.5	598.6	618.7	637.2	658.2	686.6	685.8	718.5	740.3	764.3	786.4
2	231.6	227.6	235.2	239.3	245.8	246.0	258.4	262.7	268.2	259.9	266.6	273.8	279.1
3	283.4	285.9	296.6	295.8	297.6	299.9	308.5	314.9	320.7	315.3	325.3	330.1	334.5
4	874.1	886.8	895.8	930.9	960.5	1 006.5	1 027.2	1 050.1	1 050.5	1 018.4	1 053.4	1 080.3	1 094.9
5	147.6	152.5	155.5	160.9	172.3	183.5	219.1	232.4	223.7	219.7	252.1	281.6	295.0
6	159.7	167.8	171.3	180.2	187.7	193.5	204.1	218.4	218.2	214.8	210.3	213.1	212.0
7	153.9	153.7	157.9	154.4	159.8	159.9	167.4	171.2	177.0	168.8	177.8	176.6	178.2
8	13.5	14.2	15.6	17.1	18.3	19.8	22.0	24.1	23.9	21.2	21.3	23.2	23.9
9	132.9	134.2	136.8	135.5	143.0	143.2	149.1	159.0	163.6	151.6	152.6	155.8	156.5
10	1 533.0	1 587.0	1 628.6	1 593.3	1 613.0	1 654.4	1 701.9	1 756.9	1 772.1	1 740.1	1 760.9	1 790.8	1 791.5
11	2 117.5	2 144.2	2 162.4	2 202.5	2 242.8	2 281.5	2 360.8	2 431.0	2 464.8	2 330.8	2 431.2	2 516.6	2 551.1
12	199.2	210.8	224.6	233.5	241.7	240.4	254.9	258.2	269.0	263.2	246.0	226.3	215.8
13	121.3	133.0	142.3	146.3	149.8	152.2	157.4	158.2	165.8	162.1	162.4	164.2	164.9
14	8.1	8.5	8.5	8.4	9.0	9.2	9.3	9.6	10.2	9.5	8.9	9.0	9.1
15	109.8	115.4	124.4	130.5	136.6	143.4	154.2	164.5	153.2	144.5	146.5	150.6	151.4
16	147.0	146.3	147.5	139.9	146.8	143.5	144.2	152.8	151.2	155.0	163.1	173.8	178.1
17	1 466.5	1 515.3	1 471.0	1 478.8	1 466.6	1 473.6	1 530.6	1 581.8	1 614.8	1 549.1	1 533.1	1 537.5	1 524.6
18	3 287.0	3 288.1	3 316.3	3 357.2	3 438.6	3 458.3	3 471.1	3 548.1	3 468.9	3 234.7	3 389.8	3 350.2	3 390.1
19	808.4	837.8	894.1	908.8	951.4	975.1	1 003.2	1 054.8	1 056.5	1 048.2	1 116.9	1 148.5	1 163.3
20	23.4	23.2	24.5	25.8	27.2	28.2	31.7	33.7	33.3	31.2	33.4	34.9	35.4
21	987.1	983.6	1 000.8	1 043.2	1 086.4	1 150.3	1 232.1	1 273.9	1 315.8	1 282.3	1 353.6	1 475.9	1 519.7
22	468.3	481.5	492.7	484.2	494.8	509.4	531.7	554.5	570.8	541.2	545.2	550.5	545.9
23	82.1	84.2	87.1	90.0	92.9	93.6	97.7	101.6	100.7	104.2	105.4	108	110.4
24	162.3	163.2	160.6	164.4	178.6	195.8	214.4	218.5	236.5	211.3	221.9	234.2	248.8
25	404.3	408.3	422.3	430.9	454.9	467.7	491.1	530.9	555.6	572.8	604.4	630.8	645.2
26	182.0	185.7	189.6	191.2	190.9	200.4	207.7	213.4	214.2	211.4	214.7	209.3	201.9
27	59.3	63.3	66.6	68.9	72.2	77.5	84.7	93.6	101.5	97.6	99.5	100.6	102.9
28	34.9	35.8	37.7	38.6	40.7	41.8	43.6	45.7	47.5	43.7	43.0	43.5	42.7
29	858.1	896.4	949.8	978.8	1 014.8	1 056.9	1 144.6	1 202.1	1 221.6	1 172.3	1 145.6	1 129.2	1 118.6
30	248.0	244.8	249.6	256.5	267.6	262.5	277.0	292.5	295.4	276.7	290.8	301.7	309.2
31	233.6	234.8	241	238.7	243.0	244.4	261.7	280.9	296.5	290.0	300.3	313.1	321.8
32	589.3	546.9	546.5	553.4	630.4	694.6	764.7	811.3	863.7	829.6	909.8	975.7	1024.8
33	1 552.1	1 604.6	1 646.0	1 686.4	1 756.3	1 784.4	1 839.8	1 839.7	1 816.7	1 716.6	1 758.8	1 753.8	1 788.8
34	10 290	10 356	10 489	10 833	11 243	11 643	11 854	12 040	11 904	11 517	11 817	12 032	12 270

Source: Author's calculations

In order to make the harmonized volumes and prices defined by (21) and (22) comparable to the country prices and volumes expressed in US dollar that are listed in the Annex in Tables A3 and A4, we impose a normalization on the prices defined by (22) that makes the price level for the US in 2000 equal to unity; i.e., we divide all prices defined by (22) by a constant that sets the resulting p_{H34}^{2000} equal to 1 and the quantities or volumes defined by (21) are all multiplied by this constant. The resulting normalized q_{Hn}^{t} and p_{Hn}^{t} are listed in Tables 5 and 6.

Note that $q_{H34}^{2000} = q_{34}^{2000}$ and $p_{H34}^{2000} = p_{34}^{2000} = 1$ so that country GDP volumes are measured as multiples of a bundle of US GDP in the year 2000. Thus the price levels in Table 6 measure the US dollar value of constant bundle of GDP that is (in theory) comparable across countries. The price levels in Table 6 are comparable across space and time, whereas the price levels pnt listed in Table A3 of the Annex are only comparable across time for each country.

Table 6: Harmonized OECD country GDP price levels in comparable US dollar units of measurement p_{Hn}^{t}

n	p_{Hn}^{2000}	p_{Hn}^{2001}	p_{Hn}^{2002}	p_{Hn}^{2003}	p_{Hn}^{2004}	p_{Hn}^{2005}	p_{Hn}^{2006}	p_{Hn}^{2007}	p_{Hn}^{2008}	p_{Hn}^{2009}	p_{Hn}^{2010}	p_{Hn}^{2011}	p_{Hn}^{2012}
1	0.763	0.706	0.760	0.931	1.095	1.193	1.239	1.432	1.534	1.403	1.740	1.988	1.998
2	0.829	0.842	0.883	1.061	1.186	1.240	1.258	1.428	1.544	1.476	1.417	1.519	1.414
3	0.821	0.813	0.853	1.054	1.216	1.258	1.297	1.460	1.582	1.501	1.448	1.555	1.444
4	0.829	0.807	0.820	0.930	1.033	1.127	1.245	1.356	1.431	1.313	1.497	1.609	1.626
5	0.528	0.468	0.451	0.473	0.576	0.671	0.706	0.745	0.803	0.783	0.863	0.892	0.910
6	0.368	0.384	0.458	0.529	0.607	0.672	0.727	0.826	1.033	0.918	0.944	1.014	0.927
7	1.040	1.044	1.101	1.377	1.532	1.611	1.639	1.819	1.943	1.839	1.760	1.889	1.768
8	0.420	0.438	0.471	0.577	0.659	0.702	0.764	0.913	0.993	0.918	0.892	0.971	0.936
9	0.917	0.929	0.989	1.212	1.322	1.367	1.395	1.548	1.662	1.579	1.551	1.684	1.581
10	0.865	0.843	0.892	1.125	1.274	1.292	1.325	1.470	1.598	1.506	1.457	1.554	1.458
11	0.891	0.877	0.928	1.100	1.216	1.213	1.230	1.367	1.470	1.415	1.359	1.442	1.343
12	0.632	0.616	0.650	0.826	0.943	0.999	1.027	1.183	1.270	1.220	1.196	1.281	1.153
13	0.382	0.396	0.466	0.571	0.680	0.725	0.715	0.860	0.930	0.781	0.785	0.837	0.756
14	1.072	0.937	1.043	1.309	1.466	1.769	1.788	2.124	1.651	1.273	1.409	1.556	1.492
15	0.886	0.912	0.989	1.216	1.364	1.413	1.445	1.578	1.724	1.560	1.430	1.501	1.391
16	0.844	0.836	0.765	0.847	0.861	0.932	1.006	1.089	1.333	1.257	1.335	1.402	1.353
17	0.753	0.742	0.833	1.024	1.183	1.212	1.224	1.345	1.429	1.363	1.341	1.429	1.321
18	1.439	1.265	1.200	1.282	1.354	1.322	1.255	1.228	1.398	1.557	1.621	1.760	1.758
19	0.660	0.602	0.644	0.708	0.759	0.866	0.949	0.995	0.882	0.796	0.909	0.970	0.971
20	0.866	0.871	0.921	1.130	1.251	1.333	1.343	1.522	1.642	1.586	1.557	1.662	1.556
21	0.645	0.693	0.711	0.671	0.698	0.736	0.770	0.811	0.830	0.686	0.762	0.783	0.771
22	0.822	0.832	0.889	1.112	1.233	1.253	1.275	1.411	1.526	1.471	1.426	1.513	1.411
23	0.655	0.635	0.711	0.924	1.093	1.216	1.127	1.331	1.296	1.138	1.361	1.511	1.551
24	1.037	1.047	1.195	1.368	1.456	1.553	1.586	1.801	1.919	1.793	1.897	2.095	2.008
25	0.424	0.466	0.469	0.503	0.556	0.650	0.696	0.801	0.953	0.752	0.777	0.818	0.759
26	0.645	0.648	0.698	0.847	0.971	0.957	0.972	1.086	1.176	1.107	1.066	1.137	1.051
27	0.344	0.334	0.367	0.483	0.584	0.618	0.659	0.801	0.929	0.894	0.877	0.953	0.888
28	0.572	0.572	0.614	0.756	0.831	0.855	0.892	1.036	1.149	1.127	1.093	1.156	1.063
29	0.676	0.679	0.723	0.903	1.029	1.070	1.080	1.199	1.304	1.241	1.209	1.288	1.182
30	0.997	0.929	1.005	1.227	1.353	1.412	1.441	1.581	1.646	1.466	1.592	1.776	1.694
31	1.096	1.118	1.190	1.402	1.540	1.574	1.548	1.604	1.768	1.757	1.829	2.105	1.961
32	0.452	0.358	0.426	0.548	0.622	0.695	0.694	0.798	0.846	0.741	0.804	0.794	0.769
33	0.962	0.926	0.985	1.112	1.265	1.301	1.350	1.553	1.480	1.286	1.305	1.404	1.382
34	1.000	1.026	1.047	1.063	1.092	1.125	1.169	1.203	1.237	1.252	1.266	1.291	1.324

Source: Author's calculations

From Table 6, it can be seen that the countries with the lowest price levels (in US dollar) in 2012 are countries 13, 21, 25 and 32 (Hungary, Mexico, Poland and Turkey) with price levels in the 0.76 to 0.77 range. Countries with relatively high price levels in 2012 are countries 1 (Australia, $p_{H1}^{2012} = 2.00$), 4 (Canada, $p_{H4}^{2012} = 1.62$), 7 (Denmark, 1.77), 9 (Finland, 1.58), 18 (Japan, 1.76), 20 (Luxembourg, 1.56), 23 (New Zealand, 1.55), 24 (Norway, 2.01), 30 (Sweden, 1.69) and 31 (Switzerland, 1.96). These price level estimates are [37](imperfect) indicators of the competiveness of the country on international markets, with lower price levels indicating greater competiveness.

A problem with the volume estimates listed in Table 5 is that they do not 'respect' national growth rates of GDP by country, in that they are not required to agree with the national growth rates that are produced by the individual countries; only the aggregate OECD growth rate is respected. In the following two sections, we will derive alternative country volume estimates that are comparable over time and space. These alternative estimates will respect country growth rates but they will not reproduce the real OECD country expenditure shares listed in Table 3 for all time periods.

4. OECD growth and inflation using country annual GDP volume growth rates and base period shares of OECD real GDP

We generate comparable country GDP volume estimates for OECD countries covering the period 2000-2012 by using the real GDP country volume shares for 2000, the s_n^{2000} listed in Table 3 above, along with the national growth rates of country real GDP relative to 2000, the Q_n^t/Q_n^{2000} listed in Table A2 of the Annex. This is a typical strategy in forming estimates of real GDP that rely on PPPs that are only produced infrequently. Our purpose in listing these estimates is to evaluate how different the resulting estimates are from our preferred harmonized volume estimates, q_{Hn}^t, listed in Table 5 above.

Define preliminary *base period estimates of country GDP volumes for year t and country n*, q_{Bn}^t, as follows:

$$(23)\ q_{Bn}^t \equiv s_n^{2000} (Q_n^t/Q_n^{2000})\ ;\ n = 1,\ldots,34\ ;\ t = 2000,\ldots,2012.$$

The above estimates are obviously based on the country shares of real OECD GDP that prevailed in 2000 (the s_n^{2000}) and the long term country growth rates of real GDP (the Q_n^t/Q_n^{2000}). The companion *country US dollar price levels* for country n and year t, p_{Bn}^t, are defined as follows:

$$(24)\ p_{Bn}^t \equiv v_n^t/q_{Bn}^t\ ;\ n = 1,\ldots,34\ ;\ t = 2000,\ldots,2012$$

where v_n^t is the nominal value of GDP for country n in year t converted into US dollar at market exchange rates for that year.

In order to make the volumes and prices defined by (23) and (24) comparable to the country prices and volumes expressed in US dollar that are listed Tables 5 and 6 in the previous section,

(37) The price levels p_{Hn}^t are imperfect indicators of competiveness because not all components of GDP are internationally traded. Moreover, these price levels are not independent of the choice of the numeraire currency (US dollar in this case). They are also imperfect because they depend heavily on the accuracy of the underlying PPPs and these PPPs are subject to considerable error variances due to the difficulties involved in matching product prices (and quantities) across countries.

we impose a normalization on the prices defined by (24) that makes the price level for the US in 2000 equal to unity; i.e., we divide all prices defined by (24) by a constant that sets the resulting p_{B34}^{2000} equal to 1 and the volumes defined by (23) are all multiplied by this constant. The resulting normalized p_{Bn}^{t} are listed in Table 7 ([38]).

Table 7: OECD country GDP price levels in comparable US dollar units of measurement p_{Bn}^{t} based on country growth rates of GDP volumes and year 2000 country shares of OECD output

n	p_{Bn}^{2000}	p_{Bn}^{2001}	p_{Bn}^{2002}	p_{Bn}^{2003}	p_{Bn}^{2004}	p_{Bn}^{2005}	p_{Bn}^{2006}	p_{Bn}^{2007}	p_{Bn}^{2008}	p_{Bn}^{2009}	p_{Bn}^{2010}	p_{Bn}^{2011}	p_{Bn}^{2012}
1	0.763	0.699	0.756	0.930	1.094	1.191	1.233	1.432	1.507	1.414	1.764	2.014	2.008
2	0.829	0.821	0.874	1.060	1.186	1.212	1.246	1.386	1.509	1.454	1.406	1.506	1.416
3	0.821	0.814	0.873	1.068	1.200	1.230	1.270	1.418	1.550	1.488	1.448	1.550	1.460
4	0.829	0.804	0.802	0.928	1.031	1.144	1.255	1.367	1.433	1.312	1.499	1.611	1.622
5	0.528	0.468	0.448	0.469	0.571	0.667	0.793	0.844	0.848	0.820	0.981	1.070	1.083
6	0.368	0.391	0.466	0.546	0.624	0.667	0.711	0.817	0.990	0.907	0.891	0.953	0.875
7	1.040	1.036	1.117	1.361	1.531	1.573	1.620	1.810	2.015	1.929	1.917	2.023	1.917
8	0.420	0.434	0.478	0.596	0.685	0.727	0.799	0.972	1.095	1.042	0.997	1.077	1.028
9	0.917	0.917	0.977	1.163	1.286	1.294	1.316	1.479	1.630	1.568	1.500	1.618	1.539
10	0.865	0.857	0.922	1.127	1.261	1.287	1.326	1.484	1.629	1.556	1.498	1.592	1.494
11	0.891	0.875	0.933	1.132	1.258	1.268	1.283	1.423	1.535	1.473	1.418	1.507	1.413
12	0.632	0.626	0.680	0.848	0.960	0.989	1.022	1.152	1.291	1.252	1.207	1.281	1.175
13	0.382	0.419	0.505	0.612	0.712	0.742	0.728	0.879	0.988	0.870	0.867	0.920	0.848
14	1.072	0.940	1.055	1.269	1.422	1.629	1.591	1.841	1.500	1.155	1.249	1.360	1.297
15	0.886	0.912	1.012	1.259	1.418	1.454	1.515	1.682	1.748	1.595	1.497	1.582	1.472
16	0.844	0.833	0.770	0.796	0.810	0.816	0.838	0.907	1.055	1.008	1.074	1.149	1.101
17	0.753	0.752	0.817	1.010	1.138	1.160	1.190	1.329	1.459	1.412	1.352	1.438	1.352
18	1.439	1.261	1.203	1.279	1.352	1.311	1.228	1.202	1.352	1.486	1.549	1.672	1.658
19	0.660	0.600	0.640	0.695	0.745	0.839	0.899	0.943	0.818	0.730	0.836	0.885	0.879
20	0.866	0.842	0.904	1.148	1.286	1.349	1.453	1.645	1.768	1.690	1.726	1.888	1.798
21	0.645	0.691	0.715	0.695	0.723	0.781	0.834	0.878	0.917	0.786	0.874	0.943	0.921
22	0.822	0.839	0.917	1.123	1.245	1.277	1.311	1.457	1.592	1.512	1.453	1.542	1.444
23	0.655	0.628	0.693	0.893	1.053	1.141	1.087	1.289	1.265	1.133	1.369	1.524	1.550
24	1.037	1.033	1.143	1.326	1.475	1.681	1.837	2.071	2.388	2.026	2.240	2.580	2.549
25	0.424	0.465	0.477	0.503	0.557	0.646	0.683	0.796	0.943	0.756	0.793	0.833	0.776
26	0.645	0.649	0.708	0.874	0.985	1.012	1.049	1.177	1.280	1.225	1.175	1.236	1.139
27	0.344	0.344	0.381	0.495	0.597	0.636	0.683	0.831	0.988	0.962	0.922	0.983	0.920
28	0.572	0.570	0.620	0.759	0.844	0.856	0.882	1.003	1.118	1.095	1.033	1.097	1.016
29	0.676	0.684	0.751	0.938	1.074	1.122	1.179	1.329	1.456	1.382	1.318	1.384	1.279
30	0.997	0.905	0.975	1.195	1.319	1.309	1.351	1.516	1.603	1.409	1.509	1.697	1.642
31	1.096	1.111	1.210	1.412	1.542	1.544	1.567	1.678	1.911	1.894	1.983	2.337	2.216
32	0.452	0.353	0.394	0.488	0.577	0.656	0.675	0.786	0.881	0.779	0.849	0.827	0.823
33	0.962	0.936	0.999	1.112	1.276	1.292	1.345	1.497	1.419	1.229	1.257	1.334	1.337
34	1.000	1.023	1.039	1.059	1.088	1.123	1.158	1.189	1.212	1.221	1.236	1.260	1.282

Source: Author's calculations

The differences between the entries in Tables 6 and 7 are very large. If we take each column in Table 6, subtract the corresponding entries in the same column of Table 7 and then take the

([38]) The entries in Tables 5 and 6 enable one to recover the US dollar values of GDP, equal to $v_n^t = p_{Hn}^t q_{Hn}^t$ for $n = 1,\ldots,34$ and $t = 2000,\ldots,2012$. Then the q_{Bn}^t can be recovered as v_n^t/p_{Bn}^t.

absolute value of the differences, we find that the *average* absolute difference grows from 0 in 2000 to 9.4 percentage points in 2012 ([39]). The maximum absolute difference grows from 0 in 2000 to 54.0 percentage points in 2012. These are massive differences in price levels, which translate into massive differences in GDP levels. This problem of the inconsistency with national growth rates is well known but most users of PPP adjusted country real volume estimates are not aware of the magnitude of these inconsistencies ([40]).

In the following section, we undertake a computation that is similar to those of the present section except that we use the real volume shares of 2012 as the benchmark shares instead of the shares of 2000.

5. OECD growth and inflation using country annual GDP volume growth rates and current period shares of OECD real GDP

We generate comparable country GDP volume estimates for OECD countries covering the period 2000-2012 by using the real GDP country volume shares for 2012, the s_n^{2012} listed in Table 3 above, along with the national growth rates of country real GDP relative to 2000, the Q_n^t/Q_n^{2000} listed in Table A2 of the Annex. This method for forming comparable country GDP volumes is used by the World Bank when the International Comparisons Project produces a new set of PPPs ([41]). The methodology is straightforward and follows the approach used in the previous section except that the 2012 country volume shares are used in place of the 2000 shares.

The preliminary *end of sample period estimates of country GDP volumes for year t and country n*, q_{En}^t, is defined as follows:

$$(25)\ q_{En}^t \equiv s_n^{2012}\,(Q_n^t/Q_n^{2012})\ ;\ n = 1,\ldots,34\ ;\ t = 2000,\ldots,2012.$$

The above estimates are obviously based on the country shares of real OECD GDP that prevailed in 2012 (the s_n^{2012}) and the levels of real GDP in year *t* relative to the corresponding country *n* level in 2012 (the Q_n^t/Q_n^{2012}). The companion country US dollar price levels for country *n* and year *t*, p_{En}^t, are defined as follows:

$$(26)\ p_{En}^t \equiv v_n^t/q_{En}^t\ ;\ n = 1,\ldots,34\ ;\ t = 2000,\ldots,2012$$

where v_n^t is the nominal value of GDP for country *n* in year *t* converted into US dollar at market exchange rates for that year.

In order to make the volumes and prices defined by (25) and (26) comparable to the harmonized country prices and volumes expressed in US dollar that are listed Tables 5 and 6 in Section 3,

([39]) The sequence of *average* absolute differences in percentage points over the 13 years is as follows: 0, 0.8, 1.6, 2.0, 2.2, 3.8, 5.2, 6.4, 7.6, 6.8, 8.0, 9.3, 9.4. The sequence of maximum absolute differences in percentage points over the 13 years is as follows: 0, 2.9, 5.2, 6.0, 5.4, 14.0, 25.1, 28.2, 46.9, 24.8, 34.3, 48.5, 54.0.

([40]) See Eurostat (2012, p. 18) on this point.

([41]) See Chapter 18 in the World Bank (2013). The Penn World Tables use the extrapolation methodology described in this section and the previous section to construct estimates of comparable real GDP for periods subsequent to the last available ICP round and prior to the first available ICP round; see Summers and Heston (1991) and Feenstra, Inklaar and Timmer (2013).

we impose a normalization on the prices defined by (26) that makes the price level for the US in 2000 equal to unity; i.e., we divide all prices defined by (26) by a constant that sets the resulting p_{E34}^{2000} equal to 1 and the quantities or volumes defined by (25) are all multiplied by this constant. The resulting normalized p_{En}^{t} are listed in Table 8 ([42]).

Table 8: OECD country GDP price levels in comparable US dollar units of measurement p_{En}^{t} based on country growth rates of GDP volumes and year 2012 country shares of OECD output

n	p_{En}^{2000}	p_{En}^{2001}	p_{En}^{2002}	p_{En}^{2003}	p_{En}^{2004}	p_{En}^{2005}	p_{En}^{2006}	p_{En}^{2007}	p_{En}^{2008}	p_{En}^{2009}	p_{En}^{2010}	p_{En}^{2011}	p_{En}^{2012}
1	0.735	0.674	0.728	0.896	1.055	1.148	1.188	1.380	1.452	1.363	1.700	1.940	1.935
2	0.802	0.794	0.845	1.025	1.147	1.172	1.205	1.340	1.459	1.406	1.360	1.456	1.369
3	0.786	0.779	0.836	1.022	1.149	1.178	1.216	1.358	1.485	1.425	1.386	1.484	1.398
4	0.805	0.781	0.779	0.901	1.001	1.111	1.218	1.327	1.391	1.274	1.455	1.564	1.574
5	0.430	0.381	0.365	0.381	0.465	0.542	0.645	0.686	0.690	0.667	0.798	0.871	0.881
6	0.378	0.401	0.478	0.560	0.639	0.684	0.729	0.838	1.015	0.930	0.914	0.977	0.897
7	0.929	0.925	0.998	1.215	1.367	1.405	1.447	1.617	1.800	1.723	1.713	1.807	1.713
8	0.370	0.383	0.421	0.525	0.604	0.641	0.704	0.857	0.965	0.919	0.879	0.949	0.907
9	0.912	0.912	0.972	1.157	1.279	1.287	1.309	1.471	1.621	1.560	1.492	1.610	1.531
10	0.817	0.810	0.871	1.065	1.191	1.216	1.253	1.402	1.539	1.470	1.415	1.504	1.412
11	0.820	0.805	0.859	1.041	1.158	1.167	1.181	1.309	1.412	1.355	1.305	1.387	1.301
12	0.601	0.595	0.647	0.806	0.913	0.940	0.972	1.095	1.228	1.191	1.148	1.218	1.117
13	0.330	0.362	0.436	0.528	0.615	0.640	0.628	0.759	0.852	0.751	0.748	0.794	0.732
14	1.195	1.048	1.176	1.414	1.584	1.816	1.773	2.052	1.671	1.287	1.392	1.516	1.445
15	0.811	0.835	0.927	1.153	1.298	1.331	1.387	1.540	1.601	1.460	1.371	1.448	1.347
16	1.005	0.992	0.916	0.947	0.964	0.971	0.997	1.079	1.255	1.200	1.278	1.367	1.311
17	0.712	0.712	0.773	0.955	1.076	1.097	1.126	1.258	1.380	1.336	1.279	1.360	1.279
18	1.479	1.295	1.236	1.314	1.389	1.346	1.262	1.234	1.389	1.526	1.592	1.718	1.703
19	0.706	0.642	0.684	0.744	0.797	0.898	0.961	1.008	0.875	0.781	0.894	0.947	0.940
20	0.726	0.706	0.758	0.962	1.078	1.131	1.219	1.379	1.482	1.417	1.447	1.583	1.507
21	0.523	0.560	0.580	0.563	0.586	0.633	0.676	0.712	0.743	0.637	0.709	0.764	0.747
22	0.778	0.794	0.867	1.063	1.178	1.208	1.240	1.378	1.506	1.430	1.374	1.459	1.366
23	0.635	0.609	0.671	0.866	1.020	1.106	1.053	1.249	1.226	1.098	1.328	1.477	1.502
24	0.791	0.788	0.872	1.012	1.125	1.283	1.402	1.581	1.822	1.546	1.710	1.969	1.945
25	0.401	0.441	0.452	0.477	0.527	0.612	0.648	0.755	0.894	0.716	0.751	0.789	0.735
26	0.576	0.579	0.632	0.781	0.880	0.904	0.937	1.052	1.143	1.094	1.050	1.105	1.018
27	0.322	0.322	0.356	0.462	0.558	0.594	0.639	0.777	0.924	0.899	0.861	0.919	0.860
28	0.580	0.578	0.628	0.769	0.855	0.867	0.893	1.016	1.132	1.109	1.046	1.111	1.029
29	0.605	0.613	0.672	0.840	0.961	1.005	1.056	1.189	1.303	1.237	1.180	1.239	1.145
30	0.996	0.904	0.974	1.194	1.318	1.307	1.350	1.514	1.601	1.407	1.507	1.695	1.641
31	0.940	0.952	1.037	1.211	1.322	1.324	1.343	1.438	1.638	1.624	1.700	2.004	1.899
32	0.409	0.319	0.357	0.441	0.522	0.594	0.610	0.711	0.797	0.705	0.768	0.748	0.745
33	0.963	0.938	1.000	1.113	1.278	1.294	1.347	1.498	1.420	1.230	1.258	1.335	1.338
34	1.000	1.023	1.039	1.059	1.088	1.123	1.158	1.189	1.212	1.221	1.236	1.260	1.282

Source: Author's calculations

It can be seen that there are substantial differences between the price levels listed in Table 8 as compared to the price levels listed in Table 7 and the harmonized price levels listed in Table 6. If we take each column in Table 6, subtract the corresponding entries in the same column of

([42]) As in the previous section, the q_{En}^{t} can be recovered as v_n^{t}/p_{En}^{t}.

Table 8 and then take the absolute value of the differences, we find that the average absolute difference for 2000 over the 34 countries is 6.0 percentage points, which increases to 7.9 percentage points for 2005 and then gradually decreases to 4.2 percentage points in 2012. Over all observations, the maximum absolute deviation is 35.6 percentage points ([43]). Again these are large differences in price levels, which translate into large differences in GDP levels.

For comparing real GDP levels across time and space, the results presented in this section indicate that the strategy of using national growth rates and a single cross country comparison of real GDP levels will not lead to stable comparisons. The harmonization strategy suggested in Section 3 will lead to stable comparisons and if the accuracy of the annual sequence of PPPs is roughly constant, the resulting harmonized estimates seem to be preferable to the consistent national growth rate estimates that are based on a single cross country comparison.

6. OECD growth and inflation using adjusted country annual GDP volume growth rates and OECD shares of real GDP for two benchmark years

The OECD provides annual PPPs so that estimates of relative GDP volumes can be constructed for all member countries for each year. However, the World Bank's ICP PPPs are only available at infrequent intervals ([44]). We now consider using the benchmark GDP shares for the years 2000 and 2012 along with information on national GDP growth rates in order to interpolate between the benchmark years. We propose an interpolation method that leads to country shares of real GDP that are exactly consistent with the shares s_n^{2000} for the year 2000 and the shares s_n^{2012} for the year 2012.

We begin by using the methodology of Section 3 to construct country measures of real GDP that jump from the year 2000 to the year 2012. The *long term growth factor for country n* can be defined as Q_n^{2012}/Q_n^{2000} where Q_n^t is country n's GDP volume in year t ([45]). We use these long term growth factors along with the year 2000 country shares of OECD real GDP, s_n^{2000}, in order to define the *OECD Laspeyres type long term growth factor*, Γ_L, as the following weighted average of the national long term growth factors:

$$(27)\ \Gamma_L \equiv \sum_{n=1}^{34} s_n^{2000}(Q_n^{2012}/Q_n^{2000})\ ;\ t = 2001,\ldots,2012.$$

The counterpart to the Laspeyres type formula defined by (27) is the following Paasche type formula that uses the shares of 2012 and reciprocal long term growth rates:

$$(28)\ \Gamma_P \equiv [\sum_{n=1}^{34} s_n^{2012}(Q_n^{2012}/Q_n^{2000})^{-1}]^{-1}\ ;\ t = 2001,\ldots,2012.$$

([43]) The sequence of average absolute differences over the 34 countries in percentage points over the 13 years is as follows: 6.0, 6.1, 6.2, 7.0, 7.8, 7.9, 6.2, 6.8, 5.9, 5.4, 5.1, 5.2, 4.2. The sequence of maximum absolute differences in percentage points over the 13 years is as follows: 24.6, 25.9, 32.3, 35.6, 33.1, 27.1, 20.5, 22.0, 16.0, 24.7, 18.7, 12.6, 6.3. Recall that we normalized the price level of the US to be 1 in 2000 for the p_{Hn}^t and the p_{En}^t. If instead of using the normalizations $p_{H34}^{2000} = p_{E34}^{2000} = 1$ when constructing Tables 6 and 9, we used the normalizations $p_{H34}^{2012} = p_{E34}^{2012} = 1$, we would find that the absolute differences between the resulting p_{Hn}^t and p_{En}^t would equal 0 for all countries n for $t = 2012$. Thus the choice of normalization (and hence of the units of measurement) can affect the results.

([44]) The World Bank has produced benchmark PPPs for over 150 countries for 2005 and 2011; see World Bank (2013).

([45]) These long term country growth factors are conveniently listed in the last column of Table A2 in the Annex.

A symmetric average of the two indexes leads to the following Fisher type formula for *OECD long term volume growth* going from the year 2000 to the year 2012:

(29) $\Gamma_F \equiv [\Gamma_L \Gamma_P]^{1/2}$; $t = 2001,\ldots,2012$.

The long term indexes defined by (27)-(29) turn out to be 1.2207, 1.2209 and 1.2208 respectively, so that there is practically no difference in the three indexes for this data set (46). We take the Fisher measure as our preferred measure of OECD volume growth between 2000 and 2012, and use this to define country volumes for 2012.

Preliminary estimates of country GDP volumes in comparable units for the years 2000 and 2012, q_{In}^{2000} and q_{In}^{2012} (the index I indicates that these are interpolated estimates), are defined as follows:

(30) $q_{In}^{2000} \equiv s_n^{2000}$; $q_{In}^{2012} \equiv \Gamma_F s_n^{2012}$; $n = 1,\ldots,34$.

The volumes defined by (30) will be imposed as constraints on our interpolation scheme. Define the *implied long term growth factor* over the years 2000-2012 for country *n*, g_n, that is implied by the estimates of country levels given by equations (30):

(31) $g_n \equiv q_{In}^{2012}/q_{In}^{2000}$; $n = 1,\ldots,34$.

These growth factors are not necessarily equal to the national growth factors G_n that are implied by the national growth rates listed in Table A2 of the Annex:

(32) $G_n \equiv Q_n^{2012}/Q_n^{2000}$; $n = 1,\ldots,34$.

Thus for each country *n*, there is an 'error' factor or discrepancy, $E_n \equiv g_n/G_n$ between the implied growth rates gn defined by (31) and the national growth rates between 2000 and 2012, G_n defined by (32). We will distribute these errors in a proportional manner and use the resulting adjusted national growth rates to interpolate between the two benchmark observations. Thus define the *country n proportional annualized discrepancy factor*, α_n, as follows (47):

(33) $\alpha_n \equiv [g_n/G_n]^{1/12}$; $n = 1,\ldots,34$.

The q_{In}^t for non-benchmark years *t* can now be defined as follows (48):

(34) $q_{In}^t \equiv q_{In}^{t-1}(Q_n^t/Q_n^{t-1})\alpha_n$; $n = 1,\ldots,34$; $t = 2001,\ldots,2011$.

Once the q_{In}^t have been defined, the corresponding US dollar price levels p_{In}^t are defined in the usual way:

(35) $p_{In}^t \equiv v_n^t/q_{In}^t$; $n = 1,\ldots,34$; $t = 2001,\ldots,2011$.

In order to make the volumes and prices defined by (34) and (35) comparable to the harmonized country prices and volumes expressed in US dollar that are listed Tables 5 and 6 in Section 3, we impose a normalization on the prices defined by (35) that makes the price level for the

(46) Note that (29) defines a direct comparison of the data of 2000 with the data of 2012 whereas in Section 3 above, we used chained Fisher type indexes to go from 2000 to 2012. The chained Fisher index for 2012 relative to 2000 is equal to 1.2203, which is very close to its direct counterpart, 1.2208.

(47) The average of the g_n/G_n was 1.03. The maximum ratio was 1.27 (Norway) and the minimum ratio was 0.81 (Israel). The PPP based growth rates treat changes in the terms of trade differently than the nationally based growth rates and so fluctuations in the price of oil probably explain the Norwegian divergence. For the three largest countries, the ratio was 1.05 (Germany), 0.94 (Japan) and 0.97 (US).

(48) It can be verified that if we apply definitions (34) for $t = 2012$, we obtain the q_{In}^{2012} defined by (30).

US in 2000 equal to unity; i.e., we divide all prices defined by (35) by a constant that sets the resulting p_{I34}^{2000} equal to 1 and the quantities or volumes defined by (34) are all multiplied by this constant. The resulting normalized pInt are listed in Table 9 ([49]).

Table 9: OECD country GDP price levels in comparable US dollar units of measurement p_{In}^{t} based on adjusted country growth rates of GDP volumes and year 2000 and 2012 country shares of OECD output

n	p_{In}^{2000}	p_{In}^{2001}	p_{In}^{2002}	p_{In}^{2003}	p_{In}^{2004}	p_{In}^{2005}	p_{In}^{2006}	p_{In}^{2007}	p_{In}^{2008}	p_{In}^{2009}	p_{n}^{2010}	p_{In}^{2011}	p_{In}^{2012}
1	0.763	0.699	0.755	0.928	1.092	1.189	1.229	1.427	1.502	1.408	1.756	2.004	1.997
2	0.829	0.821	0.874	1.060	1.185	1.211	1.244	1.385	1.507	1.452	1.404	1.503	1.413
3	0.821	0.813	0.871	1.064	1.195	1.224	1.262	1.409	1.538	1.475	1.434	1.533	1.443
4	0.829	0.804	0.803	0.929	1.032	1.145	1.256	1.369	1.435	1.314	1.501	1.613	1.625
5	0.528	0.461	0.435	0.449	0.539	0.620	0.726	0.762	0.754	0.719	0.848	0.912	0.909
6	0.368	0.393	0.471	0.554	0.635	0.683	0.731	0.845	1.028	0.947	0.934	1.003	0.926
7	1.040	1.029	1.102	1.333	1.490	1.521	1.556	1.727	1.909	1.815	1.792	1.878	1.768
8	0.420	0.431	0.470	0.582	0.664	0.699	0.762	0.920	1.028	0.971	0.921	0.988	0.936
9	0.917	0.919	0.981	1.171	1.297	1.308	1.334	1.502	1.658	1.599	1.533	1.658	1.580
10	0.865	0.856	0.918	1.120	1.250	1.274	1.310	1.463	1.602	1.527	1.467	1.556	1.457
11	0.891	0.871	0.926	1.117	1.237	1.241	1.251	1.381	1.483	1.417	1.359	1.438	1.342
12	0.632	0.625	0.678	0.844	0.954	0.981	1.012	1.139	1.275	1.235	1.189	1.259	1.153
13	0.382	0.415	0.495	0.594	0.685	0.707	0.687	0.822	0.914	0.798	0.787	0.827	0.755
14	1.072	0.951	1.080	1.314	1.489	1.727	1.706	1.998	1.646	1.283	1.404	1.546	1.492
15	0.886	0.908	1.003	1.241	1.391	1.420	1.473	1.627	1.684	1.528	1.428	1.502	1.391
16	0.844	0.848	0.796	0.838	0.867	0.889	0.928	1.022	1.210	1.177	1.274	1.387	1.353
17	0.753	0.751	0.813	1.004	1.129	1.149	1.176	1.311	1.436	1.387	1.325	1.407	1.320
18	1.439	1.267	1.215	1.298	1.379	1.343	1.265	1.243	1.405	1.552	1.627	1.764	1.758
19	0.660	0.605	0.650	0.713	0.770	0.874	0.944	0.999	0.874	0.786	0.908	0.969	0.971
20	0.866	0.832	0.882	1.107	1.225	1.271	1.352	1.512	1.605	1.516	1.530	1.653	1.555
21	0.645	0.681	0.694	0.664	0.681	0.725	0.763	0.791	0.814	0.688	0.754	0.800	0.771
22	0.822	0.838	0.913	1.117	1.235	1.264	1.295	1.437	1.567	1.485	1.424	1.509	1.410
23	0.655	0.628	0.693	0.893	1.053	1.142	1.087	1.289	1.266	1.134	1.370	1.524	1.550
24	1.037	1.013	1.098	1.249	1.362	1.522	1.631	1.802	2.036	1.694	1.836	2.073	2.007
25	0.424	0.465	0.476	0.500	0.552	0.640	0.676	0.786	0.930	0.743	0.778	0.816	0.759
26	0.645	0.644	0.698	0.857	0.959	0.978	1.007	1.123	1.212	1.152	1.098	1.148	1.050
27	0.344	0.343	0.379	0.490	0.590	0.626	0.671	0.814	0.965	0.936	0.894	0.951	0.888
28	0.572	0.572	0.625	0.767	0.856	0.872	0.902	1.029	1.151	1.132	1.072	1.142	1.062
29	0.676	0.680	0.741	0.920	1.046	1.086	1.133	1.269	1.381	1.302	1.234	1.287	1.182
30	0.997	0.908	0.980	1.204	1.332	1.325	1.372	1.543	1.636	1.441	1.548	1.745	1.693
31	1.096	1.100	1.186	1.370	1.480	1.467	1.474	1.562	1.761	1.728	1.790	2.089	1.960
32	0.452	0.351	0.390	0.480	0.564	0.638	0.652	0.755	0.842	0.740	0.802	0.777	0.769
33	0.962	0.939	1.005	1.121	1.290	1.310	1.367	1.525	1.450	1.260	1.292	1.374	1.381
34	1.000	1.026	1.044	1.068	1.100	1.138	1.176	1.211	1.238	1.250	1.269	1.297	1.323

Source: Author's calculations

Again, it can be seen that there are some substantial differences between the price levels listed in Table 9 as compared to the price levels listed in Table 6 but the discrepancies are much reduced as compared to the discrepancies when only one benchmark set of PPPs is used. The overall sample average absolute discrepancy is now only 1.9 percentage points. The average

([49]) As usual, the q_{In}^{t} can be recovered as v_n^t/p_{In}^t.

absolute difference for 2000 over the 34 countries is 0, which increases to 3.2 percentage points for 2005 and 2007 and then gradually decreases to 0.3 percentage points in 2012. Over all observations, the maximum absolute deviation is 12.5 percentage points (50).

Some tentative conclusions can be drawn from the tables in this section and the previous sections. First, the interpolation method which is consistent with benchmark expenditure shares for two widely separated years seems to work quite well. If the benchmark PPPs are of equal quality, the interpolation method is much better than simply projecting the country shares from a single benchmark using national growth rates (51). Second, if it is too expensive to prepare annual PPPs for a group of countries, then the interpolation method will probably generate comparable country real GDP volumes that are close to our preferred harmonized volumes described in Section 3, provided that benchmark PPPs are calculated every three to five years (52).

Feenstra, Inklaar and Timmer (2013, pp. 17–19) (FIT) use a simple interpolation method to harmonize their new Penn World Table estimates of real GDP growth (in comparable units of measurement) using both ICP information between two benchmarks and national information on GDP growth. Their interpolation method is similar to our suggested method, in that their interpolated estimates are consistent with the relative GDP levels for the PPP benchmark years. The key to their interpolation method is the construction of interpolated PPPs between ICP benchmarks. We explain their method using our notation and adapting their analysis to the problem of constructing PPPs for the years 2001-2011, given that we have PPPs for the benchmark years 2000 and 2012. Recall that the domestic price level for country n in year t was defined as $P_n^{\,t} \equiv V_n^{\,t}/Q_n^{\,t}$ for $n = 1,\ldots,34$ and $t = 2000,\ldots,2012$ (and these price levels are listed in Annex Table A1). Recall also that $PPP_n^{\,t}$ was defined as the number of units of the national currency of country n that is required to purchase one dollar of US (real) GDP in year t (and these OECD PPPs are listed in Annex Table A5). Using our notation, their interpolated PPP for country n in year t, $PPP_{FITn}^{\,t}$, is defined as follows:

$$(36)\ PPP_{FITn}^{\,t} \equiv (1 - w^t)(PPP_n^{\,2000})(P_n^{\,t}/P_n^{\,2000}) + w^t(PPP_n^{\,2012})(P_n^{\,t}/P_n^{\,2012})\ ;\ n = 1,\ldots,34;\ t = 2000, 2001, \ldots,2012$$

where the weight function w^t is defined as follows:

$$(37)\ w^t \equiv (t - 2000)/12\ ;\ t = 2000, 2001, \ldots,2012.$$

Thus w^t grows linearly in t with $w^{2000} = 0$ and $w^{2012} = 1$. Note that $PPP_{FITn}^{\,2000} = PPP_n^{\,2000}$ and $PPP_{FITn}^{\,2012} = PPP_n^{\,2012}$ so that the interpolated PPPs coincide with the actual PPPs for the two benchmark years, 2000 and 2012. Thus the interpolated PPP for country n in year t, $PPP_{FITn}^{\,t}$, is a simple weighted average of two extrapolated PPPs for country n. The first index in the weighted average uses the PPPs for 2000, $PPP_n^{\,2000}$, and pushes these PPPs forward using the normalized

(50) The sequence of average absolute differences over the 34 countries in percentage points over the 13 years is as follows: 0, 1.0, 1.8, 2.0, 2.3, 3.2, 3.0, 3.2, 2.8, 2.6, 1.7, 1.2, 0.06. The sequence of maximum absolute differences in percentage points over the 13 years is as follows: 0, 3.9, 9.7, 11.9, 9.4, 10.7, 8.3, 12.5, 12.3, 9.9, 6.1, 3.2, 0.08. The reason why the differences are not all equal to 0 for 2012 is that the direct aggregate Fisher index going from 2000 to 2012 differs slightly from its chained counterpart defined in Section 3.

(51) 'Better' means 'more consistent' with our preferred harmonized volumes that could be calculated if annual PPPs were available.

(52) The interpolation method that we described in this section is not the only possible method that could be used to calculate comparable real GDP series over time and space when benchmark PPPs are only available infrequently. In particular, econometricians may prefer to use an interpolation method that is based on the Kalman filter; see Rao, Rambaldi and Doran (2010) (2011) for the description of such a method. For additional methods for harmonizing cross sectional and time series estimates of real GDP, see Rao, Rambaldi and Balk (2013). Summers and Heston (1991, pp. 340–341) also used an econometric method to reconcile the differences between national growth rates and ICP generated estimates of relative GDP levels. However, statistical agencies are generally reluctant to adopt methods that rely heavily on econometrics so the simple method of interpolation described here is proposed as an attractive alternative.

domestic inflation rates $P_n^{\,t}/P_n^{\,2000}$ and the second index uses the PPPs for 2012, $PPP_n^{\,2012}$, and pushes these PPPs backwards using the normalized domestic inflation rates $P_n^{\,t}/P_n^{\,2012}$.

We do not construct OECD real volumes for our sample period using the complete Feenstra, Inklaar and Timmer methodology but we experiment with their method of weighting. Recall that Tables 7 and 8 listed the OECD country US dollar prices (in comparable units across time and space), $p_{Bn}^{\ t}$ and $p_{En}^{\ t}$, where the prices $p_{Bn}^{\ t}$ ($p_{En}^{\ t}$) were based on country growth rates of GDP volumes and year 2000 (2012) country shares of OECD output. Using the weights *wt* defined by (37), define the FIT type US dollar price levels, $p_{FITn}^{\ t}$, as the following weighted averages of the 2000 US dollar prices $p_{Bn}^{\ 2000}$ and 2012 US dollar prices p^{En2012}:

(38) $p_{FITn}^{\ t} \equiv (1 - w^t)p_{Bn}^{\ 2000} + w^t p_{En}^{\ 2012}$; $n = 1,\ldots,34$; $t = 2000, 2001, \ldots,2012$.

It can be seen that $p_{FITn}^{\ 2000} = p_{Bn}^{\ 2000}$ and $p_{FITn}^{\ 2012} = p_{En}^{\ 2012}$ for $n = 1,\ldots,34$. Now compare the prices $p_{FITn}^{\ t}$ to our preferred Harmonized US dollar price levels $p_{Hn}^{\ t}$ listed in Table 6. Take the absolute value of the differences, $p_{Hn}^{\ t} - p_{FITn}^{\ t}$. The sample average absolute difference (over time periods *t* and countries *n*) is 12.2 percentage points. The within year absolute difference grows from 0 in 2000 to 25.3 percentage points in 2008 ([53]). The maximum absolute difference grows from 0 in 2000 to 83.4 percentage points in 2007. These are large differences in price levels, which translate into large differences in real GDP levels.

The relative volumes generated by dividing the US dollar GDP values by the corresponding US dollar prices defined by (38) are no longer independent of the choice of the numeraire country. Thus instead of taking the weighted arithmetic means of the prices $p_{Bn}^{\ 2000}$ and $p_{En}^{\ 2012}$, take the corresponding weighted geometric means and denote the resulting prices by $p_{FITGn}^{\ t}$ ([54]) Compare the prices $p_{FITGn}^{\ t}$ to our preferred Harmonized US dollar price levels pHnt listed in Table 6 and take the absolute value of the differences, $p_{Hn}^{\ t} - p_{FITGn}^{\ t}$. The sample average absolute difference is now 13.6 percentage points, which is larger than the average differences using the weighted arithmetic means. The within year absolute difference grows from 0 in 2000 to 28.6 percentage points in 2008 ([55]). The maximum absolute difference grows from 0 in 2000 to 84.7 percentage points in 2007.

Why do the above variants of the interpolation method suggested by Feenstra, Timmer and Inklaar generate US dollar price levels (and the corresponding country real GDP levels) that are so different from the Harmonized country US dollar price levels that are listed in Table 6? The reason is that the interpolated PPPs defined by (36) (and their geometric counterparts) depend on country inflation rates, which are quite variable ([56]). In order to eliminate the effects of country inflation rates, we tried the following variant of the FIT interpolation method: instead

([53]) The sequence of within year average absolute differences in percentage points over the 13 years is as follows: 0, 6.2, 6.2, 9.3, 14.5, 15.8, 14.7, 23.0, 25.3, 14.9, 9.4, 14.8, 4.2. The differences are nonzero in 2012 even though the corresponding PPPs for 2012 are exactly consistent with the OECD PPPs for 2012. Thus while the US dollar prices $p_{Hn}^{\ 2012}$ equal $\lambda p_{FITn}^{\ 2012}$ for $n = 1,\ldots,34$, the factor of proportionality λ is not equal to one and thus the differences are nonzero in 2012.

([54]) Usually, taking geometric means (rather than arithmetic means) of two indexes leads to indexes that have better invariance and homogeneity properties. For examples of this phenomenon, see Diewert (1997) and Hill and Fox (1997).

([55]) The sequence of within year average absolute differences in percentage points over the 13 years is as follows: 0, 5.2, 5.0, 10.8, 16.7, 18.6, 17.5, 26.3, 28.6, 16.2, 11.6, 16.0, 4.2.

([56]) More precisely, the FIT method interacts country inflation rates with the linear in time weights in equations (36) and *these weights are independent of the magnitude of economic price and quantity data that pertain to the countries* whereas our interpolation method depends only on country volume indexes over the sample period and the relative volumes generated by the PPPs at the beginning and end of the sample period. If the α_n defined by (33) were all equal to unity, then the matrix of country real volumes generated by extrapolating the base period relative GDP volumes forward by national growth rates would be equal to the matrix of country real volumes generated by extrapolating the final period relative GDP volumes backwards (after normalization to a common base) and our interpolation method would generate this common matrix of comparable over time and space real GDP volumes. Under the same conditions, the FIT method would not generate the same matrix (except by chance).

of using equations (36) to interpolate the PPPs between the years 2000 and 2012, use the following equations to define the interpolated $PPP_{In}^{\ t}$ for country n and year t:

(39) $PPP_{In}^{\ t} \equiv (1 - w^t)PPP_n^{\ 2000} + w^t\, PPP_n^{\ 2012}$; $n = 1,\ldots,34$; $t = 2000, 2001, \ldots, 2012$

where the weight w^t is still defined by (37). Now return to our description for the construction of the harmonized country estimates for GDP and US dollar price levels that is in Section 3 but replace $PPP_n^{\ t}$ in equations (12) by their interpolated counterparts $PPP_{In}^{\ t}$ defined by (39). Denote the resulting US dollar price levels by pInt. We compare the prices $p_{In}^{\ t}$ to our preferred US dollar price levels $p_{Hn}^{\ t}$ listed in Table 6 and as usual, take the absolute value of the differences, $p_{Hn}^{\ t} - p_{In}^{\ t}$. The sample average absolute difference (over time periods t and countries n) is only 2.48 percentage points. The within year average absolute difference grows from 0 in 2000 to 4.3 percentage points in 2005 ([57]). The maximum absolute difference is 20.6 percentage points in 2003. The performance of this interpolation method is much better than the previous interpolation method but still not quite as good as our suggested interpolation method that was described at the beginning of this section (which generated an average absolute difference of only 1.92 percentage points).

The above numerical experiments with interpolation methods that are similar in spirit to the method used by Feenstra, Inklaar and Timmer are not conclusive since it assumes that the 'truth' is best defined by the harmonized parities $p_{Hn}^{\ t}$ defined earlier in Section 3. However, at a minimum, the numerical experiments do show that the method of interpolation between benchmark Purchasing Power Parity rounds does matter. Additional research into alternative methods of interpolation is required.

7. Conclusion

A number of interesting points emerged from our investigations. If our focus is on measuring overall OECD GDP growth and PPP information is unavailable, then the method that is explained in Section 2 may be used. The overall OECD growth measures, γ^t, do not depend on PPPs or the choice of the numeraire currency but exchange rate fluctuations can cause perhaps unwarranted fluctuations. We computed γ^t using the US and then Germany as the numeraire country and found that while the Fisher index of OECD real GDP growth remained invariant, the accompanying Fisher price indexes, P^t and $P_{EU}^{\ t}$, exhibited wildly different rates of growth. Thus these price indexes are useless as indicators of OECD inflation.

Three alternative measures of overall OECD GDP growth were defined in Section 3: the Laspeyres, Paasche and Fisher measures, $\gamma_L^{\ t}$, $\gamma_P^{\ t}$ and $\gamma_F^{\ t}$. These measures depended on the annual OECD PPP information. The Laspeyres measure is the official OECD measure for overall OECD growth but the Fisher measure seems preferable on conceptual grounds. However, for our data set, all three measures were very close to each other.

The Fisher measure of OECD growth, $\gamma_F^{\ t}$, (which used real GDP share weights constructed using PPPs) grew on average about 1/10 of a percentage point more rapidly per year over the period 2000-2012 than our Section 2 measure of OECD GDP growth γ^t, (which used exchange rate based share weights). This result was expected since the Section 3 PPP based share weights $S_n^{\ t}$

([57]) The sequence of within year average absolute differences in percentage points over the 13 years is as follows: 0, 1.1, 1.9, 2.9, 4.0, 4.3, 3.7, 3.7, 3.6 2.3, 2.4, 1.9, 0.3. The sequence of within year maximum absolute differences in percentage points over the 13 years is as follows: 0, 6.7, 13.8, 20.6, 20.3, 18.6, 15.0, 13.3, 15.8, 8.2, 6.8, 5.8, 0.5.

for rich countries (which generally have lower rates of GDP growth) are generally smaller than the corresponding Section 2 exchange rate based share weights S_n^t.

Section 3 also introduced three measures of OECD aggregate GDP price inflation, the Laspeyres, Paasche and Fisher measures ρ_L^t, ρ_P^t and ρ_F^t defined by (19). These inflation measures used PPP based country share weights to weight the country inflation rates and were much more satisfactory than the Section 2 measures of OECD aggregate inflation. The three measures differed somewhat so the choice of index matters. Our preference is for the Fisher measure ρ_F^t since it satisfies a time reversal test whereas the other two indexes do not.

We used two principles in Section 3 to generate our *harmonized estimates of real GDP for OECD countries*: (i) The resulting harmonized estimates of country volumes q_{Hn}^t must be consistent with the real volume shares s_n^t listed in Table 3 and (ii) OECD aggregate real GDP growth must be equal to the rates of aggregate growth generated by our recommended Fisher indexes γ_F^t.

Once the harmonized estimates of real GDP q_{Hn}^t have been generated, companion *US dollar country price levels* p_{Hn}^t can be generated as $p_{Hn}^t \equiv v_n^t/q_{Hn}^t$ where v_n^t is the exchange rate converted US dollar nominal value of GDP for country *n* in year *t*. These country price levels are useful (but imperfect) indicators of a country's competitiveness in year t.

In Sections 4 and 5, alternative measures of comparable levels of real GDP and the accompanying US dollar price levels were constructed. The measures constructed in Section 4 used the PPP information for 2000 and national growth rates for real GDP by country whereas the estimates constructed in Section 5 substituted the PPP information for 2012. We found tremendous discrepancies in these estimates as compared to the harmonized estimates constructed in Section 3.

The results listed in Sections 3 to 5 show that it is very hazardous for analysts interested in comparative levels of GDP across countries to use national growth rates and a *single* cross country comparison of real GDP levels. Eventually, the single cross country comparison is replaced by another single cross country comparison but the new set of comparable GDP levels across time and space can be vastly different from the earlier set of GDP levels, particularly for small countries. These results reinforce the case for using the harmonized series that were defined in Section 3. Using the Section 3 methodology, the previously constructed harmonized estimates of relative GDP levels remain unchanged as another year of data is added.

If PPP computations for a group of countries are only done on an infrequent basis (rather than on an annual basis as is the case for the OECD), then the interpolation method explained in Section 6 may prove to be a useful method for obtaining comparable GDP levels that are consistent with the GDP relative levels for the two benchmark years. The results in Section 6 also indicate that different interpolation methods can generate very different results. In particular, the present interpolation method used in the Penn World Tables did not work well with our OECD data base.

Of course, the harmonization methods that have been suggested in this paper can be applied to any other value aggregate, such as consumption, investment or domestic absorption. The results in this paper show that if countries want to compare the size of their economies or measure expenditure growth or price inflation for a group of countries, it is absolutely essential that those countries undertake regular cross country comparisons of prices.

Acknowledgements

The authors thank two anonymous referees, Bert Balk, Yuri Dikhanov, Dennis Fixler, Anne-Sophie Fraisse, Robert Inklaar, Paul McCarthy, Prasada Rao, Marshall Reinsdorf and Paul Schreyer for helpful comments on an earlier draft. Financial support from the Social Sciences and Humanities Research Council of Canada and the Australian Research Council's Discovery Program (DP150100830) is gratefully acknowledged.

References

Balk, B.M.(2008), *Price and Quantity Index Numbers*, New York: Cambridge University Press.

Diewert, W.E. (1978), 'Superlative Index Numbers and Consistency in Aggregation', *Econometrica Vol.* 46, issue 4, pp. 883–900.

Diewert, W.E. (1992), 'Fisher Ideal Output, Input and Productivity Indexes Revisited', *Journal of Productivity Analysis*, No 3, pp. 211–248.

Diewert, W.E. (1993), 'The Early History of Price Index Research', pp. 33–65 in *Essays in Index Number Theory*, Vol. 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

Diewert, W.E. (1997), 'Commentary on Mathew D. Shapiro and David W. Wilcox: Alternative Strategies for Aggregating Prices in the CPI', *The Federal Reserve Bank of St. Louis Review*, Vol. 79:3, pp. 127–137.

Diewert, W.E. (2009), 'Similarity Indexes and Criteria for Spatial Linking', pp. 183–216 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Diewert, W.E. and K.J. Fox (2017), 'Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data, Discussion Paper 17-02, Vancouver School of Economics, University of British Columbia, Canada.

Eurostat (2012), *Eurostat-OECD Methodological Manual on Purchasing Power Parities*, 2012 Edition, Luxembourg: Publications Office of the European Union.

Feenstra, R.C., R. Inklaar and M.P. Timmer (2013), 'The Next Generation of the Penn World Table', NBER Working Paper, No. 19255, Cambridge MA: NBER.

Fisher, I. (1911), *The Purchasing Power of Money*, London: Macmillan.

Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton-Mifflin.

Griliches, Z. (1985), 'Data and Econometricians — The Uneasy Alliance', *American Economic Review,* Vol. 75, pp. 196–200.

Hill, R.J. (2001), 'Measuring Inflation and Growth Using Spanning Trees', *International Economic Review,* Vol. 42, pp. 167–185.

Hill, R.J. (2004), 'Constructing Price Indexes Across Space and Time: The Case of the European Union', *American Economic Review,* Vol. 94, pp.1379–1410.

Hill, R.J. (2009), 'Comparing Per Capita Income Levels Across Countries Using Spanning Trees: Robustness, Prior Restrictions, Hybrids and Hierarchies', pp. 217–244 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Hill, R.J. and K.J. Fox (1997), 'Splicing Index Numbers', *Journal of Business and Economic Statistics*, Vol. 15, pp. 387–389.

Hill, T.P. (1988), 'Recent Developments in Index Number Theory and Practice', *OECD Economic Studies,* 10, pp. 123–148.

Hill, T.P. (1993), 'Price and Volume Measures', pp. 379–406 in *System of National Accounts 1993,* Eurostat, IMF, OECD, UN and World Bank, Luxembourg, Washington, D.C., Paris, New York, and Washington, D.C.

Lehr, J. (1885), *Beiträge zur Statistik der Preise*, Frankfurt: J.D. Sauerlander.

Marshall, A. (1887), 'Remedies for Fluctuations of General Prices', *Contemporary Review*, Vol. 51, pp. 355–375.

McCarthy, P. (2013), 'Extrapolating PPPs and Comparing ICP Benchmark Results', pp. 473–505 in *Measuring the Real Size of the World Economy*, Washington D.C., World Bank.

OECD (2001), 'Introduction of the Euro in OECD Statistics', Statistics Directorate, National Accounts, STD/NA(2001)8, Paris: OECD, September 14.

OECD (2014), 'Methodological Notes: Compilation of G20 Consumer Price Index', Paris: OECD.

Rao, D.S.P., A. Rambaldi and B.M. Balk (2013), 'Purchasing Power Parities, Price Levels and Measures of Regional and Global Inflation', preliminary paper presented at the Economic Measurement Group Workshop 2013, November 28, University of New South Wales, Sydney, Australia.

Rao, D.P., A.N. Rambaldi, and H.E. Doran (2010), 'Extrapolation of Purchasing Power Parities using Multiple Benchmarks and Auxiliary Information: A New Approach', *The Review of Income and Wealth* 56: Supplement, pp. 59–S98.

Rao, D.S.P., A.N. Rambaldi, and H.E. Doran (2011), 'An Econometric Approach to the Construction of Complete Panels of Purchasing Power Parities: Analytical Properties and Empirical Results', School of Economics, University of Queensland, St Lucia, QLD 4072, Australia.

Summers, R. and A. Heston (1991), 'The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988', *Quarterly Journal of Economics*, Vol. 106, pp. 327–368.

Szulc, B.J. (1983), 'Linking Price Index Numbers,' pp. 537–566 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

World Bank (2013), *Measuring the Real Size of the World Economy*, Washington D.C.: World Bank.

Annex

This Annex lists the underlying OECD data, some supplementary tables and notes. The source for all of the data listed in this Annex is OECD.Stat. The country price levels P_n^t using domestic currencies (normalized to equal unity in 2000) are listed in Table A1 and the corresponding volumes are listed in Table A2.

Table A1: OECD country price levels in national currencies P_n^t

n	P_n^{2000}	P_n^{2001}	P_n^{2002}	P_n^{2003}	P_n^{2004}	P_n^{2005}	P_n^{2006}	P_n^{2007}	P_n^{2008}	P_n^{2009}	P_n^{2010}	P_n^{2011}	P_n^{2012}
1	1.000	1.028	1.057	1.089	1.131	1.186	1.244	1.301	1.366	1.378	1.462	1.484	1.474
2	1.000	1.019	1.031	1.043	1.061	1.082	1.103	1.125	1.144	1.162	1.179	1.203	1.224
3	1.000	1.021	1.041	1.062	1.084	1.110	1.136	1.163	1.188	1.202	1.227	1.251	1.275
4	1.000	1.011	1.022	1.056	1.090	1.126	1.156	1.192	1.241	1.218	1.253	1.294	1.316
5	1.000	1.042	1.083	1.136	1.221	1.309	1.474	1.546	1.553	1.613	1.755	1.815	1.847
6	1.000	1.046	1.074	1.084	1.128	1.124	1.130	1.167	1.189	1.217	1.197	1.186	1.205
7	1.000	1.025	1.049	1.066	1.091	1.122	1.146	1.172	1.222	1.230	1.282	1.292	1.321
8	1.000	1.065	1.115	1.160	1.211	1.285	1.398	1.560	1.645	1.647	1.652	1.702	1.758
9	1.000	1.030	1.043	1.036	1.041	1.046	1.055	1.086	1.118	1.135	1.139	1.170	1.204
10	1.000	1.020	1.043	1.064	1.081	1.102	1.126	1.155	1.184	1.193	1.204	1.220	1.238
11	1.000	1.011	1.026	1.037	1.048	1.055	1.058	1.075	1.083	1.096	1.108	1.121	1.138
12	1.000	1.031	1.066	1.108	1.141	1.173	1.201	1.241	1.300	1.330	1.345	1.359	1.348
13	1.000	1.113	1.207	1.272	1.338	1.372	1.420	1.497	1.576	1.632	1.671	1.714	1.770
14	1.000	1.086	1.147	1.155	1.183	1.217	1.324	1.399	1.564	1.694	1.811	1.870	1.924
15	1.000	1.060	1.118	1.160	1.187	1.215	1.256	1.277	1.241	1.193	1.175	1.183	1.191
16	1.000	1.018	1.059	1.053	1.054	1.064	1.084	1.082	1.099	1.152	1.166	1.194	1.234
17	1.000	1.029	1.062	1.095	1.121	1.142	1.161	1.189	1.219	1.244	1.249	1.266	1.288
18	1.000	0.988	0.973	0.956	0.943	0.931	0.921	0.912	0.901	0.896	0.877	0.860	0.853
19	1.000	1.039	1.072	1.110	1.144	1.152	1.150	1.174	1.208	1.249	1.295	1.314	1.327
20	1.000	1.001	1.022	1.082	1.102	1.155	1.233	1.279	1.284	1.294	1.387	1.445	1.489
21	1.000	1.058	1.132	1.229	1.338	1.395	1.490	1.573	1.673	1.742	1.812	1.920	1.989
22	1.000	1.051	1.091	1.115	1.123	1.150	1.171	1.192	1.218	1.219	1.229	1.243	1.259
23	1.000	1.037	1.039	1.067	1.102	1.124	1.162	1.216	1.249	1.258	1.318	1.338	1.327
24	1.000	1.017	0.999	1.028	1.089	1.186	1.291	1.330	1.475	1.396	1.483	1.584	1.624
25	1.000	1.035	1.058	1.062	1.106	1.135	1.152	1.197	1.235	1.280	1.299	1.340	1.373
26	1.000	1.036	1.075	1.107	1.134	1.163	1.195	1.229	1.248	1.260	1.268	1.271	1.267
27	1.000	1.050	1.091	1.149	1.216	1.245	1.282	1.296	1.333	1.317	1.324	1.345	1.362
28	1.000	1.087	1.169	1.234	1.274	1.295	1.323	1.378	1.435	1.482	1.467	1.484	1.487
29	1.000	1.042	1.087	1.133	1.178	1.230	1.280	1.322	1.354	1.355	1.356	1.356	1.356
30	1.000	1.024	1.039	1.058	1.061	1.071	1.091	1.121	1.157	1.180	1.190	1.206	1.218
31	1.000	1.013	1.019	1.027	1.036	1.038	1.061	1.088	1.118	1.113	1.117	1.121	1.122
32	1.000	1.529	2.101	2.589	2.910	3.117	3.407	3.619	4.054	4.268	4.510	4.897	5.229
33	1.000	1.023	1.048	1.071	1.096	1.118	1.150	1.176	1.214	1.241	1.279	1.309	1.331
34	1.000	1.023	1.039	1.059	1.088	1.123	1.158	1.189	1.212	1.221	1.236	1.260	1.282

Source: Author's calculations

It can be seen that Country 18, Japan, had the lowest rate of domestic inflation, which was actually a deflation. Country 32, Turkey, had the highest rate of domestic inflation, which was 533 % over the sample period. The countries that exhibited the fastest rates of real GDP growth over the sample period were countries 5 (Chile), 27 (Slovak Republic), 32 (Turkey), 19 (Korea) and 25 (Poland) with growth rates equal to 168 %, 167 %, 162 %, 159 % and 156 % respectively.

Table A2: OECD Country GDP volumes relative to the corresponding 2000 volumes, $Q_n^{\,t}/Q_n^{\,2000}$

n	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1	1.000	1.039	1.072	1.116	1.152	1.187	1.232	1.278	1.299	1.327	1.359	1.405	1.456
2	1.000	1.009	1.026	1.035	1.061	1.087	1.127	1.168	1.185	1.140	1.160	1.193	1.203
3	1.000	1.008	1.022	1.030	1.064	1.082	1.111	1.143	1.155	1.122	1.148	1.169	1.167
4	1.000	1.018	1.048	1.067	1.101	1.134	1.166	1.192	1.200	1.167	1.204	1.234	1.256
5	1.000	1.033	1.061	1.101	1.178	1.250	1.322	1.390	1.436	1.421	1.503	1.591	1.679
6	1.000	1.031	1.053	1.093	1.145	1.222	1.308	1.383	1.426	1.361	1.395	1.420	1.406
7	1.000	1.007	1.012	1.016	1.039	1.064	1.101	1.118	1.109	1.046	1.061	1.072	1.068
8	1.000	1.063	1.133	1.221	1.298	1.413	1.556	1.672	1.603	1.377	1.412	1.547	1.608
9	1.000	1.023	1.042	1.063	1.106	1.139	1.189	1.252	1.256	1.149	1.187	1.220	1.210
10	1.000	1.018	1.028	1.037	1.064	1.083	1.110	1.135	1.134	1.098	1.117	1.140	1.140
11	1.000	1.015	1.015	1.011	1.023	1.030	1.068	1.103	1.115	1.058	1.100	1.137	1.145
12	1.000	1.042	1.078	1.142	1.192	1.219	1.286	1.332	1.329	1.287	1.223	1.136	1.064
13	1.000	1.037	1.084	1.126	1.180	1.226	1.274	1.276	1.287	1.200	1.213	1.232	1.211
14	1.000	1.039	1.041	1.066	1.150	1.233	1.291	1.368	1.384	1.293	1.240	1.274	1.291
15	1.000	1.050	1.107	1.148	1.196	1.269	1.339	1.405	1.375	1.287	1.274	1.301	1.303
16	1.000	0.998	0.998	1.012	1.062	1.114	1.179	1.248	1.299	1.314	1.379	1.443	1.489
17	1.000	1.019	1.023	1.023	1.040	1.050	1.073	1.091	1.079	1.019	1.037	1.042	1.016
18	1.000	1.004	1.007	1.023	1.048	1.061	1.079	1.103	1.091	1.031	1.079	1.073	1.094
19	1.000	1.040	1.114	1.145	1.198	1.246	1.310	1.377	1.409	1.413	1.503	1.558	1.590
20	1.000	1.025	1.067	1.085	1.132	1.192	1.251	1.333	1.323	1.250	1.288	1.313	1.311
21	1.000	1.000	1.007	1.021	1.063	1.098	1.153	1.192	1.207	1.134	1.195	1.241	1.288
22	1.000	1.019	1.020	1.024	1.046	1.068	1.104	1.147	1.168	1.125	1.142	1.153	1.139
23	1.000	1.037	1.089	1.133	1.175	1.214	1.235	1.278	1.256	1.274	1.276	1.304	1.346
24	1.000	1.020	1.035	1.045	1.087	1.115	1.141	1.171	1.172	1.152	1.158	1.172	1.208
25	1.000	1.012	1.027	1.066	1.123	1.164	1.237	1.320	1.388	1.411	1.465	1.532	1.561
26	1.000	1.020	1.028	1.018	1.034	1.042	1.057	1.082	1.082	1.051	1.071	1.058	1.024
27	1.000	1.035	1.082	1.134	1.191	1.271	1.377	1.521	1.609	1.529	1.597	1.644	1.674
28	1.000	1.029	1.069	1.100	1.149	1.195	1.264	1.352	1.398	1.287	1.303	1.313	1.279
29	1.000	1.037	1.065	1.098	1.134	1.174	1.222	1.265	1.276	1.227	1.224	1.225	1.205
30	1.000	1.013	1.038	1.062	1.107	1.142	1.191	1.231	1.223	1.162	1.238	1.274	1.286
31	1.000	1.012	1.014	1.015	1.039	1.067	1.107	1.150	1.175	1.152	1.186	1.207	1.220
32	1.000	0.943	1.001	1.054	1.153	1.249	1.336	1.398	1.407	1.339	1.462	1.590	1.625
33	1.000	1.022	1.045	1.087	1.121	1.157	1.189	1.230	1.221	1.157	1.177	1.190	1.191
34	1.000	1.010	1.027	1.056	1.096	1.133	1.163	1.184	1.181	1.147	1.176	1.198	1.231

Source: Author's calculations

In order to obtain the country volume levels $Q_n^{\,t}$ that match up with the price levels $P_n^{\,t}$ in Table A1, the entries in the rows labeled 1-34 need to be multiplied by the country volume levels for 2000, the $Q_n^{\,2000}$ for $n = 1,\ldots,34$. These year 2000 levels are as follows: 706.89, 208.47, 252.54, 1 076.58, 42 094.99, 2 269.70, 1 293.96, 6.16, 132.19, 1 439.60, 2 047.50, 135.04, 13 089.05, 683.75, 105.64, 506.17, 1 198.29, 509 860.00, 603 236.00, 22.00, 6 020.65, 417.96, 118.38, 1 481.24, 744.38, 1 27.32, 31.18, 18.57, 629.91, 2 265.45, 432.41, 166.66, 987.14, 10 289.70. The units are in billions of year 2000 domestic currency units.

Table A3 lists the country *n*, year *t* US dollar price levels for GDP, p_n^t, and Table A4 lists the corresponding volume levels, q_n^t. Note that $p_n^t q_n^t$ equals v_n^t, the year *t* value of country *n*'s GDP in current US dollar.

Table A3: OECD country price levels in US dollar at market exchange rates p_n^t

n	p_n^{2000}	p_n^{2001}	p_n^{2002}	p_n^{2003}	p_n^{2004}	p_n^{2005}	p_n^{2006}	p_n^{2007}	p_n^{2008}	p_n^{2009}	p_n^{2010}	p_n^{2011}	p_n^{2012}
1	1.000	0.917	0.991	1.219	1.435	1.562	1.616	1.877	1.976	1.854	2.312	2.640	2.633
2	1.000	0.990	1.054	1.278	1.430	1.461	1.502	1.671	1.819	1.753	1.695	1.815	1.707
3	1.000	0.991	1.063	1.301	1.461	1.498	1.547	1.728	1.889	1.813	1.764	1.888	1.779
4	1.000	0.970	0.967	1.119	1.244	1.379	1.513	1.649	1.728	1.582	1.807	1.942	1.955
5	1.000	0.885	0.848	0.887	1.081	1.262	1.500	1.596	1.604	1.552	1.856	2.024	2.049
6	1.000	1.062	1.266	1.483	1.693	1.810	1.930	2.220	2.689	2.463	2.420	2.587	2.377
7	1.000	0.995	1.074	1.308	1.472	1.512	1.558	1.740	1.937	1.854	1.843	1.945	1.843
8	1.000	1.034	1.139	1.420	1.632	1.733	1.902	2.316	2.610	2.483	2.375	2.566	2.450
9	1.000	1.001	1.066	1.269	1.403	1.412	1.436	1.614	1.778	1.711	1.637	1.766	1.679
10	1.000	0.991	1.065	1.303	1.457	1.488	1.533	1.716	1.883	1.798	1.731	1.840	1.727
11	1.000	0.982	1.048	1.270	1.413	1.424	1.440	1.597	1.723	1.653	1.592	1.692	1.587
12	1.000	0.990	1.076	1.341	1.519	1.564	1.616	1.822	2.042	1.981	1.910	2.026	1.858
13	1.000	1.096	1.320	1.600	1.863	1.939	1.904	2.300	2.584	2.276	2.267	2.406	2.218
14	1.000	0.877	0.984	1.183	1.325	1.519	1.483	1.717	1.398	1.077	1.165	1.268	1.209
15	1.000	1.029	1.142	1.421	1.600	1.640	1.710	1.898	1.973	1.799	1.689	1.785	1.661
16	1.000	0.987	0.911	0.943	0.959	0.967	0.992	1.074	1.249	1.194	1.271	1.360	1.304
17	1.000	0.999	1.085	1.341	1.511	1.541	1.581	1.766	1.938	1.876	1.795	1.910	1.796
18	1.000	0.876	0.836	0.889	0.939	0.911	0.853	0.835	0.939	1.032	1.077	1.162	1.152
19	1.000	0.910	0.969	1.054	1.130	1.272	1.362	1.429	1.240	1.107	1.266	1.341	1.332
20	1.000	0.972	1.044	1.325	1.485	1.558	1.679	1.900	2.041	1.951	1.993	2.180	2.076
21	1.000	1.071	1.109	1.077	1.121	1.211	1.292	1.361	1.421	1.219	1.356	1.461	1.428
22	1.000	1.021	1.115	1.366	1.514	1.553	1.594	1.771	1.936	1.838	1.767	1.875	1.756
23	1.000	0.959	1.057	1.364	1.607	1.743	1.659	1.968	1.932	1.730	2.091	2.327	2.366
24	1.000	0.996	1.102	1.278	1.422	1.621	1.772	1.997	2.302	1.953	2.160	2.488	2.457
25	1.000	1.099	1.127	1.187	1.314	1.524	1.613	1.880	2.227	1.784	1.872	1.966	1.832
26	1.000	1.006	1.098	1.356	1.529	1.570	1.627	1.826	1.985	1.900	1.822	1.918	1.767
27	1.000	1.000	1.108	1.438	1.735	1.848	1.987	2.416	2.872	2.796	2.679	2.858	2.675
28	1.000	0.997	1.083	1.326	1.474	1.496	1.541	1.752	1.953	1.913	1.805	1.916	1.775
29	1.000	1.012	1.111	1.387	1.588	1.660	1.743	1.964	2.152	2.043	1.949	2.046	1.891
30	1.000	0.908	0.978	1.199	1.323	1.312	1.355	1.520	1.608	1.413	1.513	1.702	1.647
31	1.000	1.013	1.104	1.288	1.407	1.408	1.429	1.531	1.743	1.728	1.809	2.132	2.021
32	1.000	0.780	0.871	1.079	1.277	1.450	1.491	1.737	1.947	1.722	1.876	1.828	1.820
33	1.000	0.973	1.038	1.156	1.326	1.343	1.398	1.555	1.475	1.277	1.306	1.386	1.389
34	1.000	1.023	1.039	1.059	1.088	1.123	1.158	1.189	1.212	1.221	1.236	1.260	1.282

Source: Author's calculations

Table A4: OECD country GDP volumes in US dollar units of measurement q_n^t

n	q_n^{2000}	q_n^{2001}	q_n^{2002}	q_n^{2003}	q_n^{2004}	q_n^{2005}	q_n^{2006}	q_n^{2007}	q_n^{2008}	q_n^{2009}	q_n^{2010}	q_n^{2011}	q_n^{2012}
1	409.8	425.9	439.3	457.5	472.1	486.5	504.9	523.9	532.5	543.7	556.9	575.6	596.8
2	192.1	193.7	197.0	198.7	203.9	208.7	216.4	224.4	227.6	218.9	222.8	229.1	231.1
3	232.7	234.6	237.7	239.7	247.5	251.8	258.6	266.0	268.6	261.1	267.2	271.9	271.5
4	724.9	737.8	759.4	773.7	797.8	821.9	845.1	863.7	869.7	845.6	872.8	894.8	910.1
5	78.0	80.6	82.7	85.9	91.9	97.5	103.1	108.4	112.0	110.8	117.2	124.1	131.0
6	58.8	60.6	61.9	64.3	67.3	71.9	76.9	81.3	83.8	80.0	82.0	83.5	82.7
7	160.1	161.2	162.0	162.6	166.3	170.4	176.2	179.0	177.6	167.5	169.8	171.6	171
8	5.7	6.0	6.4	6.9	7.4	8.0	8.8	9.5	9.1	7.8	8.0	8.8	9.1
9	121.8	124.6	126.9	129.4	134.8	138.7	144.8	152.5	153	139.9	144.6	148.6	147.3
10	1 326.3	1 350.7	1 363.2	1 375.5	1 410.5	1 436.3	1 471.7	1 505.3	1 504.1	1 456.8	1 481.9	1 511.9	1 512.1
11	1 886.4	1 915.0	1 915.2	1 908	1 930.1	1 943.3	2 015.2	2 081.1	2 103.7	1 995.4	2 075.5	2 144.7	2 159.4
12	125.9	131.2	135.7	143.8	150.1	153.5	162.0	167.7	167.3	162.1	154.1	143.1	134.0
13	46.4	48.1	50.3	52.2	54.7	56.9	59.1	59.2	59.7	55.7	56.2	57.1	56.2
14	8.7	9.0	9.1	9.3	10.0	10.7	11.2	11.9	12.0	11.2	10.8	11.1	11.2
15	97.3	102.2	107.7	111.7	116.4	123.5	130.3	136.8	133.8	125.3	124.0	126.6	126.8
16	124.1	123.9	123.8	125.7	131.8	138.3	146.3	154.9	161.3	163.1	171.2	179.1	184.8
17	1 104.0	1 124.6	1 129.6	1 129.1	1 148.7	1 159.4	1 184.9	1 204.8	1 190.9	1 125.4	1 144.8	1 150.3	1 121.2
18	4 731.2	4 748.0	4 761.8	4 842	4 956.3	5 020.9	5 105.9	5 217.8	5 163.5	4 878.1	5 105.0	5 075.9	5 175.2
19	533.4	554.6	594.2	610.9	639.1	664.4	698.8	734.5	751.4	753.8	801.4	830.9	847.9
20	20.3	20.8	21.6	22.0	22.9	24.2	25.3	27.0	26.8	25.3	26.1	26.6	26.6
21	636.7	636.5	641.4	650.4	676.8	699.0	734.3	759.0	768.3	722.2	760.7	790.5	820.4
22	385.1	392.5	392.8	394.1	402.9	411.2	425.1	441.8	449.8	433.3	439.9	444.0	438.5
23	53.8	55.8	58.6	61.0	63.2	65.3	66.4	68.7	67.5	68.5	68.6	70.1	72.4
24	168.3	171.6	174.2	175.9	182.9	187.6	191.9	197.0	197.2	193.9	194.9	197.2	203.3
25	171.3	173.3	175.8	182.6	192.4	199.4	211.8	226.1	237.7	241.6	251.0	262.3	267.4
26	117.3	119.6	120.5	119.4	121.3	122.2	124.0	126.9	126.9	123.2	125.6	124.1	120.1
27	20.4	21.1	22.1	23.1	24.3	25.9	28.1	31.0	32.8	31.2	32.6	33.6	34.2
28	20.0	20.6	21.4	22.0	22.9	23.9	25.3	27.0	27.9	25.7	26.0	26.2	25.6
29	580.3	601.6	617.9	637.0	657.8	681.4	709.1	733.8	740.4	712	710.6	710.9	699.2
30	247.3	250.4	256.6	262.6	273.7	282.4	294.5	304.3	302.4	287.2	306.0	315.0	318.0
31	256.0	259.2	259.7	259.8	266.0	273.2	283.5	294.4	300.7	294.9	303.6	309.1	312.3
32	266.6	251.4	266.9	280.9	307.2	333.0	356.0	372.6	375.1	357.0	389.7	423.8	433.0
33	1 493.6	1 526.2	1 561.2	1 622.9	1 674.4	1 728.5	1 776.2	1 837.0	1 822.9	1 728.6	1 757.3	1 777	1 779.2
34	10 289.7	10 387.3	10 571.8	10 866.9	11 279.6	11 657.6	11 968.5	12 182.7	12 147.3	11 806.9	12 102.9	12 326.5	12 669.0

Source: Author's calculations

It is useful to spell out in a bit more detail what the US dollar period t share weights for country n, $S_n^t \equiv v_n^t / \Sigma_{i=1}^N v_i^t$ look like in terms of domestic values and exchange rates. Let e_n^t be the number of units of domestic currency it takes for country n to purchase one dollar of US currency in year t. Then country n's domestic currency value for its GDP in year t, V_n^t, is related to the corresponding US dollar value for year t, v_n^t by the equations $V_n^t = v_n^t e_n^t$ for all t and n. Similarly, the domestic currency price level for country n in year t, P_n^t, is related to the corresponding US dollar price level for year t, p_n^t by the equations $P_n^t = p_n^t e_n^t$ for all t and n. Substituting these equations into the definition for the country US dollar shares of OECD GDP, we find that $S_n^t = [V_n^t / e_n^t] / [\Sigma_{i=1}^N (V_i^t / e_i^t)]$ for all n and t. Note that these shares do not depend on which country is chosen as the numeraire currency. The country GDP values in domestic currencies, V_n^t, will move smoothly over time but exchange rate fluctuations will introduce erratic movements in the shares S_n^t over time t.

Table A5: Annual purchasing power parities PPP_n^t for OECD countries 2000-2012, national currencies per US dollar

n	PPP_n^{00}	PPP_n^{01}	PPP_n^{02}	PPP_n^{03}	PPP_n^{04}	PPP_n^{05}	PPP_n^{06}	PPP_n^{07}	PPP_n^{08}	PPP_n^{09}	PPP_n^{10}	PPP_n^{11}	PPP_n^{12}
1	1.316	1.330	1.336	1.351	1.363	1.388	1.408	1.423	1.479	1.437	1.498	1.493	1.458
2	0.900	0.917	0.896	0.885	0.874	0.886	0.857	0.868	0.852	0.849	0.845	0.846	0.831
3	0.891	0.886	0.865	0.879	0.897	0.900	0.884	0.887	0.874	0.863	0.864	0.867	0.849
4	1.232	1.218	1.229	1.226	1.231	1.214	1.208	1.211	1.234	1.199	1.218	1.233	1.227
5	285.108	289.501	296.891	307.807	321.76	333.69	320.257	323.512	339.271	350.588	347.852	334.241	334.206
6	14.212	14.222	14.319	14.034	14.291	14.316	14.053	13.945	14.262	13.977	14.243	13.899	13.700
7	8.409	8.468	8.302	8.537	8.404	8.59	8.336	8.235	8.012	7.877	7.821	7.857	7.736
8	0.455	0.477	0.477	0.481	0.486	0.502	0.521	0.555	0.549	0.527	0.532	0.541	0.550
9	0.995	1.012	1.003	1.011	0.975	0.977	0.951	0.941	0.918	0.908	0.925	0.938	0.929
10	0.939	0.919	0.905	0.938	0.940	0.923	0.904	0.893	0.882	0.866	0.869	0.866	0.857
11	0.967	0.955	0.942	0.917	0.897	0.867	0.838	0.831	0.812	0.814	0.811	0.803	0.789
12	0.678	0.671	0.660	0.689	0.696	0.714	0.700	0.719	0.701	0.701	0.713	0.714	0.678
13	107.885	110.652	114.88	120.516	126.307	128.594	128.637	131.336	129.429	126.256	129.005	130.345	128.453
14	84.311	88.930	91.342	94.484	94.248	99.078	107.307	113.108	117.421	125.692	136.066	139.737	140.967
15	0.962	0.993	1.004	1.014	1.006	1.01	0.985	0.958	0.952	0.897	0.853	0.836	0.818
16	3.443	3.426	3.463	3.629	3.535	3.717	3.836	3.720	3.867	3.947	3.943	3.885	3.942
17	0.817	0.808	0.845	0.854	0.873	0.867	0.834	0.817	0.789	0.784	0.800	0.796	0.776
18	155.113	149.857	143.774	139.824	134.161	129.552	124.864	120.216	116.846	116.348	112.418	108.812	105.972
19	746.206	757.829	769.772	794.282	795.998	788.92	774.815	768.65	785.718	811.664	829.897	833.034	826.191
20	0.940	0.948	0.934	0.942	0.923	0.953	0.915	0.925	0.906	0.912	0.929	0.926	0.915
21	6.099	6.311	6.554	6.815	7.217	7.127	7.181	7.370	7.470	7.409	7.604	7.532	7.668
22	0.893	0.906	0.902	0.927	0.909	0.896	0.869	0.857	0.842	0.846	0.850	0.843	0.829
23	1.442	1.473	1.469	1.497	1.510	1.535	1.486	1.506	1.491	1.454	1.492	1.481	1.446
24	9.129	9.180	9.111	9.112	8.988	8.896	8.701	8.776	8.752	9.006	9.058	9.095	8.824
25	1.841	1.861	1.829	1.841	1.861	1.869	1.846	1.843	1.857	1.875	1.852	1.877	1.868
26	0.700	0.706	0.708	0.706	0.716	0.684	0.663	0.660	0.649	0.637	0.636	0.633	0.618
27	0.526	0.522	0.528	0.555	0.573	0.566	0.556	0.546	0.533	0.514	0.523	0.531	0.522
28	0.532	0.565	0.588	0.615	0.611	0.612	0.608	0.629	0.634	0.648	0.652	0.644	0.625
29	0.734	0.740	0.733	0.753	0.759	0.765	0.736	0.728	0.720	0.713	0.721	0.718	0.695
30	9.135	9.349	9.352	9.335	9.105	9.378	9.094	8.886	8.773	8.965	9.067	8.935	8.668
31	1.851	1.840	1.771	1.776	1.754	1.743	1.660	1.601	1.549	1.527	1.506	1.448	1.389
32	0.283	0.428	0.613	0.773	0.812	0.831	0.848	0.864	0.890	0.917	0.954	1.030	1.044
33	0.636	0.627	0.628	0.641	0.633	0.636	0.627	0.645	0.651	0.660	0.667	0.679	0.661
34	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Source: Author's calculations

The exchange rate based Laspeyres and Paasche price indexes P_L^t and P_L^t that appear in Table 2 are built up using the following chain link indexes: $P_{L,link}^{t} \equiv \sum_{n=1}^{N} S_n^{t-1}(p_n^t/p_n^{t-1}) = \sum_{n=1}^{N} [V_n^{t-1}/e_n^{t-1}] [(P_n^t/e_n^t)/(P_n^{t-1}/e_n^{t-1})]/[\sum_{i=1}^{N} (V_i^{t-1}/e_i^{t-1}1)]$ and $P_{P,link}^{t} \equiv [\sum_{n=1}^{N} S_n^{t} (p_n^t/p_n^{t-1})^{-1}]^{-1} = \{\sum_{n=1}^{N} [V_n^t/e_n^t] [(P_n^t/e_n^t)/(P_n^{t-1}/e_n^{t-1})]^{-1}]/[\sum_{i=1}^{N} (V_i^t/e_i^t)]\}^{-1}$. It can be seen that exchange rate fluctuations affect not only the domestic share weights in the above expressions but they also interact directly with the country inflation rates, P_n^t/P_n^{t-1}. Thus intertemporal exchange rate variation 'noise' will tend to drown out the country inflation trends.

Define the exchange rate based Laspeyres, Paasche and Fisher chain link volume indexes as $Q_{L,link}^{\ t} \equiv \sum_{n=1}^{N} S_n^{\ t-1} (q_n^{\ t}/q_n^{\ t-1}) = \sum_{n=1}^{N} S_n^{\ t-1} (Q_n^{\ t}/Q_n^{\ t-1})$, $Q_{P,link}^{\ t} \equiv [\sum_{n=1}^{N} S_n^{\ t} (q_n^{\ t}/q_n^{\ t-1})^{-1}]^{-1} = [\sum_{n=1}^{N} S_n^{\ t} (Q_n^{\ t}/Q_n^{\ t-1})^{-1}]^{-1}$ and $Q_{F,link}^{\ t} \equiv [Q_{L,link}^{\ t}\, Q_{P,link}^{\ t}]^{1/2}$. It can be seen that exchange rate fluctuations affect only the US dollar country shares (the $S_n^{\ t}$) and not the growth rates of country real GDP (the $Q_n^{\ t}/Q_n^{\ t-1}$) and thus these volume link indexes will be much more stable than their price counterparts. We note that the chained Fisher Q^t listed in Table 2 can be defined as follows: $Q^{2000} \equiv 26694.3$; $Q^t \equiv Q^{t-1}\, Q_{F,link}^{\ t}$; $t = 2001,\ldots,2012$.

2 Uses of national accounts from the 17th century till present and three suggestions for the future

FRITS BOS [1]

Abstract: This paper provides an overview of the roles and uses of national accounts since the seventeenth century. It shows that national account statistics, including their key-indicators like GDP and GDP-volume growth, have mainly been developed for fiscal and monetary policy purposes, for measuring progress in material well-being and for showing their relationship with the supply and use of goods and services in a market economy. National accounts statistics were not developed for measuring welfare, quality of life, inequality or sustainable development. This reflects that for such policy issues measurement in monetary terms is less meaningful and more difficult to compile reliably. In order to meet the complicated measurement issues and many data needs of the future, a sizable part of national accounts statistics should be 'slow-statistics', i.e. compiled regularly but not very frequently, e.g. every two or five years. To reduce misinterpretation of GDP-volume growth statistics and to stimulate sustainable growth, such slow-statistics should also include well-being accounts and net-net statistics on domestic product and national income, i.e. including a correction for the exhaustion of natural resources.

JEL codes: B00, E01, E60, Q56

Keywords: History of national accounting, guidelines on national accounting, macro-economic modeling and forecasting, measurement in economics

(1) CPB Netherlands Bureau for Economic Policy Analysis.

1. Introduction

The roles and uses of the national accounts have developed over time and have been stimulated by major events, like the economic crisis in the thirties, the Second World War, European unification and the recent financial crisis. In this paper, an overview is provided of these developments in roles and uses over time ([2]). Three different periods are distinguished.

The period 1660–1930 is labelled as the early estimates (Section 2). This period of nearly three centuries starts with the first national income estimates in England and France. All these early estimates were incidental and directed at concrete policy issues. They were therefore not only the start of measuring the national economy, but implied also the start of quantitative economic policy analysis.

The second period is very short, only two decades: 1930–1950. But they were revolutionary decades for national accounting (Section 3). Major features of this transitory period are:

- The start of official and regular national accounts statistics in a limited number of countries. These statistics were not directed at specific policy issues, but were intended to provide information relevant more to general policy and analysis;
- Invention of many new applications and uses of the national accounts approach, e.g. the Keynesian revolution in economic theory and policy, input-output analysis and econometric modelling;
- Revolution in national accounting concepts and methods, e.g. the birth of the first modern national accounting systems;
- Political and economic circumstances favouring a national accounts approach in policy and analysis (economic crisis of the thirties, the end of the gold standard, the Second World War and the need for recovery afterwards);

We are still living in the third period (1950–present). Considering their very important role, it could be labelled the era of the international guidelines (Section 4). Major features of this period are;

- A rapid extension of the number of countries for which official and regular national accounts statistics are available. This is accommodated by the appearance of private and public forecasts of national accounts statistics and their concepts. Decision-makers and researchers become gradually accustomed to using national accounts statistics;
- The publication of international guidelines. This was very important for accumulating and transferring of knowledge on national accounting. It was also important for harmonizing the concepts used by individual countries in compiling national accounts statistics.;

([2]) The best source on early national income estimates is still Studenski (1958); about early English estimates, see also Stone (1997). For the more recent developments, see e.g. Bos (2009a, chapters 2, 3 and 4), Bos (2011), Vanoli (2005), Vanoli (2014) and Kenessey (1994). Maddison (2003 and 2005) provides a historical overview mainly restricted to the measurement of economic growth. A major new insight on national accounts in Germany was provided by Tooze (2001). He suggests that Wagemann's work during the Weimar republic was a pioneering effort in empirical macro-economics. After the First World War, Wagemann was head of the German Statistical Office and of the German Institute for Business-Cycle Research. The Institute estimated and predicted the economic outlook for policy-makers based on an explicitly macro-economic model of the economy using time-series data. This work was not mentioned in the German history of national accounts provided by Reich (1994).

- Major conceptual developments, e.g. about how to measure prices and volumes. In fact, onlyrelatively recently detailed guidelines on measuring prices and volumes in the national accounts were published;
- The development of satellite accounts, e.g. on the environment or health care. This implies that the national accounts developed from not only a tool for macro-economic management but also for such specific policy areas;
- Political and economic circumstances, e.g. globalisation, European unification, the collapse of communism in Eastern Europe and China, the financial crisis and the growing role for international organisations (IMF, World Bank, United Nations, European Union and OECD), favoured a harmonized national accounts approach in policy analysis.

Despite these trends towards harmonization, there are still substantial differences between the roles and uses of national accounts statistics in countries all over the world and even within the European Union. These differences apply to the role in public decision-making, the relative importance of specific uses and the division of tasks between the national accounts, official policy reports and research outside the official national accounts. This reflects that the roles and uses of the national accounts depend also on the specific national circumstances, e.g. the political and institutional environment. This issue is illustrated by the development of the Dutch national accounts and the interaction with the budget, monetary policy and national expert institutes like the Dutch Central Bank, the Netherlands Bureau for Economic Policy Analysis and the Dutch Social and Cultural Planning Bureau (see Bos (2006) and Bos (2008a), pp. 33–51).

This paper serves various purposes. The first purpose is to serve as an antidote against simple one-dimensional views on the national accounts, e.g. 'national accounts is GDP, GDP is not a good measure of welfare and should therefore not be used for policy-making and analysis'.

The second purpose is to inform economists about the importance of national accounting and about the major recent developments in national accounting. This is necessary, as national accounting has become much more complex and a profession separate from economic theory and econometric model building. Better understanding of the national accounts may also stimulate economists to use national accounts statistics and contribute to its development.

The third purpose is to suggest a redirection of national accounting towards its major uses. What are the major uses of the national accounts in view of different economic, political and institutional circumstances? What is the link with the various types of economic analysis? What is the link with major policy debates? Clarifying this can help to make national accounts guidelines and (inter)national statistical programmes much more balanced and effective in view of these uses. At the end of this paper (Section 5), three suggestions are provided to improve the relevance of national accounts statistics in the future:

- More slow statistics for difficult to measure topics;
- Fewer rapid statistics that do not contain much information;
- Better explaining value added, use and misuse of national accounts statistics.

2. The early estimates (1660–1930)

The estimates of national income and wealth by Petty, King and Davenant in England and Boisguillebert and Vauban in France started 'Political Arithmetick' (see Table 1).

Table 1: Major events in the early estimates of national accounting

Year	Event
1660–1710	First national income estimates; in England by Petty, King and Davenant; in France by Boisguillebert and Vauban
1707	First index-numbers by Fleetwood
1760	Tableau économique by Quesnay: economic accounts used as a primitive growth and general equilibrium model; precursor of input-output tables
1770	The concept of value added invented by Young
1790–1800	First national income estimates in Russia
1798–1804	First national income estimates in the Netherlands
1805	First national income estimates in Germany
1823	First national income estimates in constant prices by Lowe
1843	First national income estimates in the USA
1886	First official national income estimates by the government (Australia, Coghlan)
1860–1900	First national income estimates in Austria, Australia, India and Greece
1915	W.I. King (USA): one of the last national income estimates combined with clear policy conclusions
1920	The economic consequences of the peace by Keynes: using national accounts statistics to assess the dramatic economic consequences of a major political agreement
1920–1930	Private institutions start publishing national income studies, e.g. university institutes in Sweden and Norway, USA: Brookings Institutions, NBER, National Industrial Conference Board; Austria WIFO
1925–1940	More official national income estimates (e.g. Greece, Canada, Soviet Russia, Germany, Netherlands, New Zealand, USA and Turkey)

2.1. England

At the end of the seventeenth century, Petty 'wanted [firstly] to prove mathematically that the State could raise a much larger revenue from taxes to finance its peace and wartime needs, and that it could do so by more equitable and less burdensome forms of taxation ... Secondly, Petty wanted to disprove once and for all the notion that England had been ruined by the Revolution and foreign wars and was no match, either militarily or commercially, for Holland and France' (Studenski (1958), p. 27, 28). King also draws clear political conclusions from his estimates:

'the Warr cannot well be sustain'd beyond the year 1698 upon the Foot it now stands, unlesse

(1) The Yearly Income of the Nation can be Increas'd.

(2) Or the Yearly Expence Diminish'd.

(3) Or a Forreign or Home Credit be Obtain'd or Establish'd.

(4) Or the Confederacy be Inlarg'd.

(5) Or the State of the Warr Alter'd.

(6) Or a General Excise, in effect Introduced.'
(King (1936), p. 47).

Nevertheless, Petty as well as King recognised also the more general advantages of estimating national wealth and income. King states that information on a country's wealth and population is a 'Piece of Political Knowledge, of all others, and at all times, the most useful, and Necessary' (King (1936), p. 13). Petty went even further by advocating that in socio-economic discussions 'no word might be used but what marks either number, weight, or measure' (Studenski (1958), p. 27).

The estimates by King can be regarded as improvements on those of Petty. We will shortly discuss the four main features of King's estimates.

Like the estimates of Petty and the earliest estimates in France, King employs *a comprehensive concept of production and income*. This concept is still used in the current international guidelines. According to this concept the production of goods as well as services generates income.

The second important feature of the estimates by King is that they already represented the three ways of estimating domestic product: net production, expenditure and distribution by type income. The estimates of Petty and most estimates up till the 1930's only covered one or two ways.

The third important feature of the calculations by King was their remarkable coverage. He presented not only the total annual national income, expenditure, and saving, but also their distribution by social and occupational groups, a breakdown of national income by type of income and an estimate of wealth (gold, silver, jewels, furniture, livestock, etc.). Like Petty, King provided a comparison of the national incomes and wealth of England, Holland and France. International comparison, which is a major objective of the international guidelines, was therefore already present in Petty's and King's pioneering estimates. King's estimates contained also time series of the period 1688–1695 of national income, expenditure and taxes received.

The fourth important feature of King's work is that he used his time series to forecast income, expenditure and tax revenue for the years 1696, 1697 and 1698. This type of use of national accounting figures dates therefore also back to the earliest estimates of national income.

From a modern perspective, the major limitation of King's work is that all estimates are in current prices. The first price index numbers were invented only some years later by Fleetwood (1707) ([3]). However, the idea of deflating national income is much more recent and originates from Lowe in 1823 (see Studenski (1958), pp. 107–109).

Lowe used his national income estimates to calculate the tax burden. He related them to the taxable income remaining after deducting the subsistence incomes that could not bear any taxation. He did this for the national incomes of specific years in the preceding three decades, calculated first at current prices and next at prices of 1792. Lowe devoted considerable space to the effects of inflation on fiscal policy. He was particularly concerned over the increase in the burden of public debt on the shrinking money value of national income. He proposed that the government enact a tabular standard that would control the value of money and the payments

([3]) The issue addressed by Fleetwood was the interpretation of values in old statutes, entitlements, laws and regulations (see Stone (1997), 117–140). For example, according to fifteenth century college statutes in Oxford, a fellow should vacate his fellowship if he came in possession of an estate by inheritance or of a perpetual pension worth of 5 pounds per year. According to Fleetwood, corrections should be made for all price changes during the past two centuries. He also mentioned similar problems with the fines for stealing mentioned in old laws and the level of income which entitled to vote in parliamentary elections. To translate such figures into current prices of several centuries later, he provided an overview of price changes for five commodities (wheat, oats, beans, ale and cloth) and recommended to calculate an unweighted index of average price change.

of interest and wages. This tabular standard was to be based on the 'power of purchase' of money in terms of basic commodities and could be applied on a voluntary basis.

For calculating national income, King used net product estimates for agriculture, i.e. he deducted from the revenues of the crop the expenditure on the seeds. However, he did not systematically use value added as a concept. This concept originates from Young (1770, see Stone (1997), pp. 141–181). For calculating value added in agriculture, Young did not only deduct the costs of seed, but also costs of maintenance and repair, e.g. of buildings, vehicles and horses.

2.2. France

In France, at the start of the eighteenth century, Vauban proposed a major tax reform (see Studenski (1958), pp. 55–60). Most of current taxes were to be abolished, as they were inequitable — the small incomes carried nearly all the tax burden — and responsible for poverty. They should be replaced by an income tax, levied at a uniform rate of 5 % to 10 %, depending on general economic conditions. The income tax should consist of two parts: a levy in kind on agricultural produce and a levy on all money incomes, i.e. on the rent of houses, profits of businesses and grain mills, the operation of public properties, wages, pensions and the fees of government offices. In addition to the income tax, current taxes on salt, postal charges, customs duties and taxes on luxury goods like tea, coffee, chocolate, brandy and gilt coaches should be retained and a tax on wine, cider and beer consumed in public houses should be added. He then estimated the potential revenue of his tax reform. To this end, he made an estimate of the major elements of French national income, but did not present an estimate of total French national income. He was also not aware that distinguishing between the gross and net value of agricultural production is necessary for ensuring fairness for tax purposes.

Vauban was a military engineer world famous for his fortifications and tactics for attacking and defending them. He was decorated by the French king Louis XIV with every conceivable order of distinction. He made his proposals for tax reform after his retirement. However, these proposals were not all appreciated by the ruling class. At the court, Vauban fell therefore into disgrace and he died some months later.

About half a century later, Quesnay published his zigzag-diagram, which is a major precursor of modern input-output tables. Its purpose was 'to construct a fundamental Tableau of the economic order for the purpose of displaying expenditure and products in a way which is easy to grasp, and for the purpose of forming a clear opinion about the organization and disorganization which the government can bring about' (Translation by Meek (1962), p. 108). Quesnay's Tableau économique was used as a primitive growth model that served to promote the idea that agriculture and not merchandising or manufacturing is the engine of economic growth ([4]). According to the physiocrats, investments in agriculture should therefore be promoted, taxes and interest rates for agriculture should be reduced and tolls and other restrictions on trade in agricultural products should be abolished. The table also served to argue that the most efficient way of taxing is to directly tax the group that ultimately pays the tax, i.e. the landlords instead of the farmers or the artisans ([5]). The table was therefore also a simple general equilibrium model.

([4]) On the interpretation of Quesnay's Tableau économique as a growth model, see Eltis (1984), in particular the first two chapters.

([5]) However, the physiocrats were certainly not enemies of the landlords, as they also argued that the net surplus of agriculture should be sufficiently high (e.g. by raising the prices of agricultural products) to pay the taxes and to give them sufficient income.

Quesnay's Tableau économique is a clear economic accounting model. As such it can be regarded as the first precursor of both the input-output tables and the sector accounts. The estimates by King, Petty and Vauban were systematic, but did not stress the circular flow of income and expenditure and the interactions between socio-economic groups.

2.3. The Netherlands

Political arithmetic had a surprisingly slow start in the Netherlands ([6]), by far the richest country in the world ([7]) in the seventeenth century. In 1798, on request of the national assembly of the new Batavian republic, Hora Siccama and van Rees made a plan for revising the tax system (Hora Siccama and van Rees (1789)). Their task was to investigate how taxes could be levied efficiently, in proportion to personal wealth, with as low as possible tax rates that were sufficient for financing government expenditure. Estimates of national income, government expenditure and current and potential tax revenue were included in the plan. The plan, including a proposal for introducing personal income and wealth tax, was not executed due to a revolt and a change in government.

In 1804, Metelerkamp published a study of 600 pages on the economic and military power of the Batavian republic over the last fifty years and in comparison with seven other countries (Great Britain, France, Russia, Prussia, Sweden, Denmark and Saxony). The study was published in Dutch and also translated in French (Metelerkamp (1804)). He compared the size of the country, the number of inhabitants, national income, national wealth, imports and exports, income and outlay of the government, government debt and the number of soldiers and warships. However, the comparison of national income was limited to the Netherlands (including Belgium), Great Britain, France and Saxony.

Metelerkamp's study was a reaction to the pessimistic mood in the Netherlands: people thought their country was in decline, full of debt, losing morale and fighting spirit and with shrinking revenue from international trade and colonies (Metelerkamp, 1804, Introduction, part I). He wanted to investigate whether facts corroborated these feelings and opinions and what policies should be pursued. He concluded that:

- The example of Russia shows that government should interfere as little as possible with domestic trade (hunting, fishing, farming, craftsmanship and commerce); this stimulates employment, income and the supply of products;
- Dutch welfare did not decline during the past fifty years; some sources of income declined indeed, but were compensated by the increase in other sources of income;
- Welfare in other European countries increased. This implied a relative decline of Dutch welfare. However, Dutch welfare still surpassed welfare in other European countries;
- During the last century, government revenue increased in other European countries much more than in the Netherlands. Relatively low taxes may have been the reason for the much more rapid increase in Dutch government debt. However, the tax burden in the Netherlands was now excessive and would in the long run certainly ruin the economy.

([6]) On the history of the Dutch national accounts, see Bos (2006) and Bos (2008a).
([7]) In 1700, Dutch GDP per capita was 70 % higher than that of England and 130 % higher than that of France (see Maddison (2003), p. 58–59).

2.4. Major features of the early estimates

All early estimates of national income were practical and directed at concrete policy issues; this was a common feature of national income studies up to the 1920's. The policy issues addressed were national economic power and performance, poverty, unfair and inefficient taxation and sustainability of public finance. Often, several of these issues were discussed and the national accounts approach was essential for demonstrating that the various issues were intimately linked.

War, substantial economic decline and wide-spread poverty were circumstances that stimulated the early estimates. Also the availability of census data or income tax data was important; this partly explains the dominant role of the English estimates.

The ruling class was often not very happy with the national income estimates and the accommodating proposals for reform. Early national accountants were sometimes exiled (e.g. Radishchev in Russia) or fell into disgrace (e.g. Vauban in France); others may have feared the consequences of publishing their work and left it therefore unpublished. However, in the twentieth century -probably linked to advancements made in democracy- estimating national income became gradually to be perceived as a task of the government. In many countries, also private institutions took the responsibility of regularly compiling national accounts statistics and producing national income studies. Examples are university institutes (e.g. in Sweden and Norway), economic research institutes (e.g. NBER, Brookings Institution and National Industrial Conference board in USA, Mitsubishi Economic Research Institute in Japan and WIFO in Austria) or private and central Banks (e.g. the Bank of Nova Scotia in Canada). Such institutions reflect the interest of researchers, business, banks, pension funds and trade unions in national income estimates. They need such information for analyzing and monitoring the economy, calculating market shares and deciding on investment, loans, mortgages and wage negotiations.

National accounting had a brilliant start with the work by Petty and King. Since then, progress in national accounting was often slow and small and there were major cases of regress, like the production boundary used.

For three quarters of a century, Adam Smith was very influential in his argument that labourers in agriculture as well as in manufacturing, commerce and the transportation of goods were to be regarded as 'productive'. However, unlike King, he still rated 'the whole civil and military personnel of government, the professions, the domestics, and others engaged in the performance of personal services and the services of dwellings' as unproductive labourers. 'He considered the national product to be constituted solely of commodities, and the national income ... to be composed of wages, rent and profit (including interest) derived from the production of these articles' (Studenski (1958), p. 19). Smith's view was supported by among others Ricardo, Malthus, James Mill and John Stuart Mill, but became increasingly subject to criticism by, e.g., Say, McCulloch, Senior, Walras and Marshall.

Despite such unfortunate intellectual detours, at the beginning of the twentieth century, the common stock of knowledge on national accounting had already become considerable. It included for example a comprehensive production boundary treating e.g. agriculture and government services as productive, the notion of three basic ways to estimate domestic product and the concepts of value added and constant prices.

2.5. Keynes: 'The economic consequences of the peace'

Directly after the first world war, Keynes wrote 'The economic consequences of the peace' (Keynes (1920)). This polemic made him instantly world-famous. He was the official British representative at the start of the Paris Peace Conference, but resigned when it became evident that there was no hope for substantial modification of the draft terms of peace. He attacked the peace-treaty and the reparation payments imposed on Germany for not taking into account facts and economic logic.

On the basis of all kinds of statistics on national wealth in e.g. Belgium, France, Great Britain and Serbia, he estimated that the treaty implied that Germany had to pay the various allied powers 2 billion pound for direct war damage ([8]) and 5 billion for pensions and allowances. He also calculated that Germany's capacity to pay was not sufficient for this. Transferable wealth (e.g. gold, silver, ships and foreign securities) was only about 0.5 billion pound and the rest had therefore to be paid over a number of years out of the net revenue from exports. However, considering the loss of major parts of former Germany (e.g. Alsace-Lorraine) and the loss in capital stock (e.g. ships, livestock, wear and tear due to lack of repair), the transferable part of national income would be less than 0.1 billion pound per year. He proposed to limit Germany's reparation payments well within her capacity to pay, to ensure a fair distribution of coal and iron and to establish a free trade union under the auspices of the League of Nations. Only in such a way, the European economy could be revived, avoid the perils of mass inflation and ensure a proper standard of living for the whole population.

Keynes' polemic could be regarded as a revolt of economics against politics and in this revolt national accounts statistics played a major role. It is still a major example of the use of national accounts statistics for (economic) policy analysis. It marked a radical shift in Keynes' thought and its prophetic lessons were learned after the Second World War, e.g. the Marshall aid and the European Union. All these developments turned out to be very important for national accounting (see Sections 3 and 4).

3. Revolutionary decades (1930–1950)

3.1. Major changes in uses

In the period 1930–1950, national accounting was drastically transformed. It was a revolution in terms of the use of the national accounts (see Table 2).

Colin Clark (see e.g. Clark (1937 and 1940)) introduced purchasing power parities and showed how to make international comparisons of real income. He also demonstrated how to make intertemporal comparisons. He even made a comparison between the level of well-being in the ancient world (Egypt, Greece and the Roman Empire at the peak of their powers) and that in the nineteenth century and the first half of the twentieth century!

On request of the US Senate, Simon Kuznets ([9]) measured the impact of the financial crisis of 1929 on the US Economy (Kuznets (1934)). He made estimates of national income for the period 1929–1932 broken down by type of income (labour income, property income and

([8]) Considering that much guesswork was included, he argued that the amount would certainly be between 1.6 billion pound and 3 billion pound.

([9]) About Kuznets and his work, see e.g. Carson (1975) and Lundberg (1984).

entrepreneurial income) and by industry (e.g. agriculture, mining, manufacturing, etc.). These estimates showed a contraction of 40–50 %. Taking into account the decline in wholesale prices, this indicated that the volume of national income was reduced by 30–40 % in the period 1929–1932. He reconstructed national income and product accounts for the USA first back to 1919 and eventually back to 1869. Such impressive measurement exercises were the input for investigating business cycles and long term economic growth. For example, what was the role of the various industries, what was the role of technology and innovations, what is the relationship between economic growth and inequality (the U-shaped Kuznets-curve) or between economic growth and urbanization?

Table 2: New applications and uses of the national accounts during 1930–1950

New application or use	Who?
Purchasing power parity: international comparison of real income	Clark (1937 and 1940)
Systematic analysis of long term growth by using national accounts time series	Kuznets (1934)
Input-output analysis	Leontief (1936)
Econometric modelling of national economies	Frisch (1933), Tinbergen (1936)
Keynesian revolution and the birth of macro-economics	Keynes (1936)
Analysing public finance in a macro-economic framework	Meade and Stone (1941)
Monetary policy linked to national income instead of gold & inflationary gap analysis	Keynes (1940)
Analysing balance of payments in a macro-economic framework	Meade (1951)

Leontief (see e.g. Leontief (1936)) developed input-output analysis, estimated detailed input-output tables for the USA and applied this new tool to all kinds of problems. For example, he estimated the resource costs of conversion to peace time production in 1945 and by calculating the relative factor intensities of imports and exports he discovered the Leontief paradox: 'why are American exports labour intensive, while American imports are capital intensive?'

Tinbergen and Frisch [10] pioneered in econometric model building covering the whole national economy (Morgan (1990)). These models need national accounts data as input and can be used to forecast the national economy or to analyse the economy or alternative policy proposals. As a consequence, the development of macro-econometric model building was a strong stimulus for the development of national accounts statistics. For example, in 1936, Tinbergen constructed the first econometric model of the business cycle covering the whole economy (Tinbergen (1936)) to assess the best policy measure to improve the business cycle: devaluation, public works or reducing wages. His conclusion was that devaluation was the best policy. This was very much against the politics of the time, i.e. not only against the official policy of the government but also against the opinion of the socialist party. Nevertheless, in September 1936, the guilder was devalued without the disastrous consequences that had been predicted by many. This contributed enormously to the reputation of Tinbergen and his new method. Tinbergen's revolutionary model of 1936 was not based on a national accounting scheme, e.g. concepts like national income and final consumption by the government were absent. Nevertheless, it gave the development of national accounting in the Netherlands a head start (Bos (2008a)). In order to provide a better empirical grounding to the econometric model, new and longer time series were needed and the quality of existing estimates was to be improved

[10] On the impact of Frisch on national accounting in Scandinavian countries, see Aukrust (1994).

Keynes published his 'General Theory' in 1936 (Keynes (1936)). This launched the Keynesian revolution and gave birth to macroeconomics. This revolution in economic theory had an enormous impact on national accounting. The Keynesian type of analysis established a direct link between economic theory and national accounting as both came to use the same macro-economic identities. The Keynesian type of analysis also threw a new light on the role of the government: a new responsibility for stabilising the economy was added. Accounting for this role of the government became necessary for economic policy analysis. This induced the introduction of accounting per sector, in particular the introduction of a government sector. As a consequence of the Keynesian revolution, the importance of national accounting figures for economic theory and economic policy increased and was more widely recognised.

On Keynes' instigation, in 1941 Stone and Meade prepared UK estimates on national income and expenditure (Meade and Stone (1941)). These estimates were used to present government expenditure and revenue as part of a system of balanced tables describing the whole national economy. In this way, they became a tool in planning the British war economy (Stone (1951), p. 84; Patinkin (1976), p. 1109). It also implied a revolution in the practice of government budgeting: since then, in most countries the government budget is presented in a macro-economic framework with explicit statistics or forecasts about economic growth and inflation ([11]).

In the thirties, countries left the gold standard. As a consequence, monetary policy needed a new anchor and was linked to national income. In order to avoid excessive inflation, the supply of money should grow in line with the nominal growth of domestic product corrected for changes in the velocity of circulation ()[12]. During and directly after the Second World War, several countries (e.g. USA, UK, France and the Netherlands) used the national accounts to estimate the inflation gap and investigate the size and effectiveness of various policy measures to be taken to avoid massive inflation, e.g. restrictions on the use of bank deposits or a big incidental capital tax. These studies include also Keynes' famous report 'How to pay for the war'? (Keynes (1940)) ([13]).

In 1950, Meade wrote his book 'The balance of payments' (see Meade (1951)). In the first three chapters, the basic concepts of the balance of payments are clarified and linked to those in the national accounts. The importance of the various types of consistency is stressed. For example the consistency between net exports and a country's indebtedness to foreigners, between the balance of payments of various countries or between the balance of payments and the national accounts, e.g. private and public expenditure and saving, monetary circulation and inflation. In the remainder of the book, he used these accounting identities to analyse the consequences of various types of monetary and fiscal policy on the balance of payments. His exposition of the balance of payments and the national accounts is still the basis of modern text book expositions on international economics (see e.g. Krugman and Obstfield, 2011).

Most of these new uses also reinforced each other. These uses were also closely linked to the economic circumstances: the economic crisis of the thirties, the Second World War and the need for recovery afterwards stimulated an active role of the government. National accounts statistics turned out to be very useful in such circumstances for analyzing, monitoring, forecasting, discussing and planning the national economy.

([11]) However, in the official budget of the UK this relationship was only introduced in 1967 (see Jones (2000)).

([12]) This follows from Fisher's equation of exchange MV =PT in which M is money supply, V is velocity of circulation (the number of times money changes hands), P is average price level and T is the volume of transactions of goods and services (see Fisher (1911)).

([13]) This report also stimulated the work by Meade and Stone (1941), see above.

3.2. Major changes in concepts and compilation methods

These decades were also a revolution in terms of the development of national accounting concepts and compilation methods. Both revolutions were not independent: they reinforced each other and often also the same persons were involved.

The works by Clark and Kuznets consisted of profound and detailed estimates that were accompanied by elaborate motivations of the concepts and statistical methods used. In the first chapter of 'National Income and Outlay' (Clark (1937)), Clark expounds the purposes of national income measurement and its basic concepts. Examples of the latter are his discussion of the inclusion of the services of owner-occupied dwellings, the exclusion of the services of consumer durables, the exclusion of holding gains and losses and a possible 'deduction for any demonstrable exhaustion of natural resources' (Clark (1937), p. 9). Kuznets had also a famous discussion with Hicks in *Economica* on subjects like the relation between changes in national income and welfare, the valuation of government output and the concept of intermediate and final product (Hicks (1940 and 1948), Kuznets (1948b)).

Commodity-flow accounting can be regarded as the statistical counterpart of input-output analysis, as commodity-flow accounts are a type of input-output table. Commodity-flow accounting was more or less simultaneously developed in Sweden (Lindahl *et al* (1937)), Denmark (Kampmann (1942)) and the United States (Kuznets (1938)).

Clark and Kuznets did not play a role in developing national accounting systems, i.e. a system in which sectors as well as accounts are used in presenting data. Kuznets saw it even as a 'dubious addition to the theoretical equipment' (Kuznets (1948a), p. 154)! The development of national accounting systems occurred simultaneously in Britain, the Netherlands and the Scandinavian countries.

In 1939, the League of Nations had requested a report with guidelines in order to improve international comparability of national accounting figures, but the war delayed the progress on the report. In September 1944, representatives of the UK, the USA and Canada met in order 'to exchange views ... and, if possible, to bring about uniformity in terminology and the treatment of controversial items' (Denison (1947), p. 3). As a result of this meeting, the national accounts of the United States and Canada were revised, which made them more compatible with the Stone/Meade proposals of 1941 and the British national accounts (see Carson (1975), p. 177). Immediately after the war, in December 1945, consultations on the United Nations report were resumed. This time also representatives from countries occupied during the war by Germany, like the Netherlands and Norway, could be present. The report was published in 1947 by the UN (UN (1947)) and consisted mainly of an appendix by Stone (Stone (1947)).

This appendix can be regarded as the first fully worked out and detailed national accounting system (see Aukrust (1986) and Carson (1975), p. 178). Furthermore, the report was also pathbreaking in that it contained for the first time international recommendations on national accounting. However, the report should not be regarded as the first official guidelines of the United Nations: it was not approved by the statistical commission of the United Nations as official guideline, but was referred to as a useful technical report; it was also not used as a guideline for submitting data to the United Nations.

4. The era of the international guidelines (1950– ...)

4.1. The successive guidelines (14)

On request of the Organisation for European Economic Co-operation(OEEC), in 1951 a guideline on national accounting was written which was to be used in planning the Marshall-aid. This guideline and its two immediate successors (OEEC (1952) and UN (1953)) can be regarded as the first generation of official international standards. In contrast to the 1947 UN report, rather simple accounting systems were aimed at. In fact, a systems approach was nearly absent as only some aggregates and their composing parts are to be compiled; the financial flows in the national economy are even nearly fully ignored.

Table 3: The successive guidelines on national accounting

Year	Event
1947	Technical report by the UN containing recommendations; including the famous annex by Stone: the first detailed and fully worked national accounting system
1951–1953	First generation of international guidelines: OEEC guidelines of 1951 and 1952; UN guideline of 1953 (SNA53); very simple tables and accounts
1968–1970	Second generation of international guidelines: UN guidelines of 1968 (SNA68), the European guidelines of 1970 (ESA70) and the Material Product System of 1969 (MPS69) for communist countries
1993–1995	Third generation of international guidelines: joint guidelines of 1993 by the international organizations (SNA93 by UN, IMF, World Bank, OECD and EC) and the European guidelines of 1995 (ESA95)
2008–2010	Fourth generation of international guidelines: updates of the joint and European guidelines (SNA 2008 and ESA 2010)

A second generation of official guidelines was issued at the end of the sixties and beginning of the seventies. It consisted of two guidelines by the UN: the Material Product System (MPS69) used by communist countries and the System of National Accounts of 1968 (SNA68) used by the rest of the World. For the special purposes of the European Communities, also separate guidelines were issued for EC-countries, i.e. the European System of economic Accounts of 1970 (ESA70). These are (almost) fully aligned with the SNA, but contain specific EU-detail. The SNA68 and ESA70 followed mainly the trails set by their predecessors but greatly expanded the accounting system, e.g. by including also input-output tables and constant prices. The MPS69 differed fundamentally from all the other guidelines by its anachronistic concept of production that focuses on measuring only material production (15). Nevertheless, with respect to the other concepts and its scope, the MPS69 was sometimes even more advanced than the SNA68 and ESA70 (see Arvay (1994)). The major example is the notion of 'total consumption of the population'. In the SNA93 and ESA95, this is known as household actual individual final consumption, i.e. the aggregate of household final consumption expenditure plus social

(14) A detailed comparison of the successive UN-guidelines (excluding the most recent one) can be found in Bos (1993) and Bos (1994).

(15) Studenski (1958), p. 22, argues that the production boundary in MPS69 is based on the ideas of Marx, and more in particular on a mistaken interpretation of it. However, also Adam Smith advocated a material production concept (see Section 2). So, it could be argued that the communist guidelines were not actually based on the ideas of Marx, but on those of the intellectual father of capitalism, i.e. Adam Smith!

transfers in kind government and non-profit institutions, e.g. with respect to education and health care.

In 1985 Richard Stone was awarded the Nobel prize in economics for his contributions to the national accounts, in particular for his leading role in the development of the first and second generation of international guidelines.

Table 4: Major changes in the scope of the successive universal guidelines

Scope	Scope
SNA53	Simple set of tables and accounts in current prices
SNA68	Extended accounting system, including input-output tables, general principles on prices and volumes and financial accounts
SNA93	Inclusion of balance sheets, employment and purchasing power parities. More detailed accounting structure (more accounts, more sub-sectors and detailed supply and use tables). Separate chapters on satellite accounts and flexible adjustments for national circumstances. Detailed discussion of general principles on prices and volumes (e.g. chaining and index formulae)
SNA 2008	More detailed discussions of many topics, e.g. government accounts, the informal sector and capital services (important for productivity measurement). But no detailed discussion of price and volumes by industry/product and no separate chapters on quarterly national accounts and regional accounts (unlike ESA95 and ESA 2010)

In the mid nineties, a third generation was issued, the SNA93 and ESA95. The SNA93 and ESA95 have again greatly expanded the scope of the international guidelines by including e.g. balance sheets, purchasing power parities, satellite accounts and Social Accounting Matrices. A revised version of the MPS69 was not necessary due to the collapse of communism in Eastern-Europe. All countries in transition in Eastern Europe implemented the SNA93 and, for those who applied for membership of the EU, the ESA95.

In recent years, a fourth generation was published, the SNA 2008 and ESA 2010. Basic concepts were hardly changed ([16]). Much attention was paid to measuring the growing complexity of economic reality, e.g. all kinds of new financial instruments, mixes of public and private insurance and an even more global production process. By adding several new chapters, e.g. on government accounts, links to monetary and financial statistics, rest of the world account, quarterly accounts, satellite accounts, European accounts and regional accounts, the focus has been shifted substantially toward a much better link with the major uses of the national accounts. The old guidelines focused on merely presenting a basic national accounts framework in terms of sector accounts and input-output tables.

Next to the official guidelines on national accounts, over the years many other reports, handbooks and guidelines have been published on specific parts of the national accounts, for example on price and volume measurement, seasonal adjustment, environmental-economic accounting and government debt and deficit. A very welcome novelty is that also the users' perspective (how to use the national accounts for analysis?) has begun to be addressed, e.g. by the EU-KLEMS-project for detailed and internationally comparable productivity data and analysis by industry (Jäger (2016)), by the report on national accounts and policy analysis (UN (2002)) and by an introductory book on national accounting by Lequiller and Blades (2014).

([16]) There were some specific changes, e.g. including also R&D as capital formation.

By their rapid expansion of scope since the guidelines of the fifties, the most recent set of guidelines have incorporated most of the major innovations in national accounting since the Second World War. If we disregard the MPS69, basic concepts in the guidelines have shown a remarkable consistency, e.g. excluding the services of unpaid household services. Some of the criticism on unchanged basic concepts has been met by introducing satellite accounts, e.g. on the link between environmental indicators and national accounts statistics. Nevertheless, some important changes in concept have also occurred, e.g. the introduction of chain indices or the expansion of the concept of capital formation to include software, R&D and mineral exploration.

For decades, the international guidelines on national accounts statistics were partly inconsistent with the international guidelines on three specific types of macro-economic statistics: balance of payments, government finance statistics and statistics on employment and population. However, nearly all these inconsistencies have now been resolved.

Also the links with international statistics on specific policy areas, like health care, social security and environment, have been improved. International guidelines on these specific statistics explicitly refer to the national accounts concepts and classifications, discuss similarities and differences and propose satellite accounts to bridge the gap between the national accounts and these specific policy areas.

4.2. The role of the international guidelines

International guidelines have been influential for many reasons. Firstly, the leading international experts of the profession have developed the systems in the international guidelines. They are therefore relatively well thought out and it is costly, time consuming and not easy to invent an alternative system. Secondly, by keeping in line with the international guidelines, national figures can be compared with figures from other countries. This is important, as international comparison is a major use of national accounting figures. Thirdly, in many countries, the national accounts have been set up by or improved with help from the international organisations issuing the guidelines (UN, OECD, EU) or with help from countries advanced in national accounting (Sweden, France). In the latter case, following the international guidelines is usually stimulated to the extent that the helping countries follow them. Fourthly, we mention that all countries are obliged to compile some figures on the basis of the international concepts. In the EU, due to some important administrative uses, the guidelines are even legally binding; the same applies to the statistical programme linked to these guidelines. The fifth reason is that the data submitted to the international organizations play a central role in international policy discussions and decision-making, e.g. about accession to the European Union, granting loans or paying contributions to the international organizations. To link national discussion and decision-making to this international context, the international concepts have to be used for national purposes as well.

The international guidelines are very successful in standardising the concepts and classifications used in compiling national accounts figures. The achievement of the guidelines is that all over the world official figures came to be based on uniform notions of the production boundary, asset boundary, the distinction between intermediate and final consumption, etc. This is evidenced by some of the earlier country practices. They all differed fundamentally from the basic concepts in the successive guidelines.

In Sweden in 1937, Lindahl published two alternative estimates of national product, one including the services of unpaid household services and one excluding them (Lindahl *et al*

(1937)). However preference was given to the latter. In Norway in 1946, the value of unpaid household services was included in output and national product. However, since 1951 they are excluded (see Aukrust, 1994). In some official Scandinavian studies (1937, 1951, 1953), the services of consumer durables like cars were included in output (see Aukrust (1994)). In France, until 1975 when the ESA70 was implemented in the French national accounts, the value of the output by banks, insurance companies and general government was not included in output and domestic product (1971 base, see Demotes-Mainard and Bournay (1994)). Due to the influence of the international guidelines, country practice to include in the national accounts' estimates of unpaid household services and the services of consumer durables was gradually extinguished.

However, as the French case shows, some drastic differences between country practice and the international guidelines existed unto the seventies. Furthermore, if we look at the changes in the international concepts and classifications, some important differences continued to exist for decades, as country practice often stuck to old concepts and classifications or continued with typical national versions of the national accounts. For example, in the USA, the national income and product accounts and flow-of-funds accounts did not comply fully with the previous international guidelines (SNA1993), e.g. expenditure on military weapons systems was treated as gross fixed capital formation. In 2006, the USA's Bureau of Economic Analysis (BEA) launched a set of 'Integrated Macroeconomic Accounts' based on the SNA 2008 and integrating the information from the national income and product accounts and the flow-of funds accounts and adding various other pieces of information. Even in this set of accounts, there are several clear differences with the SNA 2008, like government enterprises included in the government sector, but they do generally not affect GDP (see Yamashita (2013) and McCulla *et al* (2015)) ([17]). Finally, European unification has launched in European countries an enormous effort to increase comparability of their national accounts statistics and their compliance to European guidelines (see below). Nevertheless, even in Europe, some clear differences in country practices continue to exists, e.g. with respect to price- and volume changes of health care and education services.

4.3. Welfare and the guidelines

The successive guidelines all agreed that national accounts should not aim at measuring welfare, but focus on serving as a practical tool for macro-economic policy issues (e.g. public finance, balance of payments, distribution of income in terms of profits and wages). In the late sixties and the beginning of the seventies, national income was frequently criticised for not being a welfare measure (e.g. Mishan (1969)). However, the authors of the international guidelines did not intend to provide a measure of economic welfare. For example, Jaszi even regards as one of his principal contributions to have resisted successfully to 'the will-o'-the-wisp of forging national output into a measure of economic welfare. I was a minority of one in a company that included such mental giants as Simon Kuznets and John Hicks, and at one point I had to defy a forceful Secretary of Commerce who had instructed the BEA to prepare a measure of welfare' (Jaszi (1986), p. 411). According to Okun, '[the] beauty of ... present practice is that no sensible person could seriously mistake the GNP for [a measure of total social welfare]' (Okun (1971), p. 133).

In 1972, Nordhaus and Tobin (1972) illustrated in an impressive way what accounting aimed at measuring welfare would imply. They calculated a Measure of Economic Welfare (MEW) by modifying traditional national income figures in several respects. For example, they deducted an estimated value of the disamenities of urbanisation and they added tentative estimates for

([17]) However, in transmitting data on the USA to the international organization, the BEA makes corrections in order to comply to the international guidelines.

the value of unpaid household services. Since then, many measures similar to MEW have been calculated (see Eisner (1988)). Frequently, these measures were presented as part of extended or total accounts. Measuring the contribution of economic activity to welfare is only one of the reasons for drawing up such accounts. Some other motives are to obtain: 'more inclusive and relevant measures of capital formation and other factors in economic growth, and better and/or additional data to fit concepts of consumption, investment, and production relevant to economic theory and structural econometric relations' (Eisner (1988), p. 1612).

The increased use of social indicators like the Human Development Index (UNDP (1991)) is a somewhat related development. In these social indicators, national income (per capita) is only one of the variables, other variables being e.g. infant mortality, life expectancy and adult literacy rates. In contrast to measures like MEW and National Income, social indicators are not measures in money terms; they serve solely as indexes.

In 2009, on request of the French government prominent international experts investigated how to measure economic and social progress (Stiglitz *et al* (2009)). They provide two types of recommendations: those that stay within the existing framework of national accounts and those that clearly go beyond that. Major recommendations within the existing framework of the national accounts are:

- Emphasize well-established indicators other than GDP in the national accounts. For example, a) Net Domestic product adjusted to record natural resource depletion as depreciation, b) net national disposable income, which accounts for income received from and paid abroad (as business income and household income to and from abroad can be very important) and c) real disposable household income, which is much more relevant for living standards and well-being. Compiling national balance sheets for financial and non-financial assets, including natural resources, is also essential for measuring current well-being and sustainability.
- Improve the empirical measurement of key production activities, in particular the provision of health care and education services and changes in the quality and type of services, e.g. due to digitalization. Current national accounts measurement practice focuses on measuring changes in inputs or a very rough indicator of output of health care and education. Changes in the quality of goods and services is mostly ignored, e.g. the important role of new medicine and medical treatments. This reflects that this measurement task is challenging as the link between health and education services and health and education outcomes is complicated. For example, people's lifestyles will also affect health outcomes, and the time parents spend with their children will affect exam scores.
- A household satellite account should become an major part of the regular statistical programme of countries and international organisations (18). Such a satellite account is important for showing extended measures of income, consumption and wealth of households and their distribution. Household income measures should be adjusted for government services in kind, like health and education services and subsidized housing, museums and parks. When there are large changes in inequality, GDP or any other aggregate computed per capita does not provide an accurate assessment of the situation in which most people find themselves. The household satellite accounts may also serve to incorporate in household income the implicit income derived from non-market activities (e.g. cleaning, cooking and care for children, sick people and the elderly), to treat time and money spent

(18) In the international guidelines SNA 2008 and ESA 2010, satellite accounts are discussed in separate chapters, but they are not part of the central accounting framework, like the sector accounts and the supply and use tables, and they are also absent or of secondary importance in the regular statistical programme of most countries and international oragnisations.

on transport services to work as intermediate consumption and to show major changes in leisure time.

Major recommendations to better measure economic and social progress that clearly go beyond the national accounts framework are:

- Objective and subjective information should be provided about the various dimensions of well-being, including not only material living standards, health and education, but also political voice and governance, social relationships, environment and insecurity.
- A dashboard of indicators should be used to measure sustainability. The economic component of sustainability should be captured by monetary aggregation on items for which reasonable valuation techniques exist, such as physical capital, human capital and certain natural resources. For the environmental aspects of sustainability, there is a need for a clear indicator of our proximity to dangerous levels of environmental damage, like climate change or the depletion of fishing stocks.

This relatively recent report was a game-changer and has already exerted an enormous influence on public debate about measuring welfare, quality of life and sustainable development. It has also changed measurement and publication practice by international organisations, like United Nations, OECD and European Commission. For example, since 2011, OECD is publishing regularly for all OECD countries multi-dimensional statistical indicators on well-being (see OECD (2015)). Eleven different dimensions of well-being are distinguished, e.g., income and wealth, jobs and earning, health status, education and skills, social connections, environment quality and subjective well-being; for each dimension, several statistical indicators are presented. Recently, also historical time series on the development of well-being in OECD countries since 1820 have been published. This includes not only GDP per capita, but also indicators like literacy, life expectancy, people's height, crime rates (homicide) and income inequality (see van Zanden *et al* (2014)). Serious efforts have also been undertaken to start compiling national accounts statistics on human capital (see UNECE (2016)) and on the distribution of household income, consumption and saving (see Zwijnenburg *et al* (2017)).

4.4. Differences in scope of national accounts all over the world

Under the influence of the international guidelines and the international organisations, national accounts statistics are now available for practically all countries. For most countries, they have also gradually extended in scope and detail. Nevertheless, still substantial differences in scope, detail, quality and frequency exist between the national accounts statistics published by countries.

For example, since the fifties countries like Norway, Denmark, the Netherlands and France [19] publish annual input-output tables. Input-output tables were incorporated for the first time in the international guidelines in the SNA68 and ESA70. Decades later, only a few more countries published annual input-output tables. For a somewhat larger group incidental but usually rather outdated input-output tables existed. Nowadays, annual supply and use tables are part of the standard delivery programme for OECD and especially EU countries, but in most other countries such data are still not available. A similar story can be told for the detailed sector accounts proposed by the SNA68, the ESA70 and the most recent international guidelines. Even

(19) A very interesting survey of the post-war developments in France can be found in Demotes-Mainard and Bournay (1994). A more extended overview is given by Vanoli (2005).

now, a great majority of the countries in the world only apply rather simple accounting systems of the SNA53-style.

Substantial differences in country practices exist also with respect to specific national accounts statistics, like regional accounts, quarterly accounts, satellites and balance sheets. In some countries, all of them are regularly published (e.g. in the Netherlands, France and Canada). In a somewhat larger group, some of them are regularly published, while in many countries hardly any data are regularly published on any of these topics.

Our remarks with respect to the input-output tables, detailed sector accounts, balance sheets and satellite accounts reveal that most of the international guidelines have been much more ambitious and encompassing than the national accounting practice of their time. This partly reflects their role as a pedagogical device and innovative instrument ([20]).

However, in the case of the USA it may also partly reflect a fundamentally different view on the role and design of the national accounts. In line with USA national accounts practice up to the 1970s and 1980s ([21]), two quotes from American economists stress the importance of a relatively simple set of accounts and criticize the complexity and cost-inefficiency of universal accounting framework:

> *'These … accounts are designed to answer 'Who does What by means of What for What purpose with Whom in exchange for What with What changes in stocks?' Given this level of complexity, there is a distinct danger that when the revised SNA is actually put in place, it, like the Hubble telescope, may not be successful in bringing into focus a clear view of what it was designed to examine. Only professional national accountants will be able to fathom the national accounts. Furthermore, the establishment of such an elaborate system as the standard to be adopted by national and international statistical offices may result in the SNA becoming a statistical behemoth independent of its creators and with an illogic of its own-not unlike a Frankenstein monster. One of the major virtues of national accounting systems used by many countries is that they do provide a relatively simple macroeconomic overview of the economic system'* (Ruggles (1990), p. 419).

> *'For every series of any real interest that is developed, at least a dozen series of trivial or no value must be estimated to fill out the 'accounts'. Because most of the series called for are of no appreciable interest, existing systems of data collection do not provide the information required by the new SNA (F.B.: SNA68); either collection of trivial data would be required or the number would have to be imaginary. The new SNA has another weakness: it is so complicated that not even serious and expert users of national income and product data (and few producers for that matter) can be expected to understand it or the meaning of the numbers it is to contain. A very simple set of accounts…, supplemented by supporting tables to provide analytically interesting detail and alternative breakdowns, is a far better approach, in my opinion'* (Denison (1971), p. 38).

([20]) It also reflects the totally different amounts of resources in countries available for statistics in general, and for national accounts in particular.

([21]) This changed in the 1990s: in 1993, the US Bureau of Economic Analysis (BEA) announced that it 'plans to move toward the SNA.' (BEA, 1993) Since 1993, all of BEA's major revisions have incorporated features of and improved consistency with the latest versions of the SNA. Also terminology is adjusted in line with the successive versions SNA, e.g. implementing SNA 2008 involved also switching from the term 'flow-of-funds-accounts' to 'financial accounts'.

4.5. European unification

The ongoing European unification is revolutionising European national accounting ([22]). National accounts figures, like national income, government deficit and GDP volume growth have been selected to play a special role in monitoring and managing the European unification (see also Table 5). This role in European policy has also drastically increased the importance of national accounts statistics in national policy. In discussing and deciding on the national budget, national accounts statistics on the government deficit have become the central figures in all EU countries. They have often taken over this role from specifically nationally defined concepts (see text box). The EU member qtates and the European Commission have been aware that the national accounts statistics had to be strengthened for such usage. They therefore launched an ambitious programme for improving the quality and comparability of national accounts figures and for drastically extending the set of national accounts statistics that are available for all EU member states.

Table 5: Use of national accounts statistics for European policy

Policy area	Which national accounts statistic?
Monetary policy and public finance	Government deficit and debt as a percentage of GDP
	Financial accounts showing e.g. the size of new mortgages and loans by corporations
Productivity and growth policy	Economic growth, expenditure on R&D as percentage of GDP, EU KLEMS
Social policy	Social protection statistics closely linked to national accounts concepts
Regional policy	Regional product per capita is yardstick for granting regional funds
Agricultural policy	Agricultural accounts showing also the development of farmer income
Development aid	Low domestic product per capita is a yardstick for granting aid
Defense policy	Expenditure on defense as a percentage of GDP
Maximum total expenditure by EU	Percentage of GNI (about 1 %)
EU budget contribution	GNI

From a universal perspective, these European developments are in two respects revolutionary. Firstly, the development of jurisprudence on the interpretation and application of the international guidelines is a totally new development for the national accounts. Secondly, the European experience gives a concrete example of how the quality and comparability of national accounts statistics can be improved, e.g. by auditing ([23]), issuing guidelines on inputs, official requirements for compiling and publishing an extensive set of national accounts statistics and by more specific guidelines. Some of this extra specification is also very useful for countries outside Europe, e.g. with respect to the delimitation of the government sector, the definition of price and volume changes for various types of industries and product, the definition and measurement of services of owner-occupied dwellings, compiling regional accounts and compiling quarterly accounts. As a consequence, the European guidelines on national accounting could be regarded as a useful supplement of the universal guidelines, even for non-European countries.

([22]) For a more extended discussion see Bos (2009a), pp. 54–58 and Bos (2011).
([23]) For some EU-countries, this resulted in upwards revisions of official domestic product and national income statistics of 10 to 20 %!

Box 1: European norms for actual deficit and debt

The treaty of Maastricht in 1992 implied that monetary policy became a responsibility of the European central bank and that national fiscal policy should comply with the European norms of actual deficit and debt. Deficits should not exceed 3 % of GDP and debt must be below 60 % of GDP or be declining towards the 60 % norm at a satisfactory rate. According to the Stability and Growth Pact, the budget balance should be close to balance or in surplus in the long run.

As a consequence, the national concepts on public finance were replaced by the new European concepts based on the national accounts. This had several practical implications:

- A change in concepts. For example, according to the national account's concept of budget balance, revenue and expenditure like taxes and interest payments should be recorded on a transactions basis. Financial transactions like loans and the sale of equity are irrelevant and the government includes not only the state and social security funds, but also municipalities, provinces and many other non-market units mainly financed and controlled by the government.
- The concepts cannot be changed anymore over time by the government.
- A link to national accounts statistics and therefore a new role for national statistical offices and a more limited role of Ministries of Finance. The official figures reported to the European Commission and European Central Bank should be consistent with those reported by the national statistical office. In the end therefore, statistical offices are responsible for translating the general European concepts into operational concepts for their country and to make the best estimates for these operational concepts.

The transition towards European concepts does not imply that bookkeeping and bookkeeping tricks have become irrelevant. Like all national concepts of taxable income, the European concepts on public finance can affect actual behaviour (e.g. stimulate leasing of capital goods to reduce the deficit or stimulate the sale of public equity in order to reduce public debt) and the specific institutional arrangements chosen ([a]). Furthermore, they are not optimal from an economic-theoretic point of view (e.g. not forward looking and ignores financial assets and implicit liabilities like future pensions) and may not well take account of the current economic situation in a country. They are the outcome of political negotiations in view of the circumstances in Europe in 1992 and the purposes of the criteria, i.e. to provide signals that countries are willing and able to live with the discipline required by EMU (see Bovenberg and De Jong (1996), p. 18).

Following the seminal work by Auerbach, Gokhale and Kotlikoff (1991), in 1997, the Netherlands Bureau for Economic Policy Analysis started to calculate generational accounts for the Netherlands (see Bos and Teulings (2013b)). All kinds of national accounts information on government revenue, expenditure and debt and economic growth serve as input. These calculations demonstrated that policy arrangements (taxes, public expenditure on social security, education and health care, subsidies, etc.) in the Netherlands were not sustainable due to the costs of ageing and the exhaustion of the natural gas resources. They also showed the consequences of alternative policy measures to reduce this gap for different generations. This forward looking approach of generational accounting is the new paradigm for Dutch public finance. Similar calculations are now made regularly for all EU member states and play also a role in assessing European fiscal policy.

([a]) On the merits and limitations of the EMU-targets of government deficit and debt, see also Bos (2009a, chapter 8) and Bos (2008b).

4.6. The impact of the financial crisis of 2007–2009

The financial and economic crisis of 2007–2009 has had a major impact on the demand for national accounts statistics. It has stressed the importance of much better and timelier recording of financial assets and liabilities, their international dimension and their relationship to the real economy. The suppliers of national accounts statistics have responded in various ways. For example, in the USA, integrated macro-economic accounts were developed (see Yamashita, 2013), e.g. showing the importance of mortgage debt and holding gains and losses in household wealth accumulation. In the European Union, much effort was put into better recording public interventions in the financial system and their interaction with the 'member states' public finance. International organizations, like UN, IMF, World Bank, European Commission and OECD, cooperated to identify and remedy gaps in the international supply of national statistics. More comprehensive and timely data were needed. In particular improvement was recommended of the international reporting about the financial health of financial institutions, about cross-border financial relationships (e.g. about the importance of Greek government debt for French and German banks and Dutch pension funds), about government finance statistics, balance sheets of households, government and non-financial institutions and about real estate prices ([24]). Also more coordination in the publication and communication of official statistics was recommended.

The financial crisis has also raised serious doubts about the measurement of value added and real output growth of banks. Their definitions in the international guidelines may need to be remedied, e.g. in order to better account for taking more or less risks and to measure bank service output on transaction counts instead of balances of loans and deposits (see Inklaar and Wang (2013) and van de Ven (2015)).

4.7. Long run perspectives on the economy

As already shown by Kuznets, compiling long run national accounts statistics can be very important to better understand fundamental economic issues. The much more recent work by Maddison, Gordon, Reinhart and Rogoff and Piketty can illustrate this.

Maddison (2003, 2005 and 2007) collected and compiled real GDP per capita figures for countries all over the world during the last millennium. This comprehensive data set fully focused on real GDP per capita and did not show any other national accounts statistics, e.g. no breakdown by industry is presented. This data set has been used intensively by many other economists for their research to better understand economic growth and why some countries are rich and other poor (see e.g. Acemoglu and Robinson (2012)).

Gordon (2012) investigates the importance of innovation for economic growth. Its point of departure is growth in real GDP per capita in the frontier country since 1300, the UK until 1906 and the USA afterwards. Growth in this frontier gradually accelerated after 1750, reached a peak in the middle of the 20th century, and has been slowing down since. This study suggests that the growth effects of the information revolution are not of the same order of importance as those of previous technological revolutions and are, in any event, playing out. Future growth per capita could fall below 0.5 percent per year for an extended period of decades. This study plays a major role in international discussions about our future living standards. However, its analysis and conclusion depend to great extent on how well new goods and services, e.g. computers, internet and mobile phones) are measured in real GDP figures. This problem was

([24]) This is the G20 Data Gaps initiative, a set of 20 recommendations on improving economic and financial statistics.

already raised by Adam Smith more than two centuries ago and, despite a lot of research effort, has still not been resolved (see Hulten (2015)).

Up to the recent financial crisis, macro-economic forecasts on the developed countries mostly ignored the possibility of a financial crisis (see Bos and Teulings (2013)). This reflects that these forecasts were based on national accounts time series spanning some decades. Historical and long run time series on economic growth in the Netherlands and other developed countries did not seem relevant for such forecasts. As a consequence, the recent financial crisis came as a total surprise. Following (Taleb (2007)), this was regarded an example of a black swan: a very unlikely event with a huge impact. However, by compiling and collecting data on financial crises all over the world during the past two centuries (e.g. data on economic growth, government debt, household debt and capital inflows), Reinhart and Rogoff (2011) showed that from a long run perspective such black swans were much less rare than was assumed. Furthermore, they showed that sound economic policy and institutions can play an important role in avoiding financial crisis or mitigating its most negative effects.

Since the Second World War, despite the drastic expansion of the scope of national accounts statistics, inequality of income and wealth was an issue absent in national accounts statistics all over the world. Also in the international guidelines, the issue was fully ignored. However, only some years ago, Piketty (2014) showed that this is an important issue and that it can be well addressed by compiling national accounts statistics on the size and economic return on private and public wealth and on capital income and labour income as a percentage of national income. In order to reveal the amount of inequality, capital income and labour income are broken down by level of income and wealth, e.g. the poorest 50 %, the richest 10 %, the middle class of 40 % and the very rich 1 %. These statistics show that 50 % of national income in the USA in 2010 is for the richest 10 %, which is similar to the situation in France, Britain and Germany in 1910 ([25]). Starting from his new set of statistics, he advocates more progressive taxation of income and wealth.

5. Conclusions and three suggestions for the future

Since the seventeenth century, estimates of national income and GDP and other national account statistics, have mainly been developed for fiscal and monetary policy purposes, for measuring progress in material well-being and for showing their relationship with the supply and use of goods and services in a market economy. National accounts statistics were not developed for measuring welfare, quality of life, inequality or sustainable development. This reflects that for such policy issues measurement in monetary terms is less meaningful and more difficult to compile reliably and often requires also totally different data sources.

Over the years, the scope, frequency and timeliness of national accounts statistics has been expanded greatly, recently in response to the financial crisis. But the supply of current national accounts statistics is still far from being capable of serving the many data needs of the future, including those on welfare, quality of life, inequality or sustainable development. Furthermore, major changes, like globalisation ([26]), digitalisation, exhaustion of natural resources and growing

([25]) Piketty stresses that comparisons on the basis of inequality indexes, like the Gini coefficient, do not suffice to understand the development of the distribution of wealth and income over time and across countries.
([26]) See van de Ven (2017).

inequality, challenge the relevance of the current national accounting concepts. How should the data needs of the future and the complicated measurement issues best be met? What can we learn from our overview of national accounting since the seventeenth century? We provide three suggestions [27]:

- More slow statistics for difficult to measure topics.
- Fewer rapid statistics that do not contain much information.
- Better explaining value added, use and misuse of national accounts statistics

5.1. More slow statistics for difficult to measure topics

In order to meet the complicated measurement issues and many data needs of the future, a sizable part of national accounts statistics should be 'slow-statistics', i.e. compiled regularly but not very frequently, e.g. every two or five years. In order to allow for international comparison and to increase public attention for the statistic, the publication of such slow-statistics should be coordinated internationally.

To reduce misinterpretation of GDP-volume growth statistics and to stimulate sustainable growth, such slow-statistics should include well-being accounts, information on inequality of income and wealth and net-net statistics on domestic product and national income, i.e. including a correction for the exhaustion of natural resources. For understanding economic growth, a regional breakdown of statistics on the national economy (regional accounts) is also very important, as economic growth is often location specific and can differ substantially within a nation. Awarding the 2008 Nobel prize in economics to Krugman, in particular for his work on new economic geography (see e.g. Krugman (1991)), reflects the importance of the issue, but regional accounts were hardly mentioned in the SNA 2008 [28]. Another major topic is alternative measurement of price- and volume change, e.g. with respect to the quality and new types of financial, education and health care services and the impact of telecommunication revolution.

5.2. Fewer rapid statistics that do not contain much information

In many developed countries, the frequency and timing of all kinds of national accounts statistics has been increased enormously. This is a major asset. However, there seems to have been some overshooting, as not all these statistics contain really new information. This may be revealed by compiling statistics on the measurement errors in the successive estimates (see Bos (2009b)). The statistical information on measurement errors in national accounts statistics may also be used to inform users about the inaccuracy of the estimates.

5.3. Better explaining value added, use and misuse of national accounting

By providing key-indicators on the structure and development of the national economy, national accounts statistics play an important role in public debate, policy and economic

[27] For more ideas, see Bos (1996), Bos (2009a, chapter 9 and 10), Vanoli (2014) and van de Ven (2017).

[28] Some figures may illustrate the issue. In the USA and Germany the richest state has about twice the per capita income of the poorest state and this ratio is five in China and India (see Sethia (2014)). For international comparison of small countries, comparison with specific regions in large countries may be even more relevant than comparison with the total of large countries. For example, the Netherlands is a small and highly urbanized country with 17 million inhabitants and could be compared to the neighbouring German state North Rhine Westfalia or to US states like New York and Florida.

analysis. A proper use of these statistics requires a good basic understanding of the underlying concepts and the reliability of the data. Without such understanding, public debate, policy and analysis will steer on the wrong compass.

However, there is widespread illiteracy in national accounting, national accounting is generally considered to be one of the most boring topics in economics and knowledge of basic national accounts definitions and discussions seems to be irrelevant for economic researchers (see Bos, 2009a). The new universal guidelines on national accounting (SNA 2008) are so lengthy and voluminous that hardly any national accountant has read them. From a users' perspective, the problem is that most of the text is devoted to explaining the general framework of national accounting and that the links with specific uses for policy and analysis are often very indirect.

How we define and measure the world is how we see the world and that can affect how citizens, business and government will react and what we think that is realistic, important or how fair. Changes in national accounting rules, corporation and income tax rules, rules by bookkeeping and supervisory bodies on banking, pension funds, they matter greatly for what we see and how we act. Explaining the value added of national accounts statistics for different types of uses and their link to administrative concepts of taxation, bookkeeping and supervision should therefore be a major part in economic curricula and in the education of national accountants.

And this needs not to be boring. The financial crisis and new research on long run perspectives on the national economy (e.g. by Gordon, Rogoff and Reinhart and Piketty) have again revealed the merits and limitations of national accounts statistics. Recently, also several popular books debating biases in GDP and their consequences for policy ([29]) have been published.

All this information can be used to discuss the merits of changing its basic concepts, adding new perspectives or of even abolishing national accounts statistics. Would policy and life in development countries or Saudi Arabia have been different when GDP included a correction for the exhaustion of natural resources? Would policy and life in the USA have been different when GDP included a correction for the increase in inequality? How would we have managed the financial crisis in the absence of national accounts statistics? It is also important to make a link between national accounts statistics and ordinary life. For example, what could national accounts statistics tell about the life of Becky of 10 years in the USA versus that of Desta of 10 years in Ethiopia (see Dasgupta (2007)) ([30]). Or: what does a four-fold increase in GDP per capita since 1950 mean in the Netherlands? What does it tell about the life of a female economic journalist now 60 years old and that of her grandmother who died at that age in 1950? (see Bergen (2014)).

Acknowledgements

The author would like to thank the referees for their substantial and very useful suggestions. The views expressed are those of the author and do not necessarily reflect the views of his current or former employer.

([29]) See e.g. Fioramonti (2013), Coyle (2014), Lepenies (2016) and Philipsen (2017).
([30]) A similar approach is taken by Kay (2003) comparing the lives of four families in USA, Sweden, Switzerland, Russia and India.

References

Acemoglu, D. and J.A. Robinson (2012), *Why nations fail — The origins of power, prosperity and poverty*, Crown Business Profile Books, London.

Arvay, J. (1994), 'The Material Product System (MPS): a retrospective', in Kenessey (1994), pp. 109–123.

Aukrust, O. (1986), 'On the occasion of a Nobel prize', *Review of Income and Wealth*, Vol. 32, issue 1, pp. 109–112.

Aukrust, O. (1994), 'The Scandinavian contribution to national accounting', in Kenessey (1994), pp. 16–65.

Auerbach, A.J., J. Gokhale and L.J. Kotlikoff (1991), 'Generational accounts – A meaningful alternative to deficit accounting', *NBER Working Paper* No 3589.

BEA (US Bureau of Economic Analysis) (1993), 'New international guidelines in economic accounting', *Survey of Current Business*, February 1993, pp. 43–44.

Bergen, A. van, (2014), *Gouden Jaren; hoe ons dagelijks leven in een halve eeuw onvoorstelbaar is veranderd*, Atlas Contact Uitgeverij.

Bos, F. (1993), 'Standard national accounting concepts, economic theory and data compilation issues; on constancy and change in the United Nations Manuals on National Accounting (1947, 1953, 1968, 1993)', *National Accounts Occasional Paper,* No 61, Statistics Netherlands, Voorburg.

Bos, F. (1994), 'Constancy and Change in the United Nations Manuals on National Accounting (1947, 1953, 1968, 1993)', in Kenessey (1994), pp. 198–217.

Bos, F. (1996), 'The future of the national accounts', *National Accounts Occasional Paper* No 84, Statistics Netherlands, Voorburg.

Bos, F. (2006), 'The development of the Dutch national accounts as a tool for analysis and policy', *Statistica Neerlandica*, Vol. 60, issue 2, pp. 225–258.

Bos, F. (2008a), 'Uses of National Accounts: History, International Standardization and Applications in the Netherlands', paper presented at the 2008 World Conference on National Accounting and Performance Measures across Nations, 13–17 May, Washington DC, *MPRA paper* No 9387.

Bos, F. (2008b), 'The Dutch fiscal framework; history, current practice and the role of the CPB', *OECD Journal on Budgeting*, Vol. 8, issue 1, pp. 7–48.

Bos, F. (2009a), *The national accounts as a tool for analysis and policy; in view of history, economic theory and data compilation issues*, VDM Verlag Dr. Müller, Saarbrücken.

Bos, F. (2009b), 'The Art and Craft of Compiling National Accounts Statistics and their Implications for Reliability', *Review of Income and Wealth*, Vol. 55, issue 4, pp. 930–958.

Bos, F. (2011), 'Three Centuries of Macro-Economic Statistics,' *MPRA paper* No 35391.

Bos, F. (2013), 'Meaning and measurement of national accounts statistics', paper for the World Economics Association's Online Conference on the Political Economy of Economic Metrics, 28 January-25 February 2013, *MPRA paper* No 44970.

Bos, F. and C. Teulings (2013a), 'Netherlands: Fostering Consensus on Fiscal Policy', in G. Kopits (ed.), 2013, *Restoring Public Debt Sustainability: the Role of Independent Fiscal Institutions*, Oxford University Press, Chapter 7, pp. 121–141.

Bos, F. and C. Teulings (2013b), 'Short and long-term forecasting by the Netherlands Bureau for Economic Policy Analysis (CPB): science, witchcraft or practical tool for policy?', *OECD Journal of Budgeting*, Vol. 13, issue 1, pp. 45–56.

Bovenberg, L.J. and A. de Jong (1996), 'The ins and outs of European monetary union', *CPB Report* 1996/3, pp. 16–21.

Carson, C.S. (1975), 'The history of the United States national income and product accounts: the development of an analytical tool', *Review of Income and Wealth*, Vol. 21, issue 2, pp. 153–181.

Clark, C. (1937), *National income and outlay*, MacMillan and Co, Ltd, London.

Clark, C. (1940), *Conditions of economic progress*, MacMillan and Co., Ltd, London.

Coyle, D. (2014), *GDP: A brief but affectionate history*, Princeton University Press.

Dasgupta, P. (2007), *Economics: a very short introduction*, Oxford University Press.

Demotes-Mainard, M. and J. Bournay (1994), 'Revisions of the French national accounts; past experience as lessons for the future', paper presented at the 23th IARIW Conference, St Andrews, Canada.

Denison, E.F. (1947), 'Report on tripartite discussions of national income measurement', *Studies in Income and Wealth*, Vol. X, 3–22, NBER, New York.

Denison, E.F. (1971), 'U.S. National Income and Product Estimates: Evaluation and Some Specific Suggestions for Improvement', *Survey of Current Business*, 50th Anniversary Issue, pp. 36–44.

Eisner, R. (1988), 'Extended Accounts for National Income and Product', *Journal of Economic Literature*, XXVI (4), pp. 1611–1684.

Eltis, W. (1984), *The classical theory of economic growth*, MacMillan and Co., Ltd, London.

Eurostat (1995), *European System of National and Regional Accounts (ESA95)*, Office for Official Publications of the European Communities, Luxembourg.

Eurostat (2013), *European System of Accounts (ESA 2010)*, Publication Office of the European Union, Luxembourg.

Fisher, I. (1911), *The purchasing power of money and the theory of interest. Its determination and relation to credit, interest and crisis*, McMillan, New York.

Fleetwood, W. (1707), *Chronicon Preciosum: or An account of English money, the price of corn and other commodities for the last 600 years*, 1st edn (anonymous) Harper, London, reprinted in 1745.

Frisch, R. (1933), 'Propagation problems and impulse problems in dynamic economics', in *Economic essays in honour of Gustav Cassel*, Allen & Unwin, London.

Fioramonti, L. (2013), *Gross Domestic Problem: The politics behind the world's most powerful number*, Zed Books.

Gordon, R.J. (2012), 'Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds', *NBER Working Paper* No 18315.

Hicks, J.R. (1940), 'The Valuation of Social Income', *Economica*, Vol. 7, No 26, pp. 105– 124.

Hicks, J.R. (1948), 'The Valuation of Social Income – A Comment on Professor Kuznets' Reflections, *Economica*, Vol. 16, No 59, pp. 163–172.

Hora Siccama, J.H. and B. van Rees (1798), *Algemeen belastingplan*, Archieven B. van Rees, Koninklijke Bibliotheek, The Hague.

Hulten, C.R. (2015), 'Measuring the Economy of the 21st Century', *NBER Reporter*, No 4, pp. 1–7.

Inklaar, R. and J.C. Wang (2013), 'Real output of bank services: what counts is what banks do, not what they own', *Economica*, Vol. 80, issue 317, pp. 96–117.

Jäger, K. (2016), *EU KLEMS Growth and Productivity Accounts 2016 Release*, statistical module.

Jaszi, G. (1986), 'An Economic Accountant's Audit', *American Economic Association Papers and Proceedings*, pp. 411–417.

Jones, R. (2000), 'National accounting, government budgeting and the accounting discipline', *Financial Accountability & Management*, Vol. 16, issue 2, pp. 101–116.

Kampmann, V. (1942), 'Danmarks nationalindkomst i trediverne' (Denmark's national income in the 1930s), lecture to Nationaløkonomisk Forening 19. October 1942. In: *NationalOkonomisk Tidsskrift*, Vol. 80, pp. 367–384.

Kay, J. (2003), *Culture and prosperity; the truth about markets, why some nations are rich but most remain poor*, Harper Business.

Kenessey, Z (ed.) (1994), *The Accounts of Nations*, IOS Press, Amsterdam.

Keynes, J.M. (1920), *The economic consequences of the peace*, MacMillan and Co., Ltd, London.

Keynes, J.M. (1936), *The general theory of employment, interest and money*, MacMillan and Cambridge University Press, London.

Keynes, J.M. (1940), *How to pay for the war,* Harcourt Brace & co, New York.

King, G. (1936), *Two Tracts*, John Hopkins Press, Baltimore.

Krugman, P.R. (1991), 'Increasing returns and economic geography', *Journal of Political Economy*, Vol. 99, issue 3, pp. 483–499.

Krugman, P.R. and M. Obstfield (2011), *International economics; theory and policy*, Harper Collins College Publishers.

Kuznets, S., (1934), 'National income, 1929–1932', *NBER Bulletin*, No 49.

Kuznets, S. (1938), *Commodity flow and capital formation*. NBER Books, New York.

Kuznets, S. (1948a), 'National Income: a New Version (Discussion of the new Department of Commerce Income Series)', *The Review of Economics and Statistics*, Vol. 30, issue 3, pp. 151–179.

Kuznets, S. (1948b), 'On the Valuation of Social Income — Reflection on professor Hicks' article', *Economica*, XV(57), pp. 1–16 (part I), XV(58), pp. 116–131 (part II).

Lepenies, P. (2016), *The power of a single number; a political history of GDP*, Columbia University Press.

Lequiller, F. and D. Blades (2014), *Understanding national accounts*, OECD Publishing, Paris.

Leontief, W.W. (1936), 'Quantitative Input and Output Relations in the Economic System of the United States', *Review of Economic Statistics*, Vol. 28, issue 3, pp. 105–125.

Lindahl, E., E. Dahlgren and K. Kock (1937), *The National income in Sweden. 1861–1930*. Parts one and two. University of Stockholm, Institute for Social Sciences, London.

Lundberg, E. (1984), 'Lundberg on Kuznets'. In: H.W. Spiegel and W.J. Samuels, *Contemporary economists in perspective*, Part II, JAI Press Inc., London, pp. 523–542.

Maddison, A. (2003), *The World Economy: Historical statistics*, OECD Development Centre Studies, OECD, Paris.

Maddison, A. (2005), 'Measuring and interpreting world economic performance 1500–2001', *Review of Income and Wealth*, Vol. 51, issue 1, pp. 1–36.

Maddison, A. (2007), 'Contours of the World Economy, 1–2030 AD'. *Essays in Macro-Economic History*, Oxford University Press

McCulla, S.H., K.E. Moses and B.R. Moulton (2015), 'The National Income and Product Accounts and the System of National Accounts 2008, comparison and research plans', *Survey of Current Business*, June 2015, pp. 1–17.

Meade, J.A. and R. Stone (1941), 'The Construction of Tables of National Income, Expenditure, Savings and Investment', *Economic Journal* (51)202/203, pp. 216–231.

Meade, J.A. (1951), *The balance of payments*, Oxford University Press, London.

Meek, R.L. (1962), *The Economics of Physiocracy: Essays and Translations*, Harvard University Press.

Metelerkamp, R. (1804), *De toestand van Nederland in vergelijking gebragt met die van enige andere landen van Europa*, Cornel en van Baalen, Rotterdam.

Mishan, E.J. (1969), *The Costs of Economic Growth*, Penguin Books Ltd, New York.

Morgan, M.S. (1990), *The history of econometric ideas*, Cambridge University Press, Cambridge.

Nordhaus, W. and J. Tobin (1972), 'Is Growth Obsolete?' in: *National Bureau of Economic Research, Economic Growth*, Columbia University Press, New York, pp. 1–80.

OECD (2015), *How's Life? 2015: Measuring Well-being*, OECD Publishing

Okun, A.M. (1971), 'Social Welfare has no Price Tag', *Survey of Current Business*, Vol. 51, issue 7, pp. 129–133.

Patinkin, D. (1976), 'Keynes and Econometrics: on the Interaction Between the Macroeconomic Revolutions of the Interwar Period', *Econometrica*, Vol. 44, issue 6, pp. 1091–1123.

Piketty, T. (2014), *Capital in the twenty-first century,* Harvard University Press.

Philipsen, D. (2017), *The little big number: how GDP came to rule the world and what to do about it*, Princeton University Press.

Reich, U. (1994), 'German national accounts between politics and academics', in Kenessey (1994), pp. 144–168.

Reinhart, C. and K. Rogoff (2011), 'From financial crash to debt crisis', *American Economic Review*, Vol. 101, issue 5, pp. 1676–1706.

Ruggles, R. (1990), 'A Note on the Revision of the United Nations System of National Accounts' *Review of Income and Wealth*, Vol. 36, issue 4, pp. 413–420.

Sethia, D. (2014), 'Regional accounts of India: methods, new estimates and their uses', *Review of Income and Wealth*, Vol. 62, issue 1, pp. 92–119.

Stiglitz, J.E., A. Sen and J.P. Fitoussi (2009), Report by the Commission on the measurement of economic performance and social progress.

Stone, R. (1947), 'Definition and Measurement of the National Income and Related Total'. Appendix in UN (1947).

Stone, R. (1951), 'The Use and Development of National Income and Expenditure Estimates'. In: D.N. Chester (ed.), *Lessons of the British war economy*, Cambridge University Press, pp. 83–101.

Stone, R. (1997), *Some British Empiricists in the Social Sciences 1650–1900*, Cambridge University Press, Cambridge.

Studenski, P. (1958), *The Income of Nations*, New York University Press, New York.

Taleb, N.N. (2007), *The Black Swan: The Impact of the Highly Improbable*, Random House.

Tinbergen, J. (1936), Prae-advies over de vragen: Kan hier te lande, al dan niet na overheidsingrijpen, een verbetering van de binnenlandsche conjunctuur intreden, ook zonder verbetering van onze exportpositie? Welke leering kan ten aanzien van dit vraagstuk worden getrokken uit de ervaringen van andere landen?, in: Vereniging voor de Staathuishoudkunde en de statistiek, Praeadviezen, Martinus Nijhoff, 's Gravenhage, pp. 62–108.

Tooze, J.A. (2001), *Statistics and the German State: the making of modern economic knowledge*, Cambridge University Press.

UN (1947), *Measurement of National Income and the Construction of Social Accounts*, Studies and Reports on Statistical Methods No 7, UN, Geneva.

UN (1953), *A System of National Accounts and Supporting Tables ('SNA58')*, Studies in Methods, Series F No 2, Rev. 1, UN, New York.

UN (1968), *A System of National Accounts ('SNA68')*, Studies in Methods, Series F No 2, Rev. 3, UN, New York.

UN (1993), *System of national accounts ('SNA93')*, Joint publication by the European Commission, IMF, OECD, World Bank and UN.

UNDP (United Nations Development Programme) (1991), *Human Development Report 1991,* Oxford University Press, New York.

UN (2002), *Use of Macro Accounts in Policy Analysis*, Studies in Methods, Series F, No 81, UN, New York.

United Nations, European Commission, International Monetary Fund, OECD and World Bank (2008), *System of National Accounts 2008 ('SNA 2008')*, New York, USA.

UNECE (2016), *Guide on measuring human capital,* New York and Geneva.

Vanoli, A. (2005), A history of national accounting, IOS Press, Amsterdam.

Vanoli, A. (2014), 'National accounting at the beginning of the 21st century: Wherefrom? Whereto?', Eurostat *Review of National Accounts and Macroeconomic Indicators (EURONA),* 1/2014, pp. 9–38.

Ven, P. van de (2015), 'New standards for compiling national accounts: what's the impact on GDP and other macro-economic indicators?', *OECD Statistics Brief*, February 2015, No 20.

Ven, P. van de (2017), 'Present and Future Challenges to the System of National Accounts: Linking Micro and Macro', *Review of Income and Wealth*, forthcoming.

Yamashita, T. (2013), 'A guide to the Integrated Macroeconomic Accounts', *Survey of Current Business*, April 2013, pp. 12–27.

Young, A. (1770), A six months' tour through the north of England, 4 volumes, Strahan and Nicoll, London.

Zanden, J.L. van, J. Baten, M.M.d'Ercole, A. Rijpma, C. Smith and M. Timmer (2014), *How was life? Global well-being since 1820*, OECD Publishing.

Zwijnenburg, J., S. Bournot and F. Giovannelli (2017), 'Report by expert group on disparities in a national accounts framework; results from the 2015 exercise', *OECD Statistics Working Paper* 2016/10.

3

Processing scanner data by an augmented GUV index (*)

LUDWIG VON AUER ([1])

Abstract: Statistics Netherlands is a pioneer in the use of scanner data for price measurement purposes. In his paper 'A New Methodology For Processing Scanner Data in the Dutch CPI' Antonio G. Chessa (2016a) of Statistics Netherlands proposes an approach for a more coherent and automated processing of scanner data. This approach is denoted here as the 'C-approach'.

The C-approach uses a two stage aggregation procedure in which individual items are bundled into products and then products into segments. On the lower stage (items → products), for each product and each month, a unit value and an aggregate quantity is computed. These product data are used in the upper stage (products → segments), where for each segment an average price change is computed. The computations on the upper stage utilize the Geary-Khamis index. The aggregation of segments into the overall price change of the economy uses the traditional Laspeyres-type procedure. The C-approach is not concerned with this final step of price aggregation.

The C-approach is an important advancement in official price measurement. At the same time this paper argues that in some situations the C-approach might be unnecessarily complex. As an alternative, it proposes to replace the Geary-Khamis index by a simpler formula like the Gerardi index or even an augmented Generalized Unit Value index. Furthermore, it recommends to examine whether for some segments the two stage aggregation of the C-approach can be replaced by a single stage aggregation.

JEL codes: C43, E31

Keywords: CPI, scanner data, relaunch, assignment bias, assortment bias, Lehr, GUV index

([1]) Universität Trier, Germany

(*) Companion paper to 'A New Methodology For Processing Scanner Data in the Dutch CPI' by Antonio G. Chessa (2016a).

1. Introduction

In traditional price collection, employees of national statistical offices visit shops and collect the prices of a sample of items. Scanner data are an attractive alternative to this costly and time consuming approach. They provide the price analyst with information on (1) large samples of items instead of the rather restricted samples available by the traditional price collection procedures, (2) transaction prices instead of offer prices, (3) quantities sold, and (4) additional features of the items (e.g. some price-determining qualitative characteristics).

Statistics Netherlands is one of the pioneers in the use of scanner data for price measurement purposes. Other countries benefit from the experience and innovations that Statistics Netherlands accumulated over two decades. In a recent paper, Chessa (2016a) builds on this experience and proposes additional innovations. He describes a procedure for handling and using scanner data in a more efficient and coherent way. For simplicity, his procedure is denoted here as the 'C-approach'. It is applicable to different types of scanner data and different types of goods and services.

Scanner data provide the price analyst with an unusually rich set of information. The C-approach is a sophisticated method that is able to utilize this information. The price change between the base month and some comparison month is computed as the ratio of the average price level of the comparison month and the average price level of the base month. This is an alternative to the orthodoxy of price index theory, that is, to the averaging of the individual items' price ratios ([2]).

The C-approach is a valuable contribution to the discussion on the most attractive price measurement methodology. The proposed methods have been tested with scanner data provided to Statistics Netherlands. Therefore, the C-approach is not just an intellectual exercise, but a tested procedure for a more efficient and coherent computation of reliable inflation numbers.

The present paper attempts to contribute in two ways. First, it provides a streamlined exposition of the C-approach together with a thorough analysis of its strengths and weaknesses. Second, a recommendation for a simplified approach is developed. Such a simplification is desirable, because it would reduce the obstacles to a wider integration of scanner data information into official price measurement.

The C-approach is a two stage procedure, where, on the lower stage, items are aggregated into products and, on the upper stage, these products are aggregated into a segment. At the lower stage, for each product and each month a *representative price* is computed by the unit value formula. In addition, the total monthly quantity for each product is calculated. At the upper stage, this information is used to compute the *average price* change between a (fixed) base period month and the most recent month. For this purpose the Geary-Khamis index is applied. This is a multilateral price index formula that requires a simultaneous solution of several equations. With computers, the solution can be obtained by iterative computational procedures.

The present paper propose to replace the Geary-Khamis index by a simpler multilateral index formula or by an augmented bilateral index formula. For example, the multilateral Gerardi index does not require iterative solution procedures. Alternatively, one may use an augmented

([2]) As described in Auer (2009, p.3), this orthodoxy can be traced back to Fisher's (1922, p. 451) seminal work.

version of some generalized unit value index described in Auer (2014). The generalized unit value indices are bilateral. Therefore their computation is much simpler than that of the Geary-Khamis index. If the results generated by such simplified procedures do not significantly deviate from the results from some suitable benchmark procedure, the simplified procedures may represent an attractive alternative to the Geary-Khamis index.

The paper is organized as follows. Section 2 explains the C-approach in more detail. A discussion of that approach can be found in Section 3. Section 4 proposes the simplified approaches and Section 5 concludes.

2. Two stage aggregation

To understand the basic idea of the C-approach, its strengths and weaknesses, and the modifications proposed in this paper, a few preliminary remarks on the processing of scanner data at Statistics Netherlands are necessary. Suppose that Statistics Netherlands received the scanner data on transacted men's T-shirts. The data cover many different men's T-shirts. Each specific T-shirt is denoted as an *item* (e.g. C&A package of two white cotton men's T-shirts with short sleeves and round neck). Each item covered by the scanner data has its own identification number, denoted as GTIN. Statistics Netherlands bundles items (GTINs) that are sufficiently similar into *products*. For example, all packages of two white cotton men's T-shirts with short sleeves are grouped into the product 'package of two white cotton men's T-shirts with short sleeves'; see Figure 1. This is the lower stage of aggregation. This stage of aggregation will be discussed in Section 2.1.

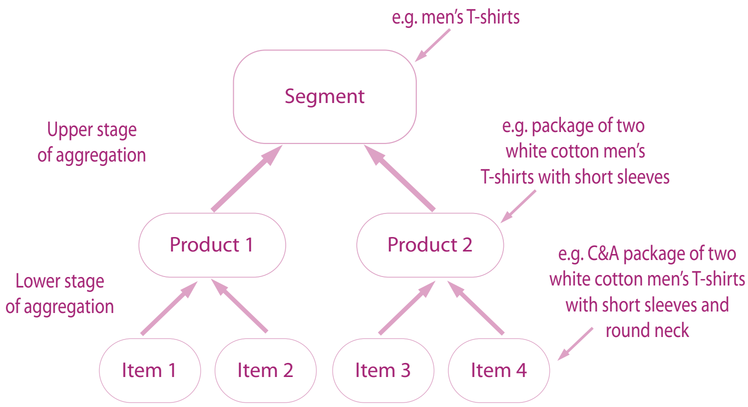
Products that are closely related to each other are bundled into a (consumption) *segment*. For example, all products representing men's T-shirts are grouped into the segment 'men's T-shirts'; see Figure 1. This is the upper stage of aggregation. It is studied in Section 2.2.

2.1. Lower stage of aggregation: bundling items into products

Suppose that we are employed by Statistics Netherlands and that we have to compute the average price change of some segment s (e.g. men's T-shirts). The provider of the scanner data has supplied us with sufficient information such that we can compute for each item j belonging to segment s a monthly representative price, p_j^t, and the monthly quantity transacted, q_j^t. Furthermore, the data contain attributes describing the items (e.g. brand, fabric, ...). The data cover several months, ranging from the base month $t = 0$ to the most recent month $t = T$. In principle, we could aggregate all data in one step to compute the average price change of segment s between months 0 and T. As we will argue below, this single stage procedure would be the preferable approach, if the set of items remained constant over time. However, many segments with unstable sets of items exist. Therefore, the C-approach applies the two stage procedure depicted in Figure 1.

At the lower stage of aggregation the segment's items are grouped into products. Which items j of segment s should be aggregated into a product? For example, which items representing men's T-shirts should be bundled into a product? In the C-approach, items are bundled into products according to their *attributes* (e.g. T-shirts per package, brand, fabric, colour, length of sleeves).

Figure 1: The two stage aggregation procedure of Statistics Netherlands.



The term *characteristic* refers to a specific value that an attribute can take (Chessa (2016a, p 53)). For example, C&A and Hugo Boss are different characteristics of the attribute 'brand'. In the experimental data set used in the paper all packages of two white cotton men's T-shirts with short sleeves are aggregated into the product 'package of two white cotton men's T-shirts with short sleeves'. The choice of this definition is based on a sensitivity analysis: an attribute was deemed as important for defining a product when the exclusion of that attribute significantly changed the price index for the whole segment 'men's T-shirts' (Chessa (2016a, p 54)). This sensitivity analysis identifies a set of relevant attributes. Items that, with respect to these relevant attributes, show the same characteristics, are grouped into one product.

After the grouping of items into products is completed, for each month *t* and each product *i* of segment *s*, some representative price, p_i^t , and some aggregate quantity, q_i^t, is computed. Let A_i^t denote the set of items *j* of segment *s* that in month *t* are bundled into product *i*. For each month *t* and each product *i*, the C-approach computes the representative price, p_i^t, as the *unit value* (Segnitz (1870, p. 184)):

$$(1)\quad p_i^t = \frac{\sum_{j \in A_i^t} p_j^t q_j^t}{\sum_{j \in A_i^t} q_j^t}.$$

The *aggregate quantity*, q_i^t, is given by

$$(2)\quad q_i^t = \sum_{j \in A_i^t} q_j^t .$$

Note that at this lower stage of aggregation no price *changes* or price *indices* are computed. Only in the upper stage of aggregation the information on the unit values, p_i^t, and aggregate quantities, q_i^t, is aggregated into the average price change of segment *s*. To this upper stage we turn now.

2.2. Upper stage of aggregation: bundling products into segments

In this upper stage of aggregation two tasks are executed simultaneously. Different products are aggregated into segments and, for that segment, an average price change is computed. In this upper stage aggregation the C-approach uses an index formula labeled as *quality adjusted unit value index* (QAUV index). To understand this index formula, it is necessary to understand the *unit value index* (UV index) introduced by Drobisch (1871a, p 39; 1871b, p 149) ([3]).

Let G_s^t denote the set of products *i* that are assigned to segment *s*. For each product *i* and each month *t*, the corresponding unit value, p_i^t , and the total number of units transacted, q_i^t, are known. They were computed at the lower stage of aggregation by formulae (1) and (2). Also the attributes describing the product (e.g. package of two white cotton men's T-shirts with short sleeves) is known.

We want to compute for segment *s* the average price change between the base month, $t = 0$ (e.g. December 2014), and the most recent month, $t = T$ (e.g. December 2015). As one possibility, we could compute for month $t = 0$ and also for month $t = T$ the segment's unit value:

$$(3) \quad P_{UV}^{t} = \frac{\sum_{i \in G_s^t} p_i^t q_i^t}{\sum_{i \in G_s^t} q_i^t} \quad \text{for } t = 0, T.$$

This gives us the segment's two unit values P_{UV}^0 and P_{UV}^T .

The UV index, $P_{UV}^{T/0}$, is the ratio of the two unit values P_{UV}^T and P_{UV}^0 . It measures the average price level change of segment *s* between months 0 and *T*:

$$(4) \quad P_{UV}^{T/0} = \frac{P_{UV}^T}{P_{UV}^0} = \frac{\left(\sum_{i \in G_s^T} p_i^T q_i^T\right) / \left(\sum_{i \in G_s^T} q_i^T\right)}{\left(\sum_{i \in G_s^0} p_i^0 q_i^0\right) / \left(\sum_{i \in G_s^0} q_i^0\right)} = \frac{\sum_{i \in G_s^T} p_i^T q_i^T}{\sum_{i \in G_s^0} p_i^0 q_i^0} \quad \frac{\sum_{i \in G_s^0} q_i^0}{\sum_{i \in G_s^T} q_i^T}.$$

Note that the sets G_s^T and G_s^0 can differ.

In the UV index (4), the product quantity summations, $\sum_{i \in G_s^t} q_i^t$, yield accurate results only if the product-identifying units being summed are identical. However, a segment comprises products of very different quality. These products have dissimilar product-identifying units. Consequently, they are unsuitable for the simple quantity summations in the UV index (4). For example, in the segment of men's T-shirts it is inappropriate to add 3 packages of black Hugo Boss T-shirts and 2 packages of white C&A T-shirts to get 5 packages of T-shirts. Therefore, the UV formula (3) and the UV index (4) are inappropriate. An amended version of the unit value concept is required. This is accomplished by the QAUV index.

The QAUV index includes quality adjustment factors, z_i. They are defined to be an appropriate number of common units per product-identifying unit. For example, if consumers considered the package of Hugo Boss T-shirts as 'three times as valuable' as the package of C&A T-shirts, one could use $z_{Hugo\ Boss} = 3$ and $z_{C\&A} = 1$. With these numbers, the common unit would be equivalent

([3]) The following exposition of the UV index and the QAUV index draws on Auer (2014).

to one package of C&A T-shirts. Then, one package of Hugo Boss T-shirts would represent three common units. Of course, $z_{Hugo\ Boss} = 2$ and $z_{C\&A} = 2/3$ would be equally appropriate. Then, the common unit would be two thirds of a package of C&A T-shirts. Accordingly, the quality adjusted prices, p_i^t / z_i, become monetary units per common unit of product *i* and the quality adjusted quantities, $q_i^t z_i$, are the number of common units transacted in the form of product *i*.

By definition, the common units of all products *i* in set G_s^t are identical and, for that reason, reliable results are now obtained from the quantity summations, $\sum q_i^t z_i$. The formula for the quality adjusted unit value becomes

$$(5) \quad P^t = \frac{\sum_{i \in G_s^t} (p_i^t / z_i) q_i^t z_i}{\sum_{i \in G_s^t} q_i^t z_i} = \frac{\sum_{i \in G_s^t} p_i^t q_i^t}{\sum_{i \in G_s^t} q_i^t z_i} .$$

Note that the value aggregate, $\sum_{i \in G_s^t} p_i^t q_i^t$, remains unaffected by this transformation, since the quality adjustment factors, z_i, cancel. Formula (5) replaces formula (3). Formula (5) is the backbone of the upper level aggregation of the C-approach (Chessa (2016a, pp 56-57)) (4).

Consequently, the QAUV index is:

$$(6) \quad P_{QAUV}^{T/0} = \frac{P^T}{P^0} = \frac{\sum_{i \in G_s^T} p_i^T q_i^T}{\sum_{i \in G_s^0} p_i^0 q_i^0} \frac{\sum_{i \in G_s^0} q_i^0 z_i}{\sum_{i \in G_s^T} q_i^T z_i} .$$

This index measures the change in the unit value of a common unit in segment *s* between months 0 and *T* (5).

If the data required for the determination of the quality adjustment factors, z_i, is not directly observable, then some form of estimation is required. The simplest estimation procedure is to use an item's observed prices, p_i^t, as indicators of the item's quality, z_i. For example, if the unit value of a package of Hugo Boss T-shirts in both months is EUR 40, whereas the unit value of a package of C&A T-shirts in both months is only EUR 10, then $z_{Hugo\ Boss} = 4$ and $z_{C\&A} = 1$ are reasonable quality adjustment factors. This approach rests on the assumption that the items' prices adequately reflect the items' differences in quality. There is some literature that suggests that this assumption is often violated. For example, when firms pursue market skimming, items are overpriced at the beginning of the life cycle. The strategy of market penetration leads to the opposite effect.

Also Chessa (2016a, p 57) assumes that an item's price adequately reflects the item's quality. Then the products' unit values can be exploited to quantify the differences in the products' quality. The idea was already mentioned in de Haan (2002, p 82). He recommends using a time period in which the products were being sold in the marketplace and they were preferably in a state of equilibrium. Furthermore, Auer (2014) provides a detailed elaboration and generalization of

(4) Chessa (2016a, p 9) denotes the quality adjustment factors, z_i, as 'quantity weights' and uses the symbol v_i. It is certainly true that the quality adjustment factors boil down to 'quantity weights'. However, that term may hide the true nature of the quality adjustment factors: they transform prices and quantities related to product-identifying units into prices and quantities related to common units. Note also that in the price index literature, the symbol v_i is usually reserved for the result of multiplying price by quantity: $v_i^t = p_i^t q_i^t$. Therefore, in the present paper we denote the adjustment factors by z_i instead of Chessa's v_i.

(5) The QAUV index (6) was originally proposed by Dalén (2001, p 11) and de Haan (2002, pp 81-82). Additional elucidations, together with some empirical applications, can be found in de Haan (2004, pp 6-7). A related proposal is provided by Silver (2010, p S220). A detailed elaboration of the QAUV index can be found in Auer (2014). Chessa (2016a, p 57) points out that the Laspeyres index and Paasche index are special cases of the QAUV index formula (5). This was shown already in Auer (2014, pp 850-51).

this basic idea. He also points out that the basic idea can be traced back to Lehr (1885, pp 37-39). However, publishing in the German language, Lehr's work went largely unnoticed. Accordingly, Davies (1924, pp 183-185) some forty years later came up with the same basic idea (6).

If the quality adjustment factors, z_i, were computed only from data associated with months 0 and T, the data of all other months ($t = 1, 2, \dots, T-1$) would be redundant and formula (5) would be applied only to months 0 and T. However, as outlined below, in the C-approach the quality adjustment factors, z_i, are computed from data of all months. This proposal is new to the literature.

Formula (5) is the basis of the upper stage aggregation of the C-approach. The formula computes, for each month t ($t = 0, 1, \dots, T$) the quality adjusted unit value of segment s. In Equation (6), the formula is used for comparing the price levels of months $t = 0$ and $t = T$. This is a *binary* comparison. However, formula (5) can be applied also for *multilateral* comparisons, and the C-approach uses such a multilateral procedure (7). Usually, multilateral indices are used for interregional price comparisons where the price levels of many different regions must be compared in a consistent way. However, Balk (1981) realizes that such multilateral indices can be applied also in intertemporal price comparisons where more than two periods are to be compared. His idea is revived by de Haan and van der Grient (2011) as well as by Ivancic *et al* (2011). By employing a multilateral price index they avoid a bias that is known as chain drift. This bias can arise from chained bilateral price indices. The C-approach does not use chaining. Instead it considers a situation with thirteen consecutive months, with the fixed base month being the December of some year and the final month being the December of the following year.

A well known multilateral index formula is the Geary-Khamis index (Geary (1958), Khamis (1972)). This index uses formula (5) and computes the quality adjustment factor, z_i, of each product i in the following way:

$$(7) \quad z_i = \sum_{t=0}^{T} \varphi_i^t \frac{p_i^t}{P^t} ,$$

with

$$(8) \quad \varphi_i^t = \frac{q_i^t}{\sum_{r=0}^{T} q_i^r} .$$

Equation (7) is a weighted average of the unit values, p_i^t, of all $T + 1$ months where the monthly weight, φ_i^t, is the quantity share of that month, and the unit value, p_i^t, of each month t, is deflated by P^t, the quality adjusted unit value of the whole segment s during that month. The deflation by P^t is the reason why formula (5) must be applied to all months simultaneously and not just to months 0 and T.

In sum, the Geary-Khamis index utilizes formula (5) to compute a sequence of quality adjusted unit values: $P^0, P^1, \dots, P^T$. The calculation of each quality adjusted unit value, P^t, is not just based on the data of month t but also on the data of all other months. This can be seen from formulae (7) and (8). They compute the quality adjustment factors, z_i, from the data of all months.

(6) A systematic elaboration of the propositions of Lehr, Davies and de Haan is provided in Auer (2014). Following Lehr (1885) and Davies (1924), he explains that formula (6) has a much wider use than enabling the aggregation of similar products. Therefore, he labels formula (6) as the family of Generalized Unit Value indices (GUV index) and the 'quality adjustment factors', z_i, as 'transformation rates'.

(7) For surveys on the literature on multilateral price comparisons see, for example, Balk (2008) or Auer (2012).

Formula (5) reveals that the quality adjusted unit values, P^t, depend on the quality adjustment factors, z_i, whereas formula (7) reveals that the opposite dependency is also true. Therefore, the values of P^t ($t = 0, 1, \dots, T$) and z_i ($i \in \{ G_s^0 \cup G_s^1 \cup \dots \cup G_s^T$) must be computed simultaneously. Computers allow us to derive the results by an iterative procedure that is described, for example, in Maddison and Rao (1996, pp 14-17).

In each month, the C-approach calculates a Geary-Khamis index that includes all months from the fixed base month (December of the previous year) to the most recent month for which data are available. For example, at the end of January 2015 only the data of January 2015 and the fixed base month December 2014 are included in the Geary- Khamis index calculation (the Geary-Khamis index simplifies to a bilateral index). This yields the quality adjusted unit values $P^{Dec}_{(Jan)}$, and $P^{Jan}_{(Jan)}$, where the brackets in the subscript indicate the last month for which data were available, here January 2015. The average price change is

$$P^{Jan/Dec}_{GK(Jan)} = \frac{P^{Jan}_{(Jan)}}{P^{Dec}_{(Jan)}} ,$$

where 'GK' stands for Geary-Khamis. The index value $P^{Jan/Dec}_{GK(Jan)}$ is the 'official measure' of the price change of segment *s* between January 2015 and the base month December 2014.

In February, the calculation of the Geary-Khamis index is based on the data of December, January, and February. This yields new quality adjustment factors, z_i, and therefore, new quality adjusted unit values $P^{Dec}_{(Feb)}$, $P^{Jan}_{(Feb)}$, and $P^{Feb}_{(Feb)}$. As a consequence, one gets new price changes $P^{Jan/Dec}_{GK(Feb)}$, $P^{Feb/Dec}_{GK(Feb)}$, and $P^{Feb/Jan}_{GK(Feb)}$. The price change $P^{Feb/Dec}_{GK(Feb)}$ is the 'official measure' of the price change of segment *s* between February 2015 and the base month December 2014. For March and each subsequent month until December the same procedure is executed. Finally, this yields a sequence of twelve index numbers: $P^{Jan/Dec}_{GK(Jan)}$, $P^{Feb/Dec}_{GK(Feb)}$, … , $P^{Dec/Dec}_{GK(Dec)}$. This sequence Chessa (2016a, p 58) denotes as the 'real time index' and it is this index that is published by Statistics Netherlands as the official price index of segment *s*.

Only in the December 2015 index computation ($P^{Dec/Dec}_{GK(Dec)}$) the information of all 13 months is included. In that month 12 x 12 index numbers are computed. Among these index numbers is the sequence of monthly index numbers that use December 2014 as their constant base: $P^{Jan/Dec}_{GK(Dec)}$, $P^{Feb/Dec}_{GK(Dec)}$, … , $P^{Dec/Dec}_{GK(Dec)}$. Chessa (2016a, p 58) denotes this sequence of index numbers as the 'benchmark index', because each of these index numbers utilizes the complete information available in 2015. Later in the paper, the benchmark index is compared to the real time index (Chessa (2016a), pp 61-62). No significant deviations occur.

In 2016 twelve new index numbers are computed. However, they use December 2015 as their fixed base month. This completes the description of the C-approach. The following section discusses the strengths and weaknesses of this procedure.

3. Discussion

The C-approach computes the average price change of the items *j* belonging to segment *s* between the base month $t = 0$ and the most recent month $t = T$. As described in the previous section, this is accomplished in a two stage procedure. At the lower stage of aggregation, items are grouped into products. For each month *t* and each product *i*, the C-approach computes a unit value, p_i^t, and an aggregated quantity, q_i^t. This is done by formulae (1) and (2). At the upper stage of aggregation, the C-approach aggregates the unit values, p_i^t, and aggregated quantities, q_i^t, of the products of segment s into the $T+1$ price levels $P_{(T)}^t$ ($t = 0, 1, \ldots, T$). For this purpose, the C-approach utilizes the Geary-Khamis index, that is, formulae (5), (7), and (8). Finally, the ratio

$$P_{GK(T)}^{T/0} = \frac{P_{(T)}^T}{P_{(T)}^0}$$

is the official measure of the average price change of segment s between months $t = 0$ and $t = T$.

Two features of this C-approach require a more elaborate discussion: (A) The approach applies a two stage aggregation instead of a single stage aggregation and (B) at the upper stage of aggregation the index computation between some fixed base month $t = 0$ and the most recent month $t = T$ uses not only the information of these two months, but relies on the Geary-Khamis index which utilizes also the information provided by all in-between months. The following discussion starts with the second feature and then moves on to the first feature.

3.1. Geary-Khamis index

The Geary-Khamis index calculates for each month t ($t = 0, 1, \ldots, T$) a quality adjusted unit value $P_{(T)}^t$. Again, the brackets in the subscript of the unit value indicate the last month included in its calculation. When $T = 2$, the quality adjusted unit values $P_{(2)}^0$, $P_{(2)}^1$, and $P_{(2)}^2$ are computed. These three numbers allow for three price comparisons:

$$P_{GK(2)}^{1/0} = \frac{P_{(2)}^1}{P_{(2)}^0}, \quad P_{GK(2)}^{2/0} = \frac{P_{(2)}^2}{P_{(2)}^0}, \text{ and } \quad P_{GK(2)}^{2/1} = \frac{P_{(2)}^2}{P_{(2)}^1}.$$

Therefore, the Geary-Khamis index is called a multilateral price index. Note also that

$$P_{GK(2)}^{2/0} = P_{GK(2)}^{1/0} \bullet P_{GK(2)}^{2/0}.$$

Therefore, the Geary-Khamis index is classified as a *transitive* multilateral price index. Transitivity allows us to compare the quality adjusted unit values of any pair of months in a consistent way.

As pointed out before, the computation of the Geary-Khamis index is not easy and requires an iterative computer procedure. However, other multilateral price indices exist that, due to their similarity to the Geary-Khamis index, are likely to generate almost identical results, but are much simpler to compute. For example, the Gerardi index proposed in Eurostat (1978, p 30) is similar to the Geary-Khamis index but replaces formulae (7) and (8) by the much simpler formula

$$(9) \quad z_i = \prod_{t=0}^{T} (p_i^t)^{1/T}.$$

This is the geometric average of all $T+1$ unit values, p_t^i, of product i ([8]).

In contrast to the Geary-Khamis index, the quality adjustment factors, z_i, of the Gerardi index are independent from the quality adjusted unit values $P_{(T)}^t$. As a consequence, the simultaneous computation of the z_i-values and the $P_{(T)}^t$-values is no longer necessary. This simplifies the

([8]) Of course, other multilateral indices exist that are also easier to compute than the Geary-Khamis index. An example is the GEKS-index proposed by Gini (1924), Éltető and Köves (1964) and Szulc (1964). However, this index has a completely different construction principle than the Geary-Khamis index and the Gerardi index (e.g. Balk (2008); Auer (2012)).

calculation considerably and it eases the implementation of scanner data procedures in the computational routines of the national statistical offices.

Does the omission of the quality adjusted unit values, $P^t_{(T)}$, introduce a source of bias in the computation of the quality adjustment factors, z_i? In interregional price comparisons the unit values may vary significantly. This is particularly true when the regions have different currencies. However, in the C-approach we compare months instead of regions. Usually, the variance of the values of $P^t_{(T)}$ is much smaller than that in interregional price comparisons. Therefore, we are unlikely to inject any significant bias when we compute the z_i-values without the $P^t_{(T)}$-values. Probably, the Gerardi index will give very similar results as the Geary-Khamis index. Admittedly, this conjecture remains to be empirically verified with scanner data such as those of Statistics Netherlands (9).

In each consecutive month, the C-approach computes a new Geary-Khamis index. This yields a sequence of Geary-Khamis indices. 'Sequence of Geary-Khamis indices' means that in each consecutive month a new set of quality adjusted unit values is computed. For example, in month $T-1$ the quality adjusted unit values

$$P^{0}_{(T-1)}\ ,\ P^{1}_{(T-1)}\ ,\ldots,\ P^{T-1}_{(T-1)}$$

are computed and from those the indices

$$P^{1/0}_{GK(T-1)}\ ,\ P^{2/0}_{GK(T-1)}\ ,\ldots,\ P^{T-1/0}_{GK(T-1)}\ .$$

In the following month the quality adjusted unit values

$$P^{0}_{(T)}\ ,\ P^{1}_{(T)}\ ,\ldots,\ P^{T-1}_{(T)}\ ,\ P^{T}_{(T)}$$

are calculated and from those the indices

$$P^{1/0}_{GK(T)}\ ,\ P^{2/0}_{GK(T)}\ ,\ldots,\ P^{T-1/0}_{GK(T)}\ ,\ P^{T/0}_{GK(T)}\ .$$

Usually one gets

$$P^{1/0}_{GK(T-1)} \neq P^{1/0}_{GK(T)}\ ,\ P^{2/0}_{GK(T-1)} \neq P^{2/0}_{GK(T)}\ ,\ldots\ .$$

In other words, the results produced by the sequence of Geary-Khamis indices are inconsistent.

The C-approach 'solves' this inconsistency problem by simply keeping only the adjusted unit values $P^{1/0}_{GK(1)}, P^{2/0}_{GK(2)}, \ldots, P^{T/0}_{GK(T)}$. This is Chessa's (2016a, p 58) real time index. To get at this real time index, a large set of inconsistent quality adjusted unit value indices is calculated and then, to ensure consistency, most of them are discarded.

If one used a sequence of Gerardi indices instead of a sequence of Geary-Khamis indices, the inconsistency problem would not go away. It is the 'sequence of multilateral indices' that causes the problem, not the applied index formula.

(9) In private correspondence Antonio Chessa reported to me that for clothing the results generated by the Gerardi index show some deviations to the Geary-Khamis results.

To sum up, for segments with stable price levels the Geary-Khamis index appears unnecessarily complex. If the results are sufficiently similar, it could be replaced by the much simpler Gerardi index. Nevertheless, an inconsistency problem would remain. However, an even simpler procedure avoids the problem of inconsistent price indices. This procedure is described in the recommendations (Section 4) ([10]).

3.2. Two stage aggregation

For computing the average price change of some segment *s*, the C-approach relies on a two stage aggregation. At the lower stage, the prices and quantities of items are aggregated into unit values of products, p_i^t, and aggregate quantities, q_i^t. This is done for each month *t* separately. At the upper stage of aggregation, the Geary-Khamis index is used. This index aggregates the unit values, p_i^t, and aggregated quantities, q_i^t, of the products of segment *s* into the price levels $P_{(T)}^t$ ($t = 0, 1, \dots, T$). Finally, the ratio $P_{GK(T)}^{T/0} = P_{(T)}^T / P_{(T)}^0$ is the official measure of the average price change of segment *s* between months $t = 0$ and $t = T$.

Suppose that the set of items (GTINs) remained constant over time: $A_s^0 = A_s^1 = \dots = A_s^T = A_s$. In such a scenario the two stage aggregation would offer no advantages whatsoever over a single stage aggregation that treats every item as an individual product. Treating each item *j* as a distinct product eliminates the first stage of the two stage aggregation and the quality adjusted unit value formula (5) becomes

$$(10) \quad P^t = \frac{\sum_{j \in A_s} p_j^t q_j^t}{\sum_{j \in A_s} q_j^t z_j},$$

where A_s is the set of items *j* belonging to segment *s*. This single stage formula replaces the three formulae (1), (2), and (5). Coupling formula (10) with the Geary-Khamis formulae (7) and (8) or with the Gerardi formula (9) gives the desired real time index: $P_{(1)}^{1/0}, P_{(2)}^{2/0}, \dots, P_{(T)}^{T/0}$.

In fact, with a constant set of items, the single stage aggregation is more reliable than the two stage aggregation, because it avoids the artificial grouping of items into products that are considered as homogeneous but, in reality, are not. For example, in the example given in Chessa's paper, the product 'package of two white cotton men's T-shirts with short sleeves' can comprise C&A and Hugo Boss packages and counts them as equivalent packages. However, for the measurement of price change it matters whether the aggregated quantity, q_i^t, consists of C&A or of Hugo Boss T-shirts. For example, suppose that all T-shirt prices remain constant over time, but that the share of Hugo Boss T-shirts increases over time. If C&A T-shirts and Hugo Boss T-shirts are assigned to the same product, formula (1) yields an increasing unit-value for that product, falsely indicating a price increase. As a consequence, the average price index computed by (5) is upward biased.

In the following, the bias caused by assigning heterogeneous items to the same product is denoted as the *assignment bias* ([11]). This bias can be upward or downward. It can arise only in the first stage of the C-approach's two stage aggregation. In the single stage aggregation conducted by formula (10) no assignment bias can arise, because no assignment is necessary.

Of course, the C-approach was explicitly designed for situations in which the universe of items changes over time. Chessa (2016a, p 56, Figure 3) presents an empirical example to show that

([10]) The Geary-Khamis index and the alternatives proposed in the present paper are not the only approaches for processing scanner data. Recent surveys include Australian Bureau of Statistics (2016), Chessa *et al* (2017), and Diewert and Fox (2017).
([11]) This effect is often denoted as 'unit value bias'. However, this terminology is somewhat misleading. The problem is not the unit value formula, but the flawed assignment of items to the group of items for which a unit value is computed.

in the presence of strong changes in the set of items the two stage aggregation is less prone to bias than a single stage aggregation. Therefore, we first discuss under which conditions the entry and exit of items causes problems for the single stage aggregation. Afterwards we examine whether the two stage aggregation is able to avoid such problems.

If an item *j* is replaced by the same item *k*, but the replacement has a different GTIN, the replacement is called a 'relaunch' of item *j*. If the analyst is able to identify the relaunch, the two items receive the same quality adjustment factors, $z_j = z_k$. No problem arises. The single stage aggregation remains more reliable than the two stage aggregation, because no assignment bias arises.

What happens in the single stage aggregation when there is an *undetected* relaunch? If the prices of the entering and exiting item are identical and stable over time, there are still no advantages from the two stage aggregation, because in the single stage aggregation we get $z_j = z_k$. This is true for the Geary-Khamis formulae (7) and (8) as well as for the Gerardi formula (9).

However, when an undetected relaunch with deviating prices occurs or when an item is replaced by a completely different item, some bias can arise in the single stage aggregation. Suppose, for example, that the price of item *j* is constant over time and then, in month $t = T$, item *j* is replaced by an equivalent relaunch item *k* that has a higher price. Since the relaunch is not detected, formula (9) would yield $z_j < z_k$ instead of $z_j = z_k$. In the denominator of formula (10), the sum $\sum_{j \in A_s^0} q_j^0 z_j$ would be too small relative to the sum $\sum_{j \in A_s^T} q_j^T z_j$. Therefore, the price level P^0 would be too large relative to P^T. As a consequence, the measured average price change of the segment, P^T / P^0, would be downward biased. This type of bias is denoted here as assortment bias. If item *j* were replaced by an equivalent relaunch item *k* that has a lower price, the assortment bias would be upward.

The G-approach with its two stage aggregation is less prone to assortment bias than the single stage aggregation. If the analyst assigns the new item *k* to a product *i* that accurately reflects item's *k* characteristics, then item k receives this product's quality adjustment factor, z_i, and the assortment bias can be significantly reduced. Of course, an appropriate assignment of the new item *k* to the best fitting product requires a careful analysis of the new item. In other words, in the two stage aggregation, the analyst must invest time and effort into this assignment process. In a single stage aggregation this time and effort could be spend on detecting relaunches, that is, on comparing exiting and entering items. Whether this time and effort would sufficiently reduce the assortment bias of the single stage aggregation is unclear.

The two stage aggregation has another advantage. Sometimes items drop out of the sample without a relaunch. As long as the product to which this item was assigned still contains other items, this does not represent a problem for the price measurement. Similarly, new items may enter the sample without an equivalent predecessor. If some product exists to which the new item can be assigned to, the inclusion of the new item causes no problems. In a single stage aggregation the entry and exit of items raises the usual problems. If the items are simply ignored, bias could arise. If the items are not ignored, the missing prices must be estimated.

In sum, the G-approach's two stage aggregation defined by formulae (1), (2), and (5) suffers from assignment bias, but, with careful analysts, may keep the assortment bias in check. The single stage aggregation defined by formula (10) does not suffer from assignment bias. However, it is prone to considerable assortment bias. The less stable the set of items over time, the more attractive becomes the two stage aggregation procedure of the G-approach relative to a single stage aggregation procedure.

4. Recommendations

The G-approach tackles some important problems in price measurement. The downside to this is the complexity of the G-approach. Can we tackle the problems by less complex approaches? This final section outlines a few ideas.

Recommendation 1: augmented GUV index instead of Geary-Khamis index

In its upper stage of aggregation the G-approach computes the average price change of some segment s between some fixed base month $t = 0$ and the most recent month $t = T$. For this purpose it uses the Geary-Khamis index, that is, formulae (5), (7), and (8). The latter two formulae ensure that the computation of the quality adjustment factors, z_i, uses not only the information provided by months 0 and T, but also the information of all other months (1, 2, …, T − 1). This is a valuable innovation of the G-approach. However, there are simpler ways of including the information of months t = 1, 2, …, T − 1 into the computation of the quality adjustment factors, z_i.

In Section 3.1 of this paper it was explained that there are two disadvantages associated with the Geary-Khamis index. First, the computation of a Geary-Khamis index is complex, and second, the sequence of Geary-Khamis indices produces a huge number of incoherent quality adjusted unit values. Very few of these are necessary to get the real time index. The first disadvantage could be mitigated by using the Gerardi index instead of the Geary-Khamis index. However, the second disadvantage would remain even with the Gerardi index.

Can we compute the quality adjustment factors, z_i, by a simpler procedure as the Geary-Khamis or Gerardi index, and still utilize the information provided by all in-between months? A good starting point for such a procedure is the family of Generalized Unit Value indices (GUV indices). This family of price index formulae is defined and explored in Auer (2014).

The Lehr index is an example. Instead of (7) and (8), Lehr (1885, p 39) computes the quality adjustment factors, z_i, by the simple unit value formula

$$(11) \qquad z_i = \frac{p_i^0 q_i^0 + p_i^T q_i^T}{q_i^0 + q_i^T} .$$

These quality adjustment factors are inserted in formula (5) to compute the quality adjusted unit values P^0 and P^T. The ratio of these two unit values is the Lehr index. Like the Geary-Khamis and the Gerardi index, it measures the price change between the base month $t = 0$ and the most recent month $t = T$. However, it does not calculate the (irrelevant) price changes between all other months. Therefore, the inconsistencies discussed in Section 3.1 of this paper cannot arise.

Closer inspection of the Lehr index and the Geary-Khamis index reveals that between these two indices a close relationship exists. Suppose we apply the Geary-Khamis index for a bilateral price comparison of months $t = 0$ and $t = T$. That is, only information of months $t = 0$ and $t = T$ is processed:

$$P_{GK(2)}^{1/0} = \frac{\sum_{i \in G_s^T} p_i^T q_i^T}{\sum_{i \in G_s^0} p_i^0 q_i^0} \; \frac{\sum_{i \in G_s^0} q_i^0 z_i}{\sum_{i \in G_s^T} q_i^T z_i}$$

(12) $$z_i = \varphi_i^0 \frac{p_i^0}{P^0} + \varphi_i^T \frac{p_i^T}{P^T}$$

(13) $$\varphi_i^0 = \frac{q_i^0}{q_i^0 + q_i^T} \quad and \quad \varphi_i^T = \frac{q_i^T}{q_i^0 + q_i^T} .$$

Inserting (13) in (12) and setting $P^0 = P^T = 1$, (12) simplifies to (11). In other words, applying the Geary-Khamis index to a bilateral price comparison and setting the price levels, P^T, in (12) equal to 1, yields the Lehr index. As pointed out before, when $t = 0$ and $t = T$ represent months instead of regions with different currencies, setting in (12) $P^0 = P^T = 1$ is unlikely to cause discernible bias ([12]).

The Lehr index is simple and the calculation of the quality adjustment factors by formula (11) is intuitively appealing. However, the Lehr index and all other GUV indices are bilateral price indices. They utilize only the information of months $t = 0$ and $t = T$, disregarding the data provided by months $t = 1, 2, \ldots, T-1$. This is a disadvantage. It is usually preferable to make use of all sensible information that is available. The computation of quality adjustment factors, z_i, is no exception. Formulae (7) and (8) utilize all information, whereas formula (11) does not.

Can this disadvantage of the GUV index family be avoided? The obvious solution is to augment the applied GUV index, such that the calculation of the quality adjustment factors utilize also the information provided by months $t = 1, 2, \ldots, T-1$. For example, we can compute the quality adjustment factors by the formula

(14) $$z_i = \frac{\sum_{t=0}^{T} p_i^t q_i^t}{\sum_{t=0}^{T} q_i^t}$$

which is a generalization of formula (11). The computed values can be inserted in formula (5). This yields the quality adjusted unit values $P_{(T)}^0$ and $P_{(T)}^T$. The ratio of these two numbers is the *augmented Lehr index* ([13]). This is still a bilateral price index. It gives the average price change between the two months 0 and T:

$$P_{aLehr(T)}^{T/0} = \frac{P_{(T)}^T}{P_{(T)}^0} .$$

In each new month a new augmented Lehr index must be computed that compares the most recent month, $t = T$, to the fixed base month, $t = 0$.

For example, the price change of segment s between December 2014 and February 2015 is measured by

$$P_{aLehr(Feb)}^{Feb/Dec} = \frac{P_{(Feb)}^{Feb}}{P_{(Feb)}^{Dec}} .$$

The average price change between March 2015 and December 2014 is given by

([12]) Applying the Gerardi index to a bilateral price comparison yields the Davies index, another member of the family of GUV indices (see Auer (2014, pp 851-52), Auer (2012, p 50)).
([13]) The same index has been independently proposed in Lamboray (2017).

$$P_{aLehr(MAr)}^{Mar/Dec} = \frac{P_{(Mar)}^{Mar}}{P_{(Mar)}^{Dec}} .$$

Note that in every month the quality adjusted unit value of the fixed base month (here December 2014) is recomputed. The other quality adjusted unit values are not recomputed. Therefore, in a sequence of augmented Lehr indices far fewer computations are executed than in a sequence of Geary-Khamis or Gerardi indices. Furthermore, the inconsistencies discussed in Section 3.1 cannot arise. In official price statistics this is an important virtue.

It would be useful to know whether the augmented Lehr index produces index numbers that coincide with those generated by a suitable benchmark index. Chessa (2016b, pp 15-16) empirically compares the augmented Lehr index with the real time index of the C-approach. He shows that in segments with a continuous downward trend in prices the augmented Lehr index shows a slightly different result than the real time index. Furthermore, he proposes a bilateral variant of the Geary-Khamis index (denoted as 'revisionless QU-GK index'). In the price level computation of period $t = T$, the price levels obtained from the computations of previous months ($T-1$, $T-2$, …) are not recomputed, but simply inserted in Equation (7). Then, inserting (7) in the quality adjusted unit value formula (5) gives for period $t = T$ an equation with P^T as its only unknown. In addition, Chessa (2016b, pp 16-17) shows for several segments that this bilateral version of the Geary-Khamis index gives virtually the same results as the real time index and the benchmark index generated by the C-approach.

Recommendation 2: single stage aggregation instead of two stage aggregation for stable sets of items

Chessa (2016a) advocates a two stage aggregation procedure (items into products and products into segments) instead of a single stage aggregation (items into segments). Obviously, a two stage aggregation is more complex than a single stage aggregation. Therefore, it should be examined whether the two stage aggregation brings more reliability and, if the answer is 'yes', whether this gain justifies the increase in complexity and intransparency.

The two stage aggregation was defended on the grounds that it reduces assortment bias. However, as argued in Section 3.2 of this paper, it introduces assignment bias. In other words, assortment bias is replaced by assignment bias, and without further investigation it is unclear which of the two is more damaging. The larger the changes in the set of items, the stronger the case for the two stage aggregation.

A single stage aggregation does not give rise to assignment bias. Furthermore, also in the context of a single stage aggregation, the assortment bias can be reduced. Comparing the characteristics of a newly entering item k to the characteristics of the existing items can yield sensible quality adjustment factors z_j. However, this requires a lot of time and effort on the side of the analyst. Furthermore, when items drop out of the sample without replacement or new items enter the sample without a predecessor, the single stage aggregation should estimate the missing prices. This requires additional time and effort.

In segments suitable for the two stage aggregation, the assignment bias should be avoided as far as possible. It is unclear, whether the automated sensitivity analysis described in Chessa's paper can achieve this. For example, it is not very convincing to drop the attribute brand on the grounds that its inclusion did not change the price index of the segment 'men's T-shirt' in a significant way. The attribute 'brand' is highly relevant for characterizing a T-shirt, regardless of the results produced by a sensitivity analysis. Therefore, this attribute should be included, if it is available. Furthermore, the automated sensitivity analysis is not well rooted in statistical theory.

Which alternative assignment rule could be implemented? If there are doubts whether an attribute affects prices and therefore the delineation of products, an automated hedonic regression analysis appears attractive. However, if such procedures consume too much time and effort, the automated sensitivity analysis of the C-approach is certainly an option.

In sum, avoiding the assignment bias of the two stage aggregation requires time and effort on the side of the analyst. With a single stage aggregation, this time and effort could be redirected to the reduction of assortment bias. Whether this reduction is sufficiently large, is an open question that can be answered only by practical experience.

Putting everything together

Recommendation 1 can be implemented even without Recommendation 2. Of course, for segments with a stable set of items a joint implementation of both recommendations appears preferable. Suppose that the average price change of some segment *s* between the base month $t = 0$ and some comparison month $t = T$ must be computed. If the assortment bias can be kept on a low level, the average price change of segment *s* can be calculated in a single stage aggregation directly from the individual items *j*. This can be accomplished by computing the following two quality adjusted unit values:

$$(15) \qquad P^0 = \frac{\sum_{j \in A_s^0} p_j^0 q_j^0}{\sum_{j \in A_s^0} q_j^0 z_j} \quad \text{and} \quad P^T = \frac{\sum_{j \in A_s^T} p_j^T q_j^T}{\sum_{j \in A_s^T} q_j^T z_j} .$$

In both unit values the same quality adjustment factors, z_i, are to be used. They can be computed by a simple formula like

$$(16) \qquad z_i = \frac{\sum_{t=0}^{T} p_i^t q_i^t}{\sum_{t=0}^{T} q_i^t} .$$

This formula utilizes all available information. Finally, the ratio

$$(17) \qquad P_{aLehr(T)}^{T/0} = \frac{P_{(T)}^T}{P_{(T)}^0}$$

gives the average price change of segment *s* between months $t = 0$ and $t = T$. Formulae (15) to (17) represent the augmented Lehr index.

When the base period ($t = 0$) is December 2014, then, at the end of January 2015, the above procedure is executed for $T = $ Jan 2015 and the resulting index number is officially published. At the end of February 2015, the procedure is repeated for $T = $ Feb 2015, and so on. The last repetition of the procedure is at the end of December 2015. This yields a sequence of 12 index numbers. No revisions are required. These 12 index numbers would replace the real time index of the C-approach.

When segment *s* is characterized by high product turnover and continuous innovations (e.g. mobile phones), the two stage aggregation of the C-approach becomes attractive. The single stage aggregation is still possible, though. The computation of the quality adjustment factors, z_i, in formula (16) could use estimated prices, $\hat{p}_i^t$, instead of observed prices, p_i^t. The estimated prices, $\hat{p}_i^t$, can be obtained from hedonic regression analysis or other reliable procedures. This were a particularly appealing procedure if the hedonic regression analysis could be partly automated.

5. Conclusions

The C-approach proposed by Chessa (2016a) is a significant step towards a more coherent and reliable processing of scanner data. The approach uses a two stage aggregation procedure in which individual items are bundled into products and then products into segments. On the lower stage, for each product and each month, a unit value and an aggregate quantity is computed. These product data are used in the upper stage, where for each month an average price level is computed. The ratios of these average price levels yield the price index numbers. The computations on the upper stage utilize the Geary-Khamis index.

It is a major virtue of the C-approach that it utilizes all available information. For this purpose it proposes a rather complex two stage procedure. The present paper has argued that in some situations the objectives of the C-approach can be achieved by somewhat simpler means.

First, the Geary-Khamis index could be replaced by a simpler procedure. Two options are the Gerardi index and the augmented Lehr index. A third promising option, the 'revisionless QU-GK method', has been proposed by Chessa (2016b) himself. Preliminary empirical investigations show that this index hardly deviates from the benchmark index of the C-approach.

Furthermore, the present paper recommends to replace the sensitivity analysis executed at the lower stage of aggregation, by a method that is deeper rooted in statistical theory (e.g. hedonic regression). Finally, this paper asks whether the first stage of the two stage procedure is really necessary. Possibly, the problem of exiting and entering items can be tackled also within a single stage procedure that redirects resources from the first stage of aggregation to the detection and qualitative evaluation of entering and exiting items. Such a move would avoid the C-approach's assignment bias and, if carefully implemented, may avoid a significant increase in assortment bias. Whether this is a viable alternative remains to be seen.

Acknowledgements

This paper originated from a consultancy project for the European Central Bank. The author wants to thank the participants of a workshop at Eurostat (Luxembourg, December 2016) and in particular Tony Chessa and Claude Lamboray for their helpful input. Furthermore, the comments of three referees are gratefully acknowledged.

References

Auer, L. von (2009), 'The Measurement of Macroeconomic Price Level Changes', Universität Trier: *Research Papers in Economics*, No. 1/09, Universität Trier.

Auer, L. von (2012), 'Räumliche Preisvergleiche: Aggregationskonzepte und Forschungsperspektiven', *AStA Wirtschafts- und Sozialstatistisches Archiv*, Vol. 6, issue 1, pp. 27–56.

Auer, L. von (2014), 'The Generalized Unit Value Index', *Review of Income and Wealth*, Vol. 60, issue 4, pp. 843–861.

Australian Bureau of Statistics (2016), *Making Greater Use of Transactions Data to Compile the Consumer Price Index*, Information Paper 6401.0.003, Canberra.

Balk, B.M. (1981), 'A Simple Method for Constructing Price Indices for Seasonal Commodities', *Statistical Papers,* Vol. 22, pp. 72–78.

Balk, B.M. (2008), *Price and Quantity Index Numbers*, Cambridge University Press, Cambridge, New York.

Chessa, A.G. (2016a), 'A New Methodology For Processing Scanner Data in the Dutch CPI', *Eurostat Review on National Accounts and Macroeconomic Indicators*, 1/2016, pp. 49–69.

Chessa, A.G. (2016b), 'Comparisons of the QU-Method With Other Index Methods for Scanner Data', paper prepared for a workshop at Eurostat, 7-8 December 2016, Luxembourg.

Chessa, A.G., J. Verburg and L. Willenborg (2017), 'A Comparison of Price Index Methods for Scanner Data', paper prepared for the 15th Meeting of the Ottawa Group, 10-12 May 2017, Eltville.

Dalén, J. (2001), 'Statistical Targets for Price Indexes in Dynamic Universes', Proceedings of the Sixth Meeting of the International Working Group on Price Indices, Canberra, Australia.

Davies, G.R. (1924), 'The Problem of a Standard Index Number Formula', *Journal of the American Statistical Association*, Vol. 19, pp. 180–188.

de Haan, J. (2002), 'Generalised Fisher Price Indexes and the Use of Scanner Data in the Consumer Price Index (CPI)', *Journal of Official Statistics*, Vol. 18, issue 1, pp. 61–85.

de Haan, J. (2004), 'Estimating Quality-Adjusted Unit Value Indexes: Evidence from Scanner Data', Paper presented at the SSHRC International Conference on Index Theory and the Measurement of Prices and Productivity, Vancouver, Canada.

de Haan, J. and H.A. van der Grient (2011), 'Eliminating Chain Drift in Price Indices Based On Scanner Data', *Journal of Econometrics*, Vol. 161, issue 1, pp. 36–46.

Diewert, E.W. and K. J. Fox (2017), 'Substitution Bias in Multilateral Methods for CPI Construction Using Scanner Data', Discussion Paper 17-02, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada.

Drobisch, M.W. (1871a), 'Ueber Mittelgrössen und die Anwendbarkeit derselben auf die Berechnung des Steigens und Sinkens des Geldwerths', *Berichte der mathematisch-physicalischen Classe der Königlich Sächsischen Gesellschaft der Wissenschaften*, Heft 1.

Drobisch, M.W. (1871b), 'Ueber die Berechnung der Veränderungen der Waarenpreise und des Geldwerths', *Jahrbücher für Nationalökonomie und Statistik*, Vol. 16, pp. 143–156.

Éltető, Ö. and P. Köves (1964), 'On an Index Computation Problem in International Comparisons' (in Hungarian), *Statiztikai Szemle*, Vol. 42, pp. 507–518.

Eurostat (1978), 'Comparison in Real Values of the Aggregates of ESA-1975', Statistical Office of the European Communities: Luxembourg.

Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton Miin, 3rd. revised ed. 1927, newly printed 1997 by Pickering and Chatto, London.

Geary, R.C. (1958), 'A Note On the Comparison of Exchange Rates and Purchasing Power Between Countries', *Journal of the Royal Statistical Society*, Series A, Vol. 121, issue 1, pp. 97–99.

Gini, C. (1924), 'Quelques Considérations au Sujet de la Construction des Nombres Indices des Prix et des Questions Analogues', *Metron*, Vol. 4, issue 1, pp. 3–162.

Ivancic, L., W.E. Diewert and K. J. Fox (2011), 'Scanner Data, Time Aggregation and the Construction of Price Indices', *Journal of Econometrics*, Vol. 161, issue 1, pp. 24–35.

Khamis, H. (1972), 'A New System of Index Numbers for National and International Purposes', *Journal of the Royal Statistical Society*, Series A, Vol. 135, issue 1, pp. 96–121.

Lamboray, C. (2017), 'The Geary Khamis Index and the Lehr Index: How Much Do They Differ?', paper prepared for the 15th Meeting of the Ottawa Group, 10-12 May 2017, Eltville.

Lehr, J. (1885), *Beiträge zur Statistik der Preise insbesondere des Geldes und des Holzes*, F. D. Sauerländer Verlag, Frankfurt a.M.

Maddison, A. and D.S.P. Rao (1996), 'A Generalized Approach to International Comparison of Agricultural Output and Productivity', Research Memorandum GD-27, Groningen: Groningen Growth and Development Centre.

Segnitz, E. (1870), 'Ueber die Berechnung der sogenannten Mittel, sowie deren Anwendung in der Statistik und anderen Erfahrungswissenschaften', *Jahrbücher für Nationalökonomie und Statistik*, Vol. 14, pp. 183–195.

Silver, M. (2010), 'The Wrong and Rights of Unit Value Indices', *Review of Income and Wealth*, Vol. 56, pp. S206–pp. S223.

Szulc, B. (1964), 'Index Numbers of Multilateral Regional Comparisons' (in Polish), *Przeglad Statystyczny*, 3, pp. 239–254.

4 Big data types for macroeconomic nowcasting

DARIO BUONO [1], GIAN LUIGI MAZZI [2], GEORGE KAPETANIOS [3], MASSIMILIANO MARCELLINO [4] AND FOTIS PAPAILIAS [5]

Abstract: In this paper we present a detailed discussion on various types of big data which can be useful in macroeconomic nowcasting. In particular, we review the big data sources, availability, specific characteristics and their use in the literature. We conclude this paper identifying the big data types which could be adopted for real applications.

JEL codes: C82, E27, E37, E47

Keywords: big data, nowcasting, data features

(1) Eurostat.
(2) retired from Eurostat.
(3) King's College London.
(4) Università Bocconi, Italy.
(5) Queen's University Management School, Queen's University Belfast, UK and quantf research.

1. Introduction

The advancements in computer technology during the last decades have allowed the storage, organisation, manipulation and analysis of vast amount of data from different sources and across different disciplines, and there is nowadays an ever growing interest in the analysis of these big data. In this paper, we summarize the results of an investigation of the specificities of various big data sources relevant for macroeconomic nowcasting and early estimates, and we discuss the characteristics of big data sources that are relevant for their application in this context.

Perhaps not surprisingly given their different origins, the literature provides various definitions of big data. One possibility to obtain a general classification is to adopt the '4 Vs' classification, originated by the IBM, which relates to: (i) Volume (Scale of data), (ii) Velocity (Analysis of streaming data), (iii) Variety (Different forms of data) and (iv) Veracity (Uncertainty of data). However, this classification seems too general to guide empirical nowcasting applications.

A second option is to focus on numerical data only, which can either be the original big data or the result of a transformation of unstructured data. Once data have been transformed, and following, e.g. Doornik and Hendry (2015), we can distinguish three main types of big data. 'Tall' datasets include not so many variables, N, but many observations, T, with $T >> N$. This is for example the case with tick by tick data on selected financial transactions or search queries. 'Fat' datasets have instead many variables, but not so many observations, $N >> T$. Large cross-sectional databases fall into this category, which is not so interesting from an economic nowcasting point of view, unless either T is also large enough or the variables are homogeneous enough to allow proper model estimation (e.g., by means of panel methods) and nowcast evaluation. Finally, 'Huge' datasets, with very large N and T, are the most interesting type of data in a nowcasting context even if, unfortunately, they are not so often available for economic variables. However, 'complexity' can help in expanding the N dimension, for example allowing for dynamics (lead-lag relations), non linearities and the micro-structure of information (calendar issues for example).

A third possibility to classify big data is to resort to some sort of official definition. A particularly useful taxonomy is provided by the statistics division of the United Nations Economic Commission for Europe (UNECE), which identifies big data according to their source.

A traditional data source, revamped by the IT developments, is represented by *Business Systems* that record and monitor events of interest, such as registering a customer, manufacturing a product, taking an order, etc. The process-mediated data thus collected by either private businesses (commercial transactions, banking/stock records, e-commerce, credit cards, etc.) or public institutions (medical records, social insurance, school records, administrative data, etc.) is highly structured and includes transactions, reference tables and relationships, as well as the metadata that sets its context. Traditional business data is the vast majority of what IT managed and processed, in both operational and BI systems, usually after structuring and storing it in relational database systems.

A novel data source is represented by *Social Networks* (human-sourced information). This information is the record of human experiences, by now almost entirely digitally stored in personal computers or social networks. Data, typically, loosely structured and often ungoverned, include those saved in proper Social Networks (such as Facebook, Twitter, Tumblr etc.), in blogs and comments, in specialized websites for pictures (Instagram, Flickr, Picasa etc.)

or videos (Youtube, etc.) or internet searches (Google, Bing, etc.), but also text messages, user-generated maps, e-mails, etc.

Yet another source of big data, and perhaps the fastest expanding one, is the so-called *Internet of Things*. Machine-generated data are derived from sensors and machines used to measure and record the events and situations in the physical world. It is becoming an increasingly important component of the information stored and processed by many businesses. Its well-structured nature is suitable for computer processing, but its size and speed is beyond traditional approaches. Examples include data from sensors, such as fixed sensors (home automation, weather/pollution sensors, traffic sensors/webcam, etc.) or mobile sensors (mobile phones, connected cars, satellite images, etc.) but also data from computer systems (logs, web logs, etc.).

The resulting many types of big data have been already exploited in many scientific fields such as climatology, oceanography, biology, medicine, and applied physics. Specific areas of economics have also seen a major interest in big data and business analytics, in particular marketing and finance. Instead, in conventional macroeconomics there have so far been limited applications, mostly concentrated in the areas of nowcasting/forecasting, on which we also focus.

A first aim of this paper is to review the main existing macroeconomic applications and, in particular, to provide a discussion of various big data types which could be useful in macroeconomic nowcasting. A survey of the econometric methods for big data is instead left for future research.

We have identified ten categories of big data which are organised as: (i) financial markets data, (ii) electronic payments data, (iii) mobile phones data, (iv) sensor data, (v) satellite images data, (vi) scanner prices data, (vii) online prices data, (viii) online search data, (ix) textual data, and (x) social media data. For each of these categories, in Section 2, we analyse the specific characteristics of the data, review the relevant macroeconomic papers based on them, and provide details regarding data sources and availability.

Next, in Section 3, we consider the dual problem: nowcasting a specific macroeconomic variable of interest, such as gross domestic product (GDP) growth, inflation or unemployment, using big data. A summary table, reported at the end of the Section, lists the main studies, providing details on the specific type of big data adopted.

In Section 4, we propose a further classification, specifically designed for numerical big data and based on their relative cross-sectional and temporal dimensions, which has implications for the proper data pre-treatment and required econometric techniques to analyse them. As mentioned above, big data are now split into three main categories: Fat, Tall, and Huge.

The final goal of the paper is to use the extensive review of existing studies and data to provide a description of the main features of big data that make them useful for nowcasting and flash estimation, and an indication of which big data categories seem particularly promising for nowcasting specific variables and for further analysis. We deal with these issues in Section 5.

Section 6 summarizes the main findings and proposes a summary classification of big data, based on a set of key features that emerge from the detailed review: source, provider, availability, continuity, type, size and sample, meta data, feature, frequency, pre-treatment, link with target, previous use, and required econometrics. This classification is finally used to create a grid for assessing the need and potential usefulness of big data for macroeconomic nowcasting and forecasting, in general and in official institutions such as Eurostat, according to the outcome of a careful cost-benefit analysis.

2. Types of big data for macroeconomic nowcasting

This section is divided into subsections, each corresponding to a big data category which could be useful for macroeconomic nowcasting. To facilitate the reading process, and in order to provide a typology of big data, we have standardised the format of each subsection as follows. At first we describe the big data and discuss its characteristics. Then, we review the relevant literature which currently uses the specific type of big data in nowcasting/forecasting or other applications in economics. Finally, we present a detailed list of data sources along with their availability.

2.1. Financial markets data

Advances in computer technology and data storage has allowed for the collection and analysis of high-frequency financial data. The most widely observed forms of financial big data are the trades and quotes. New York Stock Exchange (NYSE) initiated the collection of this data in 1992. This intraday data potentially provides detailed information which could be used in the analysis of markets efficiency, volatility, liquidity as well as price discovery and expectations. Central banks monitor activity across all financial markets and nowadays high-frequency financial data includes:

- Equities trades and quotes for all types of investors;
- Fixed-Income trades and quotes for all types of investors;
- Foreign Exchange trades and quotes for all types of investors;
- NXL and OTC Derivatives and option transactions;
- and generally all operations in financial markets.

The easiest form of the above data to obtain is intraday trades and quotes. However, this data is available for each security individually; therefore if we are interested in the construction of time series index based on intraday data, we would have to consider a large number of securities. Consequently, this increases the total cost for this data. Furthermore, data which refers to the monitoring of all market operations is very sensitive and only central banks and other regulatory bodies have access to, therefore most studies in the literature rely on anonymised data obtained by various third-party providers. Although financial big data is very important in the analysis of market microstructure, its use in macroeconomic nowcasting and forecasting is mainly on daily or weekly frequency. We consider financial data as a major data source, therefore it forms the basis of our explanatory data used in the empirical application in future. Our main aim is to examine whether one of the other nine types of big data can improve the macroeconomic nowcasting and result in improved estimates.

2.1.1. APPLICATIONS

Financial data in high-frequency form has been the main element in volatility and market microstructure studies. Moreover, the use of such data in macroeconomic forecasting, and subsequently in nowcasting, has been included in many studies, either after aggregating the data to the same frequency as the macroeconomic targets or, more recently, using mixed frequency models. Given that our focus in this project is in alternative big data types, we refer the reader to Stock and Watson (2002a), Stock and Watson (2002b), Giannone, Reichlin and Small (2008), Angelini, Camba-Mendez, Giannone and Reichlin (2011), Banbura, Giannone and Reichlin (2011), Banbura and Runstler (2011), Modugno (2013), Andreou, Ghysels and Kourtellos (2015).

2.1.2. INDICATIVE DATA SOURCES & EXAMPLES

Although most of the historical high frequency data is proprietary, end of day transaction prices (along with the daily open, high and low) are publicly available for most equities in major stock exchanges. Below, we provide an indicative list of data sources as well as live feeds which could be used to collect intraday data:

- Daily equity prices: Google Finance, Yahoo Finance, St. Louis FRED for major US indexes, etc. are some free online sources;
- Historical Intraday data: Kibot, PiTrading and Intradata provide proprietary historical intraday data for assets in major markets. Unstructured intraday data could be converted to structured time series format by constructing daily realised volatility estimators which could be included in the nowcasting exercise as risk indicators;
- Live feeds: Live-Rates.com, knoema and Oanda provide live feeds for most of the financial instruments for free. These feeds could be used as a source for a data collection which could then be used in an aggregated way in nowcasting.

2.1.3. CONCLUSIONS

Financial transactions data is a very important source of information for the analysis of markets as well as the forecasting of the economy. Given their high-frequency nature, financial markets illustrate a very fast discount of news, much faster than lower frequency macroeconomic indicators. Therefore, financial markets data is an essential source of information in macroeconomic nowcasting. We will investigate whether such data can be combined with another source of big data to provide more accurate nowcasts and short-term forecasts. Overall, if the researcher is interested in risk sentiment indexes or the construction of policy uncertainty indexes, realised volatility estimators might prove a more accurate and timely instrument compared to the standard measures, say VIX in the US.

2.2. Electronic payments data

The term electronic payments is broad and considers all kinds of electronic funds transfer. In particular, forms of electronic payments include: (i) cardholder-initiated transactions, (ii) direct deposit payments initiated by the payer, (iii) direct debit payments initiated by businesses which debit the consumer's account for the purchase of goods or services, (iv) credit transfers, (v) electronic bill payments in online stores, among others. The most heavily used form of electronic payments is the cardholder-initiated transactions, i.e. credit and debit card payments.

According to the Capgemini and BNP Paribas (2016) report, cards dominate the global non-cash market, accounting for 65 % of all non-cash transactions. In Table 1 we report the credit card usage in 2014 across six regions. Credit transfers follow with 17 % share and direct debits with 12 %. Finally, checks account for 6 % globally. China, Hong Kong, India and other Asian markets rank first in the cards usage with 84 % share in the non-cash market. Second follow the CEMEA markets with cards usage at 77 % of the non-cash transactions. Japan, Australia, Singapore and

South Korea markets follow third with 75 % cards share and North America ranks fourth with 71 % share. Latin America and Europe are last with 49 % and 47 % share of cards in the non-cash market. Even in the markets with the lowest usage, cards share is at least double of the other non-cash alternatives. For example, direct debits and credit transfers account for 23 % and 26 % in Europe compared to 47 % of cards. The above statistics highlight that cards is the main form of non-cash payments. Card payments include online as well as offline POS (6) purchases making them very useful in the tracking of consumer behaviour and retail sales (among others).

Table 1: Non-cash payments mix
(%)

	Europe	North America	JASS	CHI	Latin America	CEMEA	Global
Cards	47	71	75	84	49	77	65
Credit Transfers	26	8	17	10	32	20	17
Direct Debits	23	11	7	2	15	3	12
Checks	4	11	1	5	4	0	6

Source: Capgemini and BNP Paribas (2016). JASS: Japan, Australia, Singapore and South Korea. CHI: China, Hong Kong, India and other Asian markets. CEMEA: Poland, Russia, Saudi Arabia, South Africa, Turkey, Ukraine, Hungary, Czech Republic, Romania and other Central European and Middle Eastern markets.

The cards payments are considered as a category of big data because of high frequency of transactions. We have thousands of transactions throughout the day and, with the huge increase of e-commerce, also during the night. One specific characteristic of cards data is the weekly pattern in daily aggregated data (or intraday pattern in non-aggregated data). As indicated by the literature, consumers tend to purchase more goods and services towards the end of the week. Cards data usually is offered aggregated in order to ensure protection of personal details.

2.2.1. APPLICATIONS

The literature in economics which uses credit cards data started very recently. Galbraith and Tkacz (2007) is one of the first papers which published the results of cards data in macroeconomics. In particular, they use Canadian debit card transactions (7) in order to provide real-time estimates of economic activity. Their predictive regression analysis provides information for consumer behaviour as well as improved nowcast estimates. At first they find that household transactions have a weekly pattern (on average), peaking every Friday and falling every Sunday. The high-frequency analysis of electronic transactions around extreme events explains expenditure patterns around the September 11 terrorist attacks and the August 2003 electrical blackout. Finally, consensus forecast errors for GDP and consumption (especially non-durable) growth can be partly explained by cards data.

Carlsen and Storgaard (2010) use Dankort payments in order to nowcast the retail sales index in Denmark. Dankort is a debit card developed jointly by the Danish banks and introduced in 1983. The Dankort is free of charge to the customers, and the card is extensively used by households. This fact makes Dankort an ideal instrument for tracking household activity and thus, retail sales. Another advantage of using Dankort is the timing of publication. Dankort data is available one week after the reference month, whereas the retail sales index is published three weeks later. As mentioned in the previous studies, seasonal effects are also present which need extra care in order to end up with a clean dataset. The out-of-sample nowcast exercise is in favour

(6) Point of Sale (POS). For example, card payment at a retail store.
(7) Obtained via the Canadian interbank network, Interac.

of the two models which use cards data, however the evaluation period is again too narrow: monthly nowcasts between January, 2007 and May, 2008.

Galbraith and Tkacz (2015) tackle directly with the issue of nowcasting Canadian GDP growth using Canadian credit and debit cards transactions as well as checks. They find that, among the payments data, debit card transactions seem to produce the most improved estimates. The issue of seasonality is also present here. The authors suggest the use of X-11 methodology ([8]) in order to clean the data. Their main finding is that nowcasting using high frequency electronic payments improve by 65 % between the first and final estimates presenting supporting evidence in the use of electronic payments data.

Duarte, Rodrigues and Rua (2016) use ATM and POS high frequency data for nowcasting and forecasting quarterly private consumption for Portugal. Their ATM data is provided by Multibanco, which is the Portuguese ATM and POS network. Their methodology is based on Mixed Data Sampling (MIDAS) models and builds on the earlier work by Esteves (2009) confirming that the use of electronic payments data improves nowcasting and forecasting accuracy. Weekly payment data produce particularly good results, while daily data are too noisy.

Barnett, Chauvet, Leiva-Leon and Su (2016) derive an indicator-optimized augmented aggregator function over monetary and credit card services using credit card transaction volumes. This new indicator, inserted in a multivariate state space model, produces more accurate nowcasting of the GDP compared to a benchmark model.

Finally, Aprigliano, Ardizzi and Monteforte (2016) use a mixed frequency dynamic factor model to predict the Italian GDP growth using standard business cycle indicators (such as electricity consumption, industrial production, inflation, stock market indexes, manufacturing indexes, etc.) as well as payment systems data (cheques, credit transfers, direct debits, payment cards). They find that monthly payment data helps in tracking the economic cycle in Italy and improves nowcasting. In a separate screening exercise using the Least Absolute Shrinkage and Selection Operator (LASSO), payment system variables are indicated as potential predictors of GDP growth.

2.2.2. INDICATIVE DATA SOURCES & EXAMPLES

Based on the literature review and an extensive online search, we discuss potential data sources as well as availability for cards data.

- At first we have all credit and debit card financial services corporations. These companies facilitate electronic funds transfers throughout the world. We briefly mention them here.
  - Visa, Inc. It is not known if VISA provides aggregated data to third-parties,
  - American Express. American Express is known to sell anonymised data to third-party marketing companies ([9]). American Express also offers anonymised data to their business partners as a way to help them analyse and promote their products ([10]),
  - Mastercard. Mastercard, along with American Express, is known to sell anonymised data to third-party marketing companies ([11]);

([8]) See Kapetanios, Marcellino and Papailias (2016) for a full discussion.
([9]) See online news articles at https://goo.gl/Co3oc6 and https://goo.gl/yrj74b.
([10]) See online news article at https://goo.gl/mS9J0H.
([11]) See online news article at https://goo.gl/Co3oc6 .

- Another way to acquire electronic payments data is through interbank networks. Below we summarise the interbank networks for major economies.
  - Australia: Electronic Funds Transfer at Point Of Sale (EFTPOS). Data is not publicly available. In 2010, ANZ BANK has announced the launch of an online tool that uses aggregated data from merchant EFTPOS transactions to illustrate estimated sales patterns, market share, turnover and to provide insights into customer behaviour. However, this is available to ANZ business customers only,
  - Canada: Interac Association (as mentioned in the papers above). Not publicly available. However, given the existence of current literature, data could be potentially purchased,
  - China: China Union Pay. Not publicly available and unknown if China Union Pay have an interest in selling anonymised data,
  - France: Groupement des Cartes Bancaires CB. Payment system: EBA Clearing (Euro1) and Trans-European Automated Real-time Gross Settlement Express Transfer System (TARGET2). Not publicly available. However, given the existence of current literature, data could be potentially purchased or come to an agreement,
  - Germany: Girocard. Euro1 and TARGET2. Not publicly available. However, given the existence of current literature, data could be potentially purchased or come to an agreement,
  - Italy: BI-COMP. Euro1 and TARGET2. Not publicly available. However, given the existence of current literature, data could be potentially purchased or come to an agreement,
  - Japan: Yucho. Not publicly available. It is not known if Yucho would be interested in selling anonymised data,
  - Portugal: Multibanco (as mentioned in the papers above). Not publicly available. However, given the existence of current literature, data could be potentially purchased or come to an agreement;
- Third-party data providers, such as Envestnet Yodlee ([12]), could also provide aggregated data on electronic transactions. In most of these cases, the data is proprietary;
- Finally, the last way to collect cards data would be via access to central bank databases which are not publicly available. All central banks could provide such data in aggregated or unstructured format. Some of the central banks which could provide useful data regarding this project are: the European Central Bank (ECB), Banque de France, Deutsche Bundesbank, Banca d'Italia, De Nederlandsche Bank, Banco de Portugal and Banco de Espana. However, Central Banks do not usually provide any data to third-parties.

Some data which is publicly available and might be useful for demonstration purposes, but not for an empirical analysis due to the short sample, include:

(1) UK Corporate credit card transactions 2014-2015 in monthly frequency. This is a structured series based on all transactions on corporate credit cards in the financial year 2014-2015;

(2) Sample dataset of European credit card transactions during two days in September 2013 of Dal Pozzolo, Caelen, Johnson, and Bontempi (2015). The data is publicly available on Kaggle website ([13]). Due to confidentiality issues, no original features are presented apart from the 'Time' of the purchase, the 'Amount' and a flag which indicates if the transaction is genuine

([12]) https://goo.gl/eGaIoU.
([13]) https://goo.gl/G1saSD.

or fraudulent. In total, there are 284,808 transactions. This dataset might be useful in order to illustrate the daily aggregation process, however it is not sufficient to show the weekly seasonal pattern and its subsequent removal and data cleaning.

2.2.3. CONCLUSIONS

Electronic payments include credit and debit card transactions, credit transfers, direct debits, cheques, etc. Based on the nature of the data, it tracks well economic activity and particularly household purchases (via credit and debit card transactions). For this reason it would be very useful in the monitoring, nowcasting and forecasting of retail sales, private consumption, and other related variables. The literature provides empirical evidence in favour of cards data. However, unstructured or aggregated data is not publicly available and most of these studies are provided with confidential data from interbank networks or central banks.

2.3. Mobile phones data

With the introduction of mobile phones nearly thirty years ago, scientists across various fields were positive that mobile phones usage and information would be an important tool for statistics. The data collection from basic functions of a mobile phone, i.e. receiving and making phone calls and short text messages, already provides enough information about population density, location, economic development of particular geographic areas and use of public transport among others. The rapid development and growth of mobile phones technology during the past twenty years allows for even more specific data collection as internet activity, mobile banking, GPS tracking and other sensors data ([14]). Overall, mobile phones data provides detailed information of human behaviour and therefore can be useful in social sciences too.

The Deloitte (2012) report which uses data from the Cisco's VNI Index for 14 countries states that a doubling of mobile data use leads to a 0.5 % increase in the GDP per capita growth rate. Given the heavy use of mobile phones, the collected data is characterised as big data due to the massive volume. News coverage also provides evidence that mobile data is promising for the future ([15]).

2.3.1. APPLICATIONS

As in the case of card payments data, the literature which uses mobile phones data is also very recent. Smith-Clarke, Mashhadi and Capra (2014) employ call data in order to examine poverty in two developing countries where survey data is scarce. In particular, they use the aggregated call detail records of mobile phone subscribers and extract features that are strongly correlated with poverty indexes derived from census data. Their data consists of calls between 5 million Orange customers from Cote d' Ivoire and 928,000 calls between customers of an undisclosed network in an undisclosed region. The authors highlight the difficulty of obtaining call details records for other developed and emerging countries.

Deville, Linard, Martin, Gilbert, Stevens, Gaughan, Blondel and Tatem (2014) use mobile phone data for population mapping. They use more than 1 billion mobile phone calls from Portugal and France and show how spatially and temporarily explicit estimations of population densities can be produced at national scales. In the same topic also lies the work of Ricciato, Widhalm, Craglia and Pantisano (2015) who estimate population density distribution from network-based

([14]) In that sense, mobile phones data can also be part of sensors data.
([15]) See https://goo.gl/MYb21Q, https://goo.gl/e35lQQ and https://goo.gl/Dkdara.

mobile phone data in Europe. In the same context, De Meersman, Seynaeve, Debusschere, Lusyne, Dewitte, Baeyens, Wirthmann, Demunter, Reis and Reuter (2016) assess the quality of mobile phones data to estimate the actual population. They use Belgian mobile phone data from the major network operator, Proximus, and their results are consistent with the 2011 Census.

Mao, Shuai, Ahn and Bollen (2015) also investigate the case of Cote d' Ivoire to analyse the relations between its nation-wide communications network and the socio-economic dynamics of its regional economies. They introduce an indicator to quantify the relative importance of an area on the basis of call records, and show that a region's ratio of in- and out-going calls can predict its income level. Their results demonstrate the potential of mobile communication data to monitor the economic development and social dynamics of low-income developing countries in the absence of extensive econometric and social data.

2.3.2. INDICATIVE DATA SOURCES & EXAMPLES

Below we briefly mention some data sources as well as examples for demonstration.

- Obviously, the network operators are the main mobile phone data source as all information is sent and received through their network. However, the data is not publicly available and, as mentioned in the literature, this is due to regulatory limitations. It is not known if a private network operator company would be interested in selling anonymised data.
- Third-party software developers, such as HelloSpy ([16]), which offer their customers the ability to track their call history. This is proprietary data.

2.3.3. CONCLUSIONS

Mobile phone data also seems to be an interesting source. It could be used in order to provide timely estimates for population and census data, moreover, as we see in Toole *et al* (2015), the analysis of phone calls can provide details about behaviour which could be used for unemployment and economic activity forecasting. But, generally speaking, it seems difficult to find a detailed and long enough dataset to be tested in a nowcasting/forecasting application.

2.4. Sensor data and the internet of things

Sensor data mainly refers to any kind of output of a device which detects and responds to input sources from the physical environment. One of the oldest examples is the temperature monitoring via climate sensors. Sensors have been long used in manufacturing of plants, cars, ships, military equipment and sensor data has been used by operational technology engineers for quite many years. Information technology and the rapid development of internet greatly facilitate the collection and distribution of sensor data and gives rise to the so-called 'Internet of Things (IoT)'. A 'thing' includes any item which can be attached with sensors. Internet access allows the 'thing' to automatically transmit the data via networks and store it to public clouds and databases. This, in-turn, provides easy access to sensor data for mining and analytics purposes.

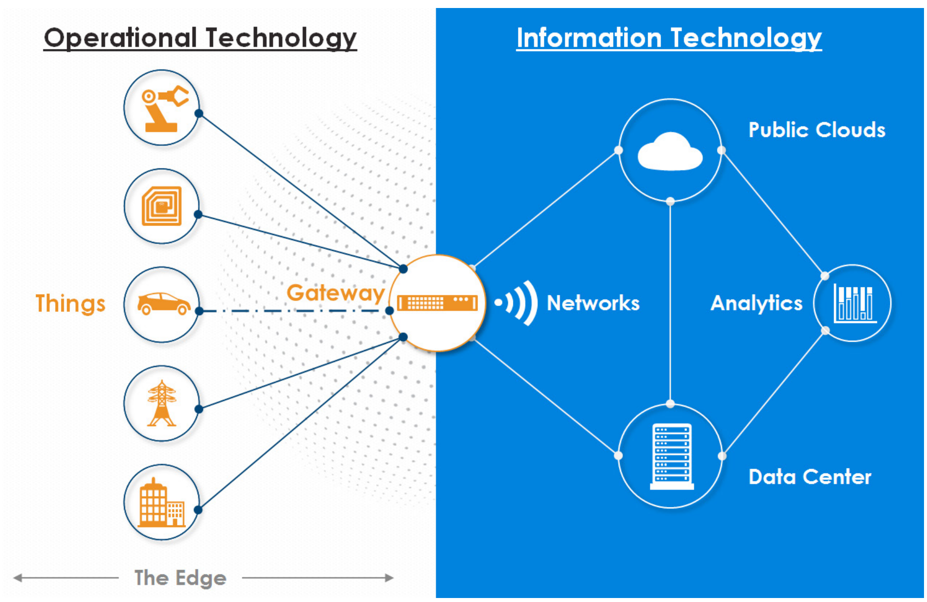
Mobile things, e.g. smart phones, computers, tablets, watches, home appliances, cars, drones, etc., as well as stationary things, e.g. wind turbines, temperature sensors, etc., all come attached with sensors which can send and receive information online. In the this category, we include

([16]) https://goo.gl/iGvLJM.

datasets which are collected from sensors and other devices which do not track directly human activity as opposed to mobile phones data where all information is returned by human actions. According to Ebner (2015) ([17]), Gartner, a leading information technology research and advisory company, forecasts 25 billion connected things by 2020. The data generated by these sensors could impact almost every major industry, including healthcare, transportation, and energy. Cisco estimates that IoT could have an economic impact of $14 trillion by the early 2020s.

Apple and Samsung are already developing platforms which gather data from the watches and other instrumented objects of the IoT. This data is then made available to developers for the creation of new applications and analyses ([18]). A recently start-up company, ThingSpeak ([19]), provides an open-source platform which allows the user to have his things transmit data to their servers and then offers his data via an API. Mathworks and MATLAB are integrated in the platform which can be used directly for analysis and visualisation of the data.

Figure 1: The internet of things



Source: Graham (2016).

2.4.1. APPLICATIONS

Unfortunately, it seems that there does not exist a literature on economic forecasting or nowcasting based on sensor data. We believe that this is due to the unavailability of data. Most of the literature is based on weather forecasting, where ample data exists, traffic and other geo-related applications. Some indicative sources include Jones (2010), Bazzani, Giovannini, Galotti and Rambaldi (2011), Chan, Khadem, Dillon, Palade, Singh and Cheng (2012), Mass (2012) and Fernandez-Ares, Mora, Arenas, de las Cuevas, Garcia-Sanchez, Romero, Rivas, Castillo and Merelo (2016). Also, Suryadevara, Mukhopadhyay, Wang and Rayudu (2013) forecast human

([17]) https://goo.gl/9h3Zf8.
([18]) See https://goo.gl/9h3Zf8.
([19]) https://goo.gl/z4y9by.

behaviour using wireless sensors data in a smart home. For example, if this could be generalised for households in an economy, sensor data could be used for private consumption nowcasting/forecasting. Papadimitriou, Sun, Faloutos and Yu (2013) review dimensionality reduction and econometric techniques which could be applied in the analysis and filtering of time series sensor streams which could be used based on data availability.

However, IoT will be very useful in the future. According to Manyika, Chui, Bisson, Woetzel, Dobbs, Bughin, and Aharon (2015) IoT will mainly contribute to:

- Factories, e.g. operations management, predictive maintenance;
- Cities, e.g. public safety, traffic control, resource management;
- Human, e.g. improving wellness;
- Retail, e.g. self-checkout, layout optimisation, etc;
- Outside, e.g. logistics routing, autonomous (self-driving) vehicles ([20]);
- Homes, e.g. energy management, safety and security, chore automation;
- Offices, e.g. organisation redesign and worker monitoring, augmented reality for training.

In the future, IOT related data could become relevant also for nowcasting/forecasting.

For example, sensors could be used to monitor the number of people entering shops or of commercial vehicles moving along specific routes, or the intensity usage of machinery, which could be relevant for variables such as retail sales, exports, or IP.

2.4.2. INDICATIVE DATA SOURCES & EXAMPLES

Below we provide an indicative list of sources and examples. The majority is focused on temperature data and traffic control.

- Open Cities Project, https://goo.gl/XBnYuZ. Open Cities is a project co-founded by the European Union that aims to validate how to approach Open & User Driven Innovation methodologies to the Public Sector in a scenario of Future Internet Services for Smart Cities. Data is not publicly available but given that nature of the funding, data could be made available;
- Array of Things (AoT), https://goo.gl/qajel1. AoT is an urban sensing project, a network of interactive, modular sensor boxes that will be installed around Chicago to collect real-time data on the city's environment, infrastructure, and activity for research and public use. AoT will essentially serve as a 'fitness tracker' for the city, measuring factors that impact liveability in Chicago such as climate, air quality and noise. Visualised data is available online and downloadable nodes data will be available from the City of Chicago Data Portal in early 2017;
- Smart Santander, https://goo.gl/jw0rjU. The project envisions the deployment of 20,000 sensors in Belgrade, Guildford, Lübeck and Santander (12,000), exploiting a large variety of technologies;
- CityPulse Dataset Collection, https://goo.gl/9gBqxs. It includes road traffic data, weather data, cultural, social and library event data and parking data for Aarhus, Surrey and Brasov;

([20]) Amazon.co.uk already performed the first drone flight delivery in December 2016 in Cambridge.

- ThingSpeak, https://goo.gl/QXDQS3. Publicly available channels with sensor data and examples. They are supplied by private users, therefore their accuracy is not guaranteed. Most of the data is meteorological.

2.4.3. CONCLUSIONS

As discussed in this section, sensor data and IoT will have a great impact in various business sectors in the next five years or sooner. At this point, there does not exist a literature on sensor data in economic nowcasting and forecasting, however this might change if, for example, retail sales related data becomes available. Currently, there is a big interest in marketing and tailor-made services as well as meteorological applications where sufficient data already exists.

2.5. Satellite images data

Satellite imagery consists of images of Earth or other planets collected by satellites. The satellites are operated by governments and businesses around the world. Satellite images are licensed to government agencies and businesses such as Apple and Google. One of the first image satellites was launched in 1946 taking one picture every 1.5 seconds. At the end of August, 2015 it was estimated that there were 4,077 satellites orbiting the Earth ([21]), of course not all of them were imaging satellites.

Satellites images have many applications in meteorology, oceanography, agriculture, forestry, geology, intelligence, warfare and others. Recently, satellite imaging has attracted the interest of economists as well. Photos of homes with metal roofs can indicate transition from poverty, night lights can show economic growth and monitoring of factory trucks and deliveries can be used for industrial production nowcasting; See Florida (2014) and Kearns (2015) for more details.

Satellite image data presents the following features: (i) the use of high quality images and the frames taken per second make satellite image databases very big and cumbersome, and (ii) in some cases, as in city night lights, the data is slowly changing and, thus, not useful for nowcasting. Lowe (2014) provides a brief guide in satellite data handling which eases the use of this data.

2.5.1. APPLICATIONS

The literature has recently started including satellite image data in economic applications, mostly for measuring long-term variables such as poverty, inequality and long-term growth.

Elvidge, Sutton, Ghosh, Tuttle, Baugh, Bhaduri and Bright (2009) use population count and satellite observed lighting in order to build a global poverty map. They claim that this construction of poverty maps is more accurate as it should improve over time through the inclusion of new reference data as improvements are made in the satellite observation of human activities related to economics and technology access.

Henderson, Storeygard and Weil (2011) first put the night lights satellite images in the economic growth context. A scatterplot of the average annual percentage change in GDP and the average annual percentage change in lights indicates a strong and positive relationship. Also, the paper suggests that night lights could be even more useful in developing countries where there is lack of accurate official statistics. This claim is further confirmed in Chen and Nordhaus (2011) who also use luminosity as a proxy for economic output. They find that luminosity has

([21]) https://goo.gl/elqcWi.

informational value particularly for countries with low-quality statistical systems or no recent population or economic censuses. In their second paper on this topic, Henderson, Storeygard and Weil (2012) show that lights data can measure growth not only at country level but also for sub- and supranational regions. As an example, their empirical analysis shows that the coastal areas in sub-Saharan Africa are growing slower than the hinterland.

Mellander, Lobo, Stolarick and Matheson (2015) use night lights data for Sweden finding the correlation between luminosity data and economic activity is strong enough to make it a relatively good proxy for population and establishment density, but the correlation is weaker in relation to wages. Also, Keola, Andersson and Hall (2015) argue that nighttime lights alone may not explain value-added by agriculture and forestry, however by adding land cover data, the estimate of economic growth in administrative areas becomes more accurate.

Finally, Alesina, Michalopoulos and Papaioannou (2016) study how night lights can be used to explore the issue of ethnic inequality in a region. Donaldson and Storeygard (2016) is a very detailed review paper which discusses all the uses of satellite image data which include climate, weather, agricultural use, urban land use, buildings, natural resources and pollution monitoring.

2.5.2. INDICATIVE DATA SOURCES & EXAMPLES

As discussed above, the use of satellite image data in economics is very promising. However, its uses in macroeconomic nowcasting are more limited mainly due to the slowly changing nature of the underlying subjects. For example, monitoring the agricultural use could be used for long-term agricultural production value-added figures. Or, monitoring plants entry/exit traffic could provide some insights regarding long-term forecasts of industrial production. Following Donaldson and Storeygard (2016) we provide below a list of data sources.

- Landsat, https://goo.gl/Xhqya5. This dataset is publicly available and provides satellite images for urban land cover, beaches, forest cover, mineral deposits;
- MODIS, https://goo.gl/NhHU6x. This dataset is publicly available and provides satellite images for airborne pollution and fish abundance;
- NightLights, https://goo.gl/vdIksu. This is one of various luminosity datasets which is publicly available. It provides satellite images of electricity use;
- SRTM, https://goo.gl/6zKR4x. This dataset is publicly available and provides satellite images for elevation and terrain roughness;
- DigitalGlobe, https://goo.gl/0rL1nW. This dataset is not publicly available and provides satellite images for urban land cover and forestry. Data could be purchased;
- Publicly available datasets (including those mentioned above) are also provided by Google Earth ([22]).

2.5.3. CONCLUSIONS

Satellite image data is very promising in economics. Satellite images can supplement official statistics and can be very important in developing countries ([23]) where official figures are difficult to estimate. Luminous data from satellite images has been studied by the literature indicating that night lights are a good proxy for economic activity. Therefore, satellite images

([22]) Available at https://goo.gl/Xk0o16.
([23]) An EU project is concerned with similar tasks.

are very useful in economic activity and structural analysis, poverty or inequality forecasting, however, due to the slowly changing nature of the captured subjects might not be useful in macroeconomic nowcasting or short-term forecasting on a monthly or quarterly basis. However, on a local economy level, such as regional industry forecasting or consumption, this data could also be used in nowcasting as it would illustrate the activity of particular businesses during the night.

2.6. Scanner prices data

Scanner prices data consists of bar-code scanned data mainly provided by retailers. Prices could be scanned daily allowing for a high-frequency measurement of the supply side in the retail market. Price changes can reveal information in macroeconomic level as well as industry-specific and, more particular, retailer-specific level. Small price changes might be due to measurement error, the analyst who deals with this data must take this into account. Slowly changing prices, on average, often indicate upcoming changes in inflation, however extreme relative prices typically reflect the retailer's conditions rather than changes in average prices. Also, scanner data allow examining different regions inside an economy. Therefore, scanner prices data can be very useful in macroeconomic nowcasting, particularly for inflation subcomponents, such as food prices.

2.6.1. APPLICATIONS

Silver and Heravi (2001) is one of the first papers which associates inflation estimation with scanner prices data. Scanner data provides improved coverage compared to data collected by price collectors and the availability of up-to-date weights at a highly detailed level. This facilitates the creation of superlative price indexes which can incorporate substitution effects. Another use of scanner data in micro-level is to investigate why prices do not rise during peak demand periods. Chevalier, Kashyap and Rossi (2003) find that this is largely due to changes in retail margins which fall during peak demand periods.

Ivancic, Diewert and Fox (2011) use scanner data from 100 stores of four super market chains focusing on 19 item categories. Time aggregation choices lead to a difference in price change estimates for chained and superlative indexes suggesting that traditional index number theory appears to break down when high-frequency data is used. Statistics Netherlands is about to start using regularly scanner data for the compilation of the Dutch CPI. de Haan (2015) proposes a framework which could be employed in order to use scanner data for CPI aggregation.

Lloyd, McCorriston, Morgan, Poen and Zgovu (2012) use scanner data of food retailers in the UK to examine retailer characteristics as well as price dynamics. As mentioned in the data description above, scanner data can reveal information in macro as well as micro level. Lloyd *et al* (2012) find that the frequency of price adjustment and the duration of prices vary across retailers. Also, price promotions vary across retailers as well; some use sales regularly as a promotion tool whereas others rarely use them. These findings can help researchers analyse the local industry and examine consequences for CPI inflation.

Statistics New Zealand (2014) use scanner data to measure price inflation in a particular sector: consumer electronics. Scanner data allows more accurate price measurement, reflects seasonalities in quantities and product substitution. Therefore, such data can be a very powerful tool in nowcasting and short-term forecasting subcomponents of CPI inflation. On the other hand, as electronic goods are more and more sold directly on the web, scanner prices could progressively lose their relevance in favour of internet based prices.

Pandya and Venkatesan (2015) reveal another use of retail scanner prices: consumer behaviour. They use data from 135 super market chains, which accounts for about 79 % of the US super market sales during that period. They show that during the 2003 US-France dispute over the Iraq War the market share of French-sounding, US supermarket brands declined due to consumers' boycott.

2.6.2. INDICATIVE DATA SOURCES & EXAMPLES

As expected, scanner data is not publicly available. Two sources seem to dominate in the literature, those mentioned below. Official Statistical Agencies (e.g. Statistics Netherlands) should own data which is used for in-house analysis.

- The Nielsen Datasets on Consumer Panel Data and Retail Scanner Data have been used by many papers in the literature. Nielsen also provides POS data. The pricing ranges from $4,000 to $7,000 for institutional subscription. More information is available at https://goo.gl/ZOIHP4;
- The second most used dataset on scanner prices is the IRi dataset. This includes 11 years of weekly store data (2001-2011) for chain grocery and drug stores in 47 markets. The academic license data is $1,000 and the size of dataset is more than 350GB. More information is available at https://goo.gl/P5fcda.

2.6.3. CONCLUSIONS

Based on the above discussion, we can conclude that there is ample evidence in favour of scanner data in macro and micro level analysis. Nowcasting and short-term forecasting of specific CPI sub-components could benefit from scanned prices of retailers. Also, industry analysis can be carried out in a both frequent and regional basis. However, access to scanner data, and especially EU data, is very limited.

2.7. Online prices data

The development of internet gave rise to online shopping. According to Abramovich (2014), online shopping retail sales are predicted to grow to $370 billion in 2017, up from $231 billion in 2012. Therefore, since online shopping substitutes, or at least supplements, offline shopping, online prices can also be used as a substitute, or supplement, of offline prices. Data collection over the internet is called web scraping. This technique provides flexibility and extreme automation. As in the previous case with scanner data, scraped prices is a potentially useful instrument in nowcasting and short-term forecasting of CPI inflation as well as retail sales variables.

Online prices are also characterised by seasonalities and stylised facts, which as with scanner data, need to be taken into account by the researcher. Daily access to online super markets and retailers, which is publicly allowed, can lead to a mass collection of data. For example, a major UK retailer, Sainsbury's, offers 12 groceries categories for online shopping with about 50 ([24]) products per category. This leads to about 600 products online, thus 600 prices have to be collected from this retailer. Usually, there are 4 or more major retailers in a country which leads

([24]) This is a rough estimate as in specific categories there can be as much as 100 or more products.

to about 2,400 prices to be collected. Over the course of a calendar year, this sums up to about 864,000 prices per year.

2.7.1. APPLICATIONS

Academic papers in economics have started using web scraped data during the last six to seven years. Lunnemann and Wintr (2011) collected more than 5 million price quotes from price comparison websites for France, Italy, Germany, the UK and the US. Their data was collected daily for a year (December, 2004 – December, 2005). They find that for some product categories, prices change more frequently in the European countries. They also find that scraped prices are not more flexible than offline prices and, as mentioned in the scanner data section, there is heterogeneity in the frequency of price changes across online retailers.

Cavallo (2013) used web scraping to collect online prices from the largest supermarket retailers in Argentina, Brazil, Chile, Colombia and Venezuela. The time frame spans from October, 2007 to March, 2011. The paper finds that for Brazil, Chile, Colombia, and Venezuela, indexes using the online prices approximate both the level and main dynamics of official inflation. This is evidence that scraped prices could be used for inflation nowcasting. However, this might not be true for all economies. The paper finds that for Argentina, the online inflation rate is nearly three times higher than the official estimate. This data collection is part of the MIT Billion Prices Project ([25]). Rigobon (2015) and Cavallo and Rigobon (2016) provide a brief discussion of the project which is now expanded and prices are collected for European countries as well.

Boettcher (2015) describes in detail technological, data security and legal requirements of web crawlers focusing on Austria. The paper finds that web crawling technology provides an opportunity to improve statistical data quality and reduce the overall workload for data collection. Automatic price collection methods enable statisticians to react better to the increasing amount of data sources on the internet.

Cavallo (2016) uses again scraped prices to study the impact of measurement bias on three common price stickiness statistics: (i) the duration of price changes, (ii) the distribution of the size of price changes, and (iii) the shape of their hazard function over time. The paper finds that online prices have longer durations, with fewer price changes close to zero, and hazard functions that initially increase over time. The author claims that the differences with the literature are due to time-averaging and imputed prices in scanner and CPI data.

Metcalfe, Flower, Lewis, Mayhew and Rowland (2016) introduce the CLIP, which is an alternative approach to aggregating large data sets into price indices using clustering. The CLIP uses all the data available by creating groups (or clusters) of similar products and monitoring the price change of these groups over time. Unsupervised and supervised machine learning techniques are used to form these product clusters. The index is applied on web scraped data. The authors explicitly say that this index does not replace official statistics, however it clearly shows the interest of official statistics agencies, the UK ONS in this case, in online prices. Also, Radzikowski and Smietanka (2016) try to construct a CPI for Poland based entirely on online prices. Cavallo (2017) compares the online and offline prices of 56 large multi-channel retailers in 10 countries: Argentina, Australia, Brazil, Canada, China, Germany, Japan, South Africa, the UK and the US. He finds that price levels are identical about 72 percent of the time. Price changes are not synchronised but have similar frequencies and average sizes. These results show that potentially scanner prices, which are more difficult to collect on a daily basis, can be substituted by online prices.

([25]) Available at https://goo.gl/xb4H95.

2.7.2. INDICATIVE DATA SOURCES & EXAMPLES

We list below a short list of ready databases with scraped prices. Alternatively, an official statistics agency would start collecting/scraping the online prices from scratch, i.e. using major super markets, retailers, etc. In a few years, this procedure could generate enough data to cross-validate the ability of scraped prices for nowcasting CPI inflation, retail sales and other price related variables.

- MIT Billion Prices Project, https://goo.gl/xb4H95. Most of the data mentioned in the work of Cavallo is available online. Particularly:
  - Daily prices for all goods sold by 7 large retailers in Latin America and the US: 2 in Argentina, 1 in Brazil, 1 in Chile, 1 in Colombia, 1 in Venezuela, and 4 in the US. Daily data from 2007 to 2010. Used in Cavallo (2016),
  - Daily prices for all goods sold by APPLE, IKEA, ZARA, and H&M. Daily data for 85 countries ([26]) from 2008 to 2013,
  - Online and offline prices for individual goods sold by 56 large multi-channel retailers in 10 countries: Argentina, Australia, Brazil, Canada,
  - China, Germany, Japan, South Africa, the UK, and the US. Mixed frequency data from 2014 to 2016. Used in Cavallo (2017);
- PriceStats, https://goo.gl/zqQoaS. This is a private company. Cavallo (2016) used scraped data for 181 retailers in 31 countries provided by PriceStats. The dataset could be purchased.

2.7.3. CONCLUSIONS

Online prices is definitely a source of big data which seems to have potential in CPI inflation nowcasting and forecasting and is something which has not been extensively studied before. Data for European countries does exist by private companies, however it is not publicly available. Official statistical agencies will likely experiment with this method of price collection in the near future. It might be useful in that case to complement price data with volumes of sales.

2.8. Online search data

Online search data consists of searches for particular keywords on the world wide web. The user typically inserts a keyword or a phrase in the search field of a search engine website. Then, the web search engine returns the information which mostly relates to the keyword. The information may be a mix of web pages, images, and other types of files. Search engines maintain real-time information by running an algorithm on a web crawler, thus a newly uploaded website must be easily accessible by search engine robots in order to be included in the databases for future web searches.

Between 1993 and 1995, Lycos, Altavista and Yahoo were some of the first web search engines that gained popular attention and daily visits. However, the search results were based mainly on the web directory of each engine rather than its full-text copies of web pages. Some years later, in about 2000, Google was introduced. The company achieved better results for many searches with an innovative procedure called PageRank. This algorithm ranks web pages based on the number and PageRank of other web sites and pages that link there, on the premise that

([26]) Coverage of countries varies by retailer and time.

good or desirable pages are linked to more than others. Since then, Google search engine has dominated followed recently by Microsoft's Bing and Baidu ([27]).

Google launched a public web facility, Google Trends, which shows how often a particular keyword is searched relative to the total search-volume across various regions of the world, and in various languages. The procedure is simple for all internet users and Google Trends data is publicly available. At first, the user specifies the keyword or search-items he/she wants to look for. Then, Google Trends returns a time series line plot with time on the horizontal axis and search frequency on the vertical axis. The time series data is offered at weekly frequency starting in 2004 and can be downloaded in .csv format. Thinking about the behaviour of internet users, who might search for particular search-items multiple times throughout the day, it is easy to understand that the raw search data Google has is a type of big data. Therefore, Google Trends is a weekly aggregated, and thus structured, form of big data (even though it might not be 'big' itself).

Google also offers another tool which aims to help the user with specific keywords and search-items, Google Correlate available at https://goo.gl/Uj1mox. This service, which is part of Google Trends, finds search patterns which correspond with real-world trends. There are two ways a researcher can use this tool. First, if there exists a weekly or monthly time series data which is of interest, the researcher can upload this data on Google and identify search-items and keywords which are correlated with the time series. This, in principle, is useful as the keywords, which will then be used to extract Google Trends, are almost automatically selected. The second use of Google correlates would be between Google Trends and keywords. In case a researcher has already identified a particular search-item, Google Correlate can be used in order to provide a list of correlated keywords which could be used in the analysis in order to decrease selection bias ([28]). However, given that this is an automatic procedure, Google Correlate is not able to filter not appropriate keywords. For example, Google Correlate might return as a result celebrities' names which happen to be trending during the same time period.

As we briefly describe below, Google Trends have been used in various applications in economics, finance, health sector, etc. with substantial success. Some specific characteristics of online search data which need to be taken into account by the researcher are: (i) clear and intuitive keywords and search-items (see Ross, 2013), (ii) weekly patterns and seasonalities and (iii) extreme events, such as physical phenomena or politics, which might result to outliers (even though sometimes could be useful as they illustrate agents behaviour, see Vosen and Schmidt, 2011).

2.8.1. APPLICATIONS

The literature which incorporates the use of online search data is already vast, even though it started very recently. Below, we briefly mention some indicative papers in economics, finance, tourism industry, health and politics. However, the use of online search data expands to other interdisciplinary fields as well. Chamberlain (2010), Choi and Varian (2012), Varian and Stephens-Davidowitz (2014) are among the first to provide nowcasting and forecasting applications using Google Trends. Also, Kapetanios *et al* (2016) provide a detailed overview and application of Google Trends data in EU macroeconomic nowcasting.

Firstly, internet data has applications in health. Ginsberg, Mohebbi, Patel, Brammer, Smolinski and Brilliant (2009) present a method of analysing large numbers of Google search queries

([27]) This is mainly used in China.
([28]) This could be done by aggregating or averaging the Google Trend time series, or by extracting their common factors.

to track influenza-like illness in a population. Also see Yuan, Nsoessie, Lv, Peng, Chunara and Brownstein (2013) for influenza epidemics monitoring in China using Baidu search results. Tefft (2011) uses Google Trends to create a depression search index to provide insights on the relationship between unemployment, unemployment insurance, and mental health. The results indicate a positive relationship between the unemployment rate and the depression search index.

Then, Schmidt and Vosen (2011) propose a new indicator for the US private consumption based on search query time series provided by Google Trends. They find that the Google indicator outperforms the relevant survey-based indicators, the University of Michigan Consumer Sentiment Index and the Conference Board Consumer Confidence Index. This suggests that incorporating information from Google Trends may offer significant benefits to forecasters of private consumption.

Koop and Onorante (2013) suggest to nowcast using dynamic model selection (DMS) methods which allow for model switching between time-varying parameter regression models. This is potentially useful in an environment of coefficient instability and over-parametrisation which can arise when forecasting with Google variables. They allow for the model switching to be controlled by the Google variables through Google probabilities. That is, instead of using Google variables as regressors, they allow them to determine which nowcasting model should be used at each point in time. In an empirical exercise involving nine major monthly US macroeconomic variables, they find that DMS methods provide large improvements in nowcasting. The use of Google model probabilities within DMS often performs better than conventional DMS. Their application includes various US macroeconomic variables.

Bontempi, Golinelli and Squadrani (2016) develop a new uncertainty indicator based on Google Trends with applications to US variables. Their results suggest that the Google Trends indicator shocks embody timely information about people's perception of uncertainty and sometimes earlier than other indexes. This is evidence that online search data could act as a leading indicator and used in macroeconomic nowcasting and forecasting.

Choi and Varian (2009) and Choi and Varian (2012) illustrate the ability of Google Trends to predict the present using daily and weekly reports of Google Trends. In particular, they claim that people who lose their jobs search the internet for job ads. Therefore, the increasing volume of Google search queries for job-related keywords potentially has an impact on forecasting/nowcasting the initial claims. Their applications include US variables.

Reis, Ferreira and Perduca (2015) analyse the web activity as a big data source. Electronic traces left by users while they use web services could be used as data either in real time or with very small time lags. As many human activities measured by official statistics are closely related to people's behaviour online, this data on people's web activity offers the potential to produce predictions for socio-economic indicators with the purpose to increase the timeliness of the statistics. Papers in the literature have shown evidence that these predictions can be made. However, this type of data should be further checked about its transparency, continuity, quality and potential to be integrated with official statistics traditional methods. The empirical application they implement is an improved nowcasting of French and Italian unemployment. More recently, Smith (2016) use a mixed-frequency model in UK unemployment forecasting.

Wirthmann and Descy try to understand demand of labour and skills with a particular focus to the job requirement evolution web scraping of job vacancies. The aim is explore ways to better utilise existing data to produce derived measures of skills demand, skills supply, mismatches and skills development. A prototype system, which was successfully tested in five countries (UK, DE, CZ, IT, IE). The prototype was designed to retrieve from job portals selected on the basis of quality and relevance. Adopting an ethical behaviour, authors only scraped portals for which permission was granted or, even better, obtained direct access to job portals' databases. Data processing transforms 'documents' — original job posting as read by by the crawler — into 'vacancies' — pre-structured single job openings. This is done by expanding job posting — one job posting can contain more than one job opening - and de-duplication — one job opening can be posted in more than one website. Vacancies are processed using text mining and machine learning algorithms to identify classify jobs into occupations and gather information about contract type, working hours, job location as well as skills and job requirements. The European skills, competence, qualification and Occupations taxonomy (ESCO) was used as multilingual taxonomy. Results are accessible via OLAP cube as well as through simple graphic interface.

2.8.2. INDICATIVE DATA SOURCES & EXAMPLES

Perhaps the most intuitive, publicly available and accurate source of internet search data is Google Trends: https://goo.gl/64mcg6.

2.8.3. CONCLUSIONS

Internet search data and particularly Google Trends is a very promising source of structured big data. The literature provides ample scientific evidence that web search data has predictive abilities in various fields such as economics, finance, politics and health. This is not surprising as web search data is input of humans and thus, reflects agents' behaviour. However, it must be noted that online search data must be carefully used as the value it adds depends on the nowcasting exercise. For example, unemployment rate benefits from online search data whereas GDP growth might not, unless perhaps a large universe of manually selected keywords is employed.

2.9. Textual data

This includes any kind of dataset providing summarised information in the form of text. Examples of textual data include news and media headlines, information related to specific events, e.g. central banks' board meetings, Twitter data (this is analysed in more details in the next section) and Wikipedia information.

2.9.1. APPLICATIONS

The literature on textual data in economics mainly uses newspaper text and text data from the FOMC minutes of the FED.

General information about text mining is provided in Smith (2010) and Bholat, Hansen, Santos and Schonhardt-Bailey (2015). Predictive Analytics ([29]) website lists a number of software which can be used in text analysis.

Schumaker and Chen (2006) investigate 9,211 financial news articles and 10,259,042 stock quotes covering the S&P 500 stocks during a five week period. They show that the model containing both article terms and stock price at the time of article release had the best performance in

([29]) https://goo.gl/SkNnVs.

closeness to the actual future stock price, the same direction of price movement as the future price and the highest return using a simulated trading engine.

Moat, Curme, Avakian, Stanley and Preis (2013) use the frequency of Wikipedia page views. The paper presents evidence that financially-related Wikipedia page views have predictive ability over firecessions.

Levenberg, Pulman, Moilanen, Simpson and Roberts (2014) present an approach to predict economic variables using sentiment composition over text streams of Web data. Their results show that high predictive accuracy for the Nonfarm Payroll index can be achieved using this sentiment over big text streams.

Baker, Bloom and Davis (2015) ([30]) develop an index of economic policy uncertainty based on newspaper coverage frequency.

Using firm-level data, the authors find that policy uncertainty raises stock price volatility and reduces investment and employment in policy-sensitive sectors like defense, healthcare, and infrastructure construction. At the macro level, policy uncertainty innovations foreshadow declines in investment, output, and employment in the United States and, in a panel VAR setting, for 12 major economies. Using the same indicators for policy uncertainty, Bacchini, Bontempi, Golinelli and Jona-Lasinio (2017), provide similar results for slowdown of Italian investments. Ericsson (2015) and Ericsson (2016) construct indexes that quantify the FOMC views about the U.S. economy, as expressed in the minutes of the FOMC's meetings. Steckler and Symington (2016) quantify the minutes of the FOMC and show that the FOMC saw the possibility of a recession but did not predict it. Using textual analysis, the authors are able to determine which variables informed the forecasts.

Thorsrud (2016) constructs a daily business cycle index based on quarterly GDP and textual information contained in a daily business newspaper. The newspaper data is decomposed into time series representing newspaper topics. The textual part attributes timeliness and accuracy to the index and provides the user with broad based high frequent information about the type of news that drive or reflect economic fluctuations. Eckley (2015) develops a news-media textual measure of aggregate economic uncertainty using text from the Financial Times. This index is documented to have a strong relationship with stock volatility on average.

Textual analysis can also be used in political economy. Acemoglu, Hassan and Tahoun (2015) use textual data from GDELT project to proxy street protests in Arab countries and investigate the relationship between protests and stock market returns. Using daily variation in the number of protesters, they document that more intense protests in Tahrir Square are associated with lower stock market valuations for firms connected to the group in power relative to non-connected firms, but have no impact on the relative valuations of firms connected to other powerful groups.

2.9.2. INDICATIVE DATA SOURCES & EXAMPLES

Below we provide an indicative list of textual data sources.

- First, an obvious source of newspaper headlines and media coverage is access to newspaper archives and other archive websites. There are two difficulties associated with archives: (i) the first is that archives offer scanned images or photographs of newspaper pages, thus a transformation of newspaper page images to text is necessary in order to create a textual

([30]) Their index is available online with real-time information and updates at: https://goo.gl/ZCWx92.

database [31], (ii) language is the second difficulty as national newspapers use the official language in each country:

- The British Newspaper archive, https://goo.gl/gdnt0s. It provides access to a historical archive which spans 200 years for 717 UK newspapers titles. It is publicly available but unlimited access requires a £12.95/month subscription,
- National Library of Australia — Trove, https://goo.gl/4yFmFC. This database offers access to 500 newspaper titles covering the period 1803–2011,
- Google News Archive, https://goo.gl/7IPVwL. Google provides free access to scanned archives of newspapers. Some of the news archives date back to 1700s and include many national and regional newspapers in multiple languages,
- Wikipedia: List of online newspaper archives, https://goo.gl/Ig2WPl. Wikipedia provides a detailed list of available online sources for each country;

• Reuters: Reuters Online News Archive — US Edition: https://goo.gl/G3lb30. The Reuters Online News Archive is a collection of news articles published on the United States edition of the Reuters website, starting from January 1st 2007. The archive is continuously updated as content is published on the website. The online repository is structured as a plain HTML web-page and can be browsed by date: subpages simply list articles published on a single day in inverse chronological order. Although the archive is not searchable, its simple design allows for easy web-scraping. Since articles are listed as they are published and/or updated, the archive contains duplicates and a large number of links return a 404 error (page not found). The Reuters News Archive has been used mainly to investigate interrelations among financial institutions (see Rönnqvist and Sarlin, 2015);

• GDELT Project [32], https://goo.gl/6FWcQl. The Global Database of Events, Language, and Tone (GDELT) is 'an initiative to construct a catalog of human societal-scale behavior and beliefs across all countries of the world, connecting every person, organization, location, count, theme, news source, and event across the planet into a single massive network that captures what's happening around the world, what its context is and who's involved, and how the world is feeling about it, every single day'. It is mainly used for political conflict (see, e.g. Acemoglu, Hassan and Tahoun (2014), Heaven (2013) among others), however it could be used potentially in the building of an uncertainty index as it provides data every 15 mins. It must be highlighted that the data volume is very big and noise could be present (see Beieler, 2013).

2.9.3. CONCLUSIONS

Financial applications using textual information are mostly concerned with uncertainty indexes and market forecasting based on newspaper headlines and quantified FOMC minutes. Newspaper headlines could be potentially used for macroeconomic nowcasting on a daily basis. The GDELT project could also be a potential instrument in the creation of an uncertainty index based on big data. This, in turn, could be used in macroeconomic nowcasting and forecasting (provided that its predictive abilities have been identified).

(31) There exist, at least, two Optical Character Recognition (OCR) packages available in R: abbyyR and Tesseract. In both cases, the use loads a vector with image data (.png, .jpeg, .tiff) and R recognizes and scraps the main text.

(32) GDELT provides an online interface to make queries about particular keywords. However, GDELTTools, https://goo.gl/IqDnl9, can also be used to load and normalise data inside R without the need of other user interfaces.

2.10. Social media data

Since the introduction of the internet, users were finding ways to communicate with each other. Message boards, guestbooks, chat platforms and personal sites have been online for many years. These services, which could be considered as primal social networks, set the grounds for modern social platforms and the new age of online social interaction. Facebook was officially launched in 2004, however it was not until 2006 that it was widely open to all internet users aged 13 or older. It has about 1.86 billion monthly active users as of December 31, 2016 ([33]). Since its introduction, social networks have been rapidly expanded and become an integral part our lives. The social networks are accessible via computers as well as mobile devices allowing for continuous connectivity and interaction to news and events. Following Facebook, Twitter was introduced in 2006 as an online news and social networking service where users post and interact with messages, or 'tweets', restricted to 140 characters. Tumblr, launched in 2007, is a service which allows users to post multimedia and other content to a short-form blog (or 'micro-blogging'). Instagram, going online in 2010, started as a photo-sharing site which now allows the post videos as well.

Social media, as in the case with online search data, illustrate human activity and reactions. Discussions or posts on Facebook, Twitter or Instagram include a variety of topics from personal issues to politics and breaking news. Therefore it would be reasonable to assume that, as in the case with Google Trends, social media data could have a predictive ability towards social and economic variables. Based on the types of social media, Twitter seems to be the most appropriate to use for scientific analysis. Below we list some key reasons:

- First, Twitter mainly uses short text streams which are often very specific about an event. In this sense, Twitter could also be part of the Textual Analysis data described in the previous section;
- The use of hashtags (#) makes Twitter discussions easier for monitoring and tracking events. This allows identifying which discussions are 'trending';
- Twitter data, due to its 'higher'-frequency nature can offer more information. See Paul, Dredze and Broniatowski (2014) who argue that influenza forecasting using Twitter data is improved compared to Google Flu Trends;
- Politicians ([34]), reporters and analysts have included Twitter as one of their main means of communication. For example, reports correspondence during Federal Open Market Committee meetings are often on Twitter with multiple tweets for breaking news. The Guardian newspaper uses Twitter feeds on their website. Particularly, the use of Twitter during the Brexit referendum and US elections was very successful.

However, as in all big data types, Twitter data might include a lot of noise. Therefore a careful selection of topics/hashtags must be done. An alternative way to using hashtags would be to follow specific users or organisations. For example, a researcher could monitor the twitter feeds from various newspaper and media organisations, government agencies and key reporters and analysts and, then, filter the feeds for particular hashtags or keywords.

([33]) See https://newsroom.fb.com/company-info/.
([34]) See Chi and Yang (2010) for more information about the use of Twitter in the Congress.

2.10.1. APPLICATIONS

Social media data use in finance and economics has rapidly increased. The University of Michigan has established economic indicators based on social media with encouraging results on forecasting. They publish online their nowcasts on initial claims for unemployment insurance using their Social Media Job Loss Index ([35]). Ramli (2012) and Griswold (2014) are just some examples of the media interest on the use of social networks data in economic analysis. The literature indicates three main areas of application: (i) financial markets; mainly stock market and foreign exchange, (ii) politics ([36]), and (iii) public mood ([37]). Below, we provide some additional information focusing more on the financial markets.

Economics and financial researchers have realised the usefulness of social networks since 2004. Antweiller and Frank (2004) are among the first to use a big data of online message boards to examine the impact of social networks on stock returns. They analyse 1.5 million messages from Yahoo!Finance and Raging Bull for the 45 Dow Jones Industrial Average. They confirm that stock messages help predict market volatility and their effect on stock returns is statistically significant (although economically small).

The stock market forecasting greatly benefits from the use of Twitter. The key reason for this is that the hashtag keyword is the stock's ticker, which allows for a clear and precise feed. Many papers in the academic and professional literature demonstrate the usefulness of Twitter data in stock directional movements or returns. Bollen, Mao and Zheng (2011) and Mittal and Goel (2012) extract investors' sentiments from Twitter data to predict the Dow Jones Industrial Average. Both papers, although they follow different methodologies, report an improvement in DJIA forecasting accuracy by 86.7-87 %. Makrehchi, Shah and Liao (2013) also create Twitter sentiment data to predict the S&P500. They use two rating scale systems for moods: OpinionFinder and Google-Profile-of-Mood States. The authors build a trading strategy using the signals from the sentiment data outperforming the standard buy-and-hold by about 20 %. Due to the proprietary nature of the Google-Profile of Mood States algorithm, Chen and Lazer experiment with simpler methods which could provide similar results.

Arias, Arratia and Xuriguerra (2013) present empirical work using Twitter forecasting in (i) US box-office sales, and (ii) Apple, Google, Yahoo, Microsoft, S&P100 implied volatility index, S&P500 index and S&P500 implied volatility index closing prices. As in the previous cases, sentiment indicators are created based on Twitter data. As a general result, they show that nonlinear models do take advantage of Twitter data when forecasting trends in volatility indices, while linear ones fail systematically when forecasting any kind of financial time series. In the case of predicting box office revenue trend, it is support vector machines that make best use of Twitter data.

Sprenger, Tumasjan, Sandner and Welpe (2014a) also use Twitter sentiment to predict stock returns. The authors use data for an undisclosed S&P 100 company and find an association between tweet sentiment and stock returns, message volume and trading volume, as well as disagreement and volatility. Their results demonstrate that users providing above average investment advice are retweeted more often and have more followers, which amplifies their share of voice. Sprenger, Tumasjan, Sandner and Welpe (2014b) use computational linguistics to a dataset of more than 400,000 S&P 500 stock-related Twitter messages, and distinguish between good and bad news. The results indicate that the returns prior to good news events

([35]) Visit https://goo.gl/jfvpDm for more information.
([36]) See Tumasjan, Sprenger, Sandner and Welpe (2010), Conover, Goncalves, Ratkiewicz, Flammini and Menczer (2011), Wang, Can, Kazemzadeh, Bar and Narayanan (2012) and Makhazanov and Rafiei (2013) among others.
([37]) See Bollen, Mao and Pepe (2011) and Lansdall-Welfare, Lampos and Cristianini (2012) among others.

are more pronounced than for bad news events. Pineiro-Chousa, Vizcaino-Gonzalez and Perez-Pico (2016) use data from Stocktwits, a social network similar to Twitter, where users share posts about stocks, indexes, and financial markets. They focus on investors' activity through social media and these media's influence over the Chicago Board Options Exchange Market Volatility Index. The results show that social media sentiment influences stock markets.

2.10.2. INDICATIVE DATA SOURCES & EXAMPLES

Following the discussion above, we can claim that the most obvious, and possibly useful, source of social media data for economic applications is Twitter:

- Twitter public streams data can be downloaded for free using the Twitter API, see https://goo.gl/YYaTaA. An R package, twitteR ([38]), can be integrated and Twitter data can be downloaded automatically in R for further editing. However, the Twitter API data can go back 1400 days;
- Alternative services which offer Twitter data which can go back to the first day Twitter was online are provided by third-party agents. A reliable company is gnip, https://goo.gl/dmKJzb, and sifter, https://goo.gl/uHh1oc. Both these services are proprietary;
- For financial market instrument with ticker symbols, StockTwits could be used as well. However, Stocktwits social media market share is much smaller compared to Twitter's.

2.10.3. CONCLUSIONS

There has been documented by now plenty of evidence in favour of Twitter use in the creation of sentiment indexes. Most applications are in financial markets instruments, elections and public mood. There does not seem to exist empirical work in macroeconomic nowcasting or forecasting, however Twitter could be one of the ingredients in a sentiment indicator, possibly with Google Trends as well.

3. Nowcasting specific macroeconomic variables using big data

In Section 2, we have considered a typology of big data potentially relevant for macroeconomic analysis and, in particular, for nowcasting/forecasting, and for each of them discussed some relevant applications. As mentioned in the Introduction, we now consider the dual problem: for a specific macroeconomic variable of interest, such as unemployment, GDP and components, inflation, surveys, financial variables, we list studies based on nowcasting them using big data. More specifically, for each study we provide: (i) author(s) and their affiliation, (ii) brief description of the paper, (iii) data characteristics (big data and standard macro/financial variables), (iv) econometric methodology. A summary of the most useful papers and their characteristics is described in Table 2.

([38]) See https://goo.gl/cuLZTL for more information.

Table 2: Indicative list of most important work in macroeconomic nowcasting using big data

	Title	Auhor(s)	Data	Method(s)	Country	Big data Added Value	Limitation(s)
Unemployment	Predicting Initial Claims for Unemployment Benefits and Predicting the Present with Google Trends	Choi, Varian	Google	Linear Regression	US	Improvement of fore(now-) casting	Limited number of other indicators, Methodology
Unemployment	Google Econometrics and Unemployment Forecasting	Askitas, Zimmermann	Google	VECM	DE	Improvement of fore(now-) casting	Limited number of other indicators, Methodology
Unemployment	The Predictive Power of Google Searches in Forecasting Unemployment	D'Amuri, Marcucci	Google	ARMA	US	Improvement of fore(now-) casting	Methodology
Unemployment	Nowcasting with Google Trends: a Keyword Selection Method	Ross	Google	Linear Regression	UK	Improvement of fore(now-) casting	Methodology
Unemployment	The Use of Web Activity Evidence to Increase the Timeliness of Official Statistics Indicators	Reis, Ferreira, Perduca	Google	Linear Regression	FR, IT	Big data in official statistics, Improvement of fore(now-) casting	Methodology
Unemployment	Improving Prediction of Unemployment Statistics with Google Trends: Part 2	Ferreira	Google	Linear Regression, Factor Models	PT	Latent variables, Improvement of forecasting	Methodology
GDP and components	Nowcasting GDP with Electronic Payments Data	Galbraith, Tkacz	Debit card and Credit Card transactions, Cheques, housing index, employment, stock exchange index, money supply, average work week (hours), new orders, durables, inventories, retail trade	Linear Regression, Factor Models	CA	Nowcasting improvement using big data	Data, Methodology
GDP and components	Forecasting Private Consumption: Survey-based Indicators vs Google Trends	Schmidt, Vosen	Google, Consumer Sentiment, Consumer Confidence	Linear Regression	US	Activity Indicator, forecasting of macro variables	Methodology
GDP and components	Are ATM/POS data relevant when nowcasting private consumption?	Esteves	ATM/POS	Linear Regression	PT	Nowcasting improvement using big data	Data, Methodology
GDP and components	Dankort payments as a timely indicator of retail sales in Denmark	Carlsen, Storgaard	Debit card data		DK	Nowcasting improvement using big data	Data, Methodology
GDP and components	A mixed frequency approach to forecast private consumption with ATM/ POS data	Duarte, Rodrigues, Rua	ATM/POS	MIDAS	PT	Nowcasting improvement using big data	Data, Methodology

Table 2: Indicative list of most important work in macroeconomic nowcasting using big data

	Title	Auhor(s)	Data	Method(s)	Country	Big data Added Value	Limitation(s)
GDP and components	Using the payment system data to forecast the Italian GDP	Aprigliano, Ardizzi, Monteforte	Payments data	MIDAS, LASSO	IT	Nowcasting improvement using big data	
GDP and components	Macroeconomic Nowcasting Using Google Probabilities	Koop, Onorante	Inflation, Wage Inflation, Unemployment, Term Spread, FCI, Commodities Price Inflation, Industrial Production, Oil Price Inflation, Money Supply	Model Averaging	US	Use of Google Data as proxy for improved weighting scheme	
Inflation	The Billion Prices Project: Research and Inflation Measurement Applications	Cavallo	Google, CPI, Gas	VAR	Argentina, US, EA	Use of big data in inflation forecasting	Data
Inflation	Automatic Data Collection on the Internet (Web Scraping)	Boettcher	Prices	-	AT	Use of Web scraping	Algorithms, Storage, Quality of data
Inflation	Collecting Clothing Data from the Internet	Griffioen, de Haan, Willenborg	CPI, Apparel	-	NL	Use of Web scraping	Algorithms, Storage, Quality of data
Inflation	Using Web Scraped Data to Construct Consumer Price Indices	Breton, Swier, O'Neil	Inflation, CPI, RPI	Aggregation	UK	Use of Web scraping in aggregate index construction	Quality of data

3.1. Unemployment

3.1.1. PREDICTING INITIAL CLAIMS FOR UNEMPLOYMENT BENEFITS AND PREDICTING THE PRESENT WITH GOOGLE TRENDS

Author(s): Choi, H., Varian, H. (Google Inc.)

Brief Description: Choi and Varian (2009) and Choi and Varian (2012) illustrate the ability of Google Trends to predict the present (nowcasting) using daily and weekly reports of Google Trends. In particular, they claim that people who lose their jobs search the internet for job ads. Therefore, the increasing volume of Google search queries for job-related keywords potentially has an impact on forecasting/nowcasting the initial claims.

Data: Google Search Insights and Google Trends, Retail Sales, Automotive Sales, Home Sales, Travel

Methodology: Linear Regression Models with/without lags and independent variables

3.1.2. GOOGLE ECONOMETRICS AND UNEMPLOYMENT FORECASTING

Author(s): Askitas, N. (IZA), Zimmermann, K. (Bonn University, DIW Berlin, IZA)

Brief Description: The paper suggests an innovative method of using data on internet activity to predict economic behaviour in a timely manner, which is difficult at times of structural change. They show a strong correlation between keyword searches and unemployment rates using monthly German data.

Data: Seasonally unadjusted monthly unemployment rate of Germany (01/01/2004- 01/04/2009). Google Insights. German Keywords (in English): unemployment office/agency', 'unemployment rate', Personnel Consultant', popular job search engines in Germany (stepstone, jobworld, jobscout etc.).

Methodology: Time series causality analysis using Error Correction Model (ECM)

3.1.3. THE PREDICTIVE POWER OF GOOGLE SEARCHES IN FORECASTING UNEMPLOYMENT

Author(s): D'Amuri, F., Marcucci, J. (Bank of Italy, Economic Research and International Relations)

Brief Description: D'Amuri and Marcucci (2012) suggest the use of an index of Internet job-search intensity (the Google Index, GI) as the best leading indicator to predict the US monthly unemployment rate. They perform a deep out-of-sample forecasting comparison analysing many models that adopt their leading indicator, the more standard initial claims or combinations of both. They find that models augmented with the GI outperform the traditional ones in predicting the unemployment rate for different out-of-sample intervals that start before, during and after the Great Recession. Google-based models also outperform standard ones in most state-level forecasts and in comparison with the Survey of Professional Forecasters. These results survive a falsification test and are also confirmed when employing different keywords.

Data: Google Data (Unit Root testing and transformations)

Methodology: ARMA, ARMAX (various combinations)

3.1.4. NOWCASTING WITH GOOGLE TRENDS: A KEYWORD SELECTION METHOD

Author(s): Ross, A. (Fraser of Allander Institute, University of Strathclyde)

Brief Description: Ross (2013) investigates the issues of identifying and extracting keywords from Google Trends relevant to economic variables. He suggests the backward induction method which identifies relevant keywords by extracting these from variable relevant websites. This backward induction method was applied to nowcast UK unemployment growth using a small set of keywords. The majority of keywords identified using the backward induction method outperformed the competing models in terms of in-sample and out-of-sample tests of predictability indicating that the backward induction method is effective in identifying relevant keywords.

Data: Google Data, Unemployment

Methodology: Linear Regressions

3.1.5. THE USE OF WEB ACTIVITY EVIDENCE TO INCREASE THE TIMELINESS OF OFFICIAL STATISTICS INDICATORS

Author(s): Reis, F. (Eurostat), Ferreira, P. (Eurostat), Perduca, V. (Universite Paris Descartes, CNRS)

Brief Description: Reis *et al* (2015) analyse the web activity as a big data source. Electronic traces left by users while they use web services could be used as data either in real time or with very small time lags. As many human activities measured by official statistics are closely related to people's behaviour online, this data on people's web activity offers the potential to produce predictions for socio-economic indicators with the purpose to increase the timeliness of the statistics. Papers in the literature have shown evidence that these predictions can be made. However, this type of data should be further checked about its transparency, continuity, quality and potential to be integrated with official statistics traditional methods. The empirical application they implement is an improved nowcasting of French and Italian unemployment.

Data: French and Italian Job related keywords

Methodology: Linear Regression

3.1.6. IMPROVING PREDICTION OF UNEMPLOYMENT STATISTICS WITH GOOGLE TRENDS: PART 2

Author(s): Ferreira, P. (Eurostat)

Brief Description: Ferreira (2015) uses a dynamic factor model to extract a latent variable from Google Trends data which is a good proxy for the unemployment dynamics. Prediction models for unemployment that make use of the estimated latent variable have performed better than the proposed approaches in previous works, in particular during a period where there was an abrupt change in the trend.

Data: Google Trends

Methodology: Linear Regression, Dynamic Factor Models

3.2. GDP and components

3.2.1. NOWCASTING GDP WITH ELECTRONIC PAYMENTS DATA

Author(s): Galbraith, J. W., Tkacz, G. (ECB)

Brief Description: Galbraith and Tkacz (2015) assess the usefulness of a large set of electronic payments data comprising debit and credit card transactions, as well as cheques that clear through the banking system, as potential indicators of current GDP growth in Canada. These variables capture a broad range of spending activity and are available on a very timely basis, making them suitable current indicators. While every transaction made with these payment mechanisms is in principle observable, the data are aggregated for macroeconomic forecasting. Controlling for the release dates of each of a set of indicators, they generate nowcasts of GDP growth for a given quarter over a span of five months, which is the period over which interest in nowcasts would exist. They find that nowcast errors fall by about 65 per cent between the first and final nowcast. Among the payments variables considered, debit card transactions appear to produce the greatest improvements in forecast accuracy.

Data: Debit card and Credit Card transactions, Cheques, lagged GDP, housing index, business and personal sales employment, stock exchange index, money supply, average work week (hours), new orders, durables, inventories, retail trade.

Methodology: Mostly based on linear regressions

3.2.2. FORECASTING PRIVATE CONSUMPTION: SURVEY-BASED INDICATORS VS GOOGLE TRENDS

Author(s): Schmidt, T. (RWI), Vosen, S. (RWI)

Brief Description: Schmidt and Vosen (2011) introduce an indicator for private consumption based on search query time series provided by Google Trends. The indicator is based on factors extracted from consumption-related search categories of the Google Trends application Insights for Search. The forecasting performance of this indicator is assessed relative to the two most common survey-based indicators — the University of Michigan Consumer Sentiment Index and the Conference Board Consumer Confidence Index. The results show that in almost all conducted in-sample and out-of-sample forecasting experiments the Google indicator outperforms the survey-based indicators.

Data: Google Insight/Trends, Michigan Consumer Sentiment, Conference Board Consumer Confidence

Methodology: Linear Regression Models

3.2.3. MACROECONOMIC NOWCASTING USING GOOGLE PROBABILITIES

Author(s): Koop, G. (University of Strathclyde), Onorante, L. (ECB)

Brief Description: Koop and Onorante (2013) suggest to nowcast using dynamic model selection (DMS) methods which allow for model switching between time-varying parameter regression models. This is potentially useful in an environment of coefficient instability and over-parametrisation which can arise when forecasting with Google variables. They allow for the model switching being controlled by the Google variables through Google probabilities. That is, instead of using Google variables as regressors, they allow them to determine which nowcasting model should be used at each point in time. In an empirical exercise involving nine major monthly US macroeconomic variables, they find that DMS methods provide large improvements in nowcasting; the variables are: inflation, industrial production, unemployment, wage inflation, money, supply, financial conditions index (FCI), oil price inflation, commodity price inflation and the term spread. The use of Google model probabilities within DMS often performs better than conventional DMS.

Data: Inflation, Wage Inflation, Unemployment, Term Spread, FCI, Commodities Price Inflation, Industrial Production, Oil Price Inflation, Money Supply (see Table 1 of the paper for more details and transformations)

Methodology: Dynamic Model Averaging

3.3. Inflation

3.3.1. THE BILLION PRICES PROJECT: RESEARCH AND INFLATION MEASUREMENT APPLICATIONS

Author(s): Cavallo, A. (MIT), Rigobon, R. (MIT)

Brief Description: Cavallo and Rigobon (2016) examines ways to deal with price data. Potential sources for micro price data include: Statistical Offices, Scanner Data (e.g. Nielsen), Online data (e.g. Billion Prices Project) etc. CPI data is useful in measuring inflation whereas Scanner and Online data can be used in marketing analytics (e.g. market shares). The Billion Prices Project is an automated web-scraping software where a robot downloads a public page, extracts the prices information and stores it in a database. A direct outcome from the papers is that online data is also useful for nowcasting inflation in the US, Latin America and Euro Area. Links between online data and CPIs are tracked using VAR models and calculating the cumulative Impulse Response Functions. The forecasting examples use predictive regressions.

Data: Google Data, CPI, Gas Prices

Methodology: VAR (for Impulse Responses), Linear Regressions for Forecasting

3.3.2. COLLECTING CLOTHING DATA FROM THE INTERNET

Author(s): Griffioen, R., de Haan, J., Willenborg, L. (Statistics Netherlands)

Brief Description: The paper is concerned with the usability of online apparel prices for CPI analysis. This study falls in the web scraping category and reports the findings and difficulties of online price collection during a two years period. The advantages of web scraping clothing prices are: (i) online price collection is cheaper than price collection in physical stores, (ii) given the relatively low collection costs, there is an incentive to rely on 'big data' and circumvent small sample problems (e.g. high sampling variance), (iii) the quality of online data tends to be very good and (iv) some item characteristics can be easily observed. The main disadvantages of conducting a data collection of this type are: (i) website changes can lead to data problems, (ii) the choice of web scraping strategy can affect the information collected and item representativeness, (iii) weighting information is unavailable, and (iv) the available information on characteristics may be insufficient, depending on the need for quality adjustment.

Data: CPI, Clothing Prices

3.3.3. USING WEB SCRAPED DATA TO CONSTRUCT CONSUMER PRICE INDICES

Author(s): Breton, R., Swier, N., O'Neil, R. (Office for National Statistics (ONS), UK)

Brief Description: The purpose of this paper is to provide an overview of ONS research into the potential of using web scraped data for consumer price statistics. The research covers the collection, manipulation and analysis of web scraped data. As before, the main benefits of web scraped data are identified as follows: (i) reduced collection costs, (ii) increased coverage (i.e. more basket items), (iii) increased frequency, (iv) production of new or complimentary outputs/indices, and (v) improved ability to respond to new challenges. ONS use web scraped data to calculate price indices which: (i) expand the number of items used, (ii) expand the number of days considered, and (iii) expand both the number of items and days considered. The construction of this sort of indices can be useful for economists and policymakers.

Data: Inflation, CPI, RPI, Web Scraping

3.4. Surveys

3.4.1. NEWS AND NARRATIVES IN FINANCIAL SYSTEMS: EXPLOITING BIG DATA FOR SYSTEMIC RISK ASSESSMENT

Author(s): Nyman, R. (UCL), Gregory, D. (Bank of England), Kapadia, S. (Bank of England), Smith, R. (UCL), Tuckett, D. (UCL)

Brief Description: Nyman *et al* (2014a) investigate ways to use big data in systemic risk management. News and narratives are key drivers behind economic and financial activity. Their news data consists of (i) daily comments on market events, (ii) weekly economic research reports and (iii) Reuters news. Machine Learning and Principal Components are included in the methodology in order to calculate the consensus indexes based on the above sources. Their findings include that weekly economic research reports could potentially forecast the Michigan Consumer Index and daily comments on market events could potentially forecast market volatility.

Data: Broker Reports, Bank of England Internal Reports, Reuters News Archive

Methodology: Emotion Dictionary Words

3.4.2. BIG DATA AND ECONOMIC FORECASTING: A TOP-DOWN APPROACH USING DIRECTED ALGORITHMIC TEXT ANALYSIS

Author(s): Nyman, R. (UCL), Smith, R. (UCL), Tuckett, D. (UCL)

Brief Description: Nyman *et al* (2014b) introduce the Directed Algorithmic Text Analysis and show that this methodology can improve considerably on consensus economic forecasts of the Michigan Consumer Index Survey. The approach is based upon searching for particular terms in textual data bases. In contrast to econometric approaches, their methodology is based upon a theory of human decision making under radical uncertainty. The search is directed by the theory. This direction dramatically reduces the dimensionality of the search. They look for words which convey a very limited number of emotions. As in other approaches, they also use regression analysis, but the choice of variables comes from the underlying theory of decision making.

Data: Text Data, Michigan Cons. Index

Methodology: Linear Regression, Building regressors upon a theory

3.5. Financial variables

3.5.1. HOW TO MEASURE THE QUALITY OF FINANCIAL TWEETS

Author(s): Cerchiello, P., Giudici, P. (University of Pavia)

Brief Description: Apart from Google Trends, economic and financial researchers have also started using Twitter posts about various economics and financial news. Cerchiello and Giudici (2014) investigate how the quality of financial tweets can be measured. They suggest that a Google Scholar 'h-index' type measure allows for improved nowcasting of financial variables using Twitter texts. The Twitter users are ranked according to their 'h-index' and confidence intervals are constructed to decide whether top Twitter users are significantly different. Twitter data are collected and R language's TwitteR package is adopted. Their methodology lies in the field of loss data modelling.

Data: Twitter data

Methodology: Proposal of an h-index (similar to Google Scholar) using tweets.

3.5.2. NEWS VERSUS SENTIMENT: COMPARING TEXTUAL PROCESSING APPROACHES FOR PREDICTING STOCK RETURNS

Author(s): Heston, S. L. (University of Maryland), Sinha, N. R. (Board of Governers of the Federal Reserve System)

Brief Description: Heston and Sinha (2014), even though it is not a macroeconomics application, use a dataset of over 900,000 news stories to test whether news can predict stock returns. They find that firms with no news have distinctly different average future returns than firms with news. Confirming previous research, daily news predicts stock returns for only 1-2 days. But weekly news predicts stock returns for a quarter year. Positive news stories increase stock returns quickly, but negative stories have a long-delayed reaction.

Data: News stories, Stock Returns

Methodology: News Sentiment, Regressions (Cross-Sectional)

3.6. Other studies

3.6.1. STATISTICAL MODEL SELECTION WITH 'BIG DATA'

Author(s): Doornik, J. A., Hendry, D. F. (Institute for New Economic Thinking, Economics Department, Oxford University)

Brief Description: big data offers potential benefits for statistical modelling, but confronts problems like an excess of false positives, mistaking correlations for causes, ignoring sampling biases, and selecting by inappropriate methods. Doornik and Hendry (2015) consider the many important requirements when searching for a data-based relationship using big data. Paramount considerations include embedding relationships in general initial models, possibly restricting the number of variables to be selected over by non-statistical criteria (the formulation problem), using good quality data on all variables, analysed with tight significance levels by a powerful selection procedure, retaining available theory insights (the selection problem) while testing for relationships being well specified and invariant to shifts in explanatory variables (the evaluation problem), using a viable approach that resolves the computational problem of immense numbers of possible models.

Data: Artificial Data

Methodology: Multiple Testing, Autometrics, Lasso

3.6.2. MEASURING ECONOMIC POLICY UNCERTAINTY

Author(s): Baker, S. R. (Northwestern), Bloom, N. (Stanford), Davis, S. J. (The University of Chicago)

Brief Description: This paper develops a new index of economic policy uncertainty (EPU) based on newspaper coverage frequency. Several types of evidence — including human readings of 12,000 newspaper articles — indicate that this index proxies for movements in policy-related economic uncertainty. The index spikes near tight presidential elections, Gulf Wars I and II, the 9/11 attacks, the failure of Lehman Brothers, the 2011 debt-ceiling dispute and other major battles over fiscal policy. Using firm-level data, they find that policy uncertainty raises stock price volatility and reduces investment and employment in policy-sensitive sectors like defence, healthcare, and infrastructure construction.

Data: Newspaper data with selected keywords such as: regulation, budget, spending, deficit, tax etc.

Methodology: Vector Autoregressions

4. Types of big data by dominant dimension

In this Section we focus on numerical data only, which can either be the original big data or the result of a transformation of unstructured data, and focus on the specific dimensions of the dataset, which are also important to identify the required econometric techniques. This classification is also relevant as will deal, respectively, with how to transform unstructured data into structured numerical data, and how to pre-treat big numerical data to eliminate outliers, recurrent temporal patterns and other data irregularities. It is also a convenient classification for the review of econometric techniques for numerical big data.

Following, e.g. Doornik and Hendry (2015), and as mentioned in the Introduction, we can distinguish three main types of big data: Fat (big cross-sectional dimension, N, small temporal dimension, T), Tall (small N, big T), or Huge (big N, big T). We discuss their main features in the following subsections.

4.1. Fat datasets

Fat datasets are characterized by a huge cross-sectional dimension but a limited temporal dimension, often it is just $T = 1$. Large cross-sectional databases (for example, coming from census or administrative records or medical analyses) fall into this category, which is not so interesting from an economic nowcasting point of view, unless either T is also large enough or the variables are homogeneous enough to allow proper econometric model estimation (e.g. by means of panel methods) and nowcast evaluation. However, Fat datasets can be of interest in many other applications of big data, both inside official statistics, e.g. for surveys construction, and outside, e.g. in marketing or medical studies.

As the collection of big data started only rather recently, Fat datasets are perhaps the most commonly available type. Actually, statistical methods for big data are mainly meant for Fat datasets, e.g. those developed in the machine learning literature, as they only require a large cross-section of i.i.d. variables.

When a (limited) temporal dimension is also present, panel estimation methods are typically adopted in the economic literature, but factor based methods can be also applied (possibly in their 'sparse' version). Classical estimation methods are not so suitable, as their finite (T) sample properties are generally hardly known, while Bayesian estimation seems more promising, as it can easily handle a fixed T sample size and, with proper priors, also a large cross-sectional dimension.

4.2. Tall datasets

Tall datasets have a limited cross-sectional (N) dimension but a big temporal dimension, T. This is for example the case with tick by tick data on selected financial transactions, and indeed high frequency data have been long used in financial econometrics. Most of the data types considered in Section 2, if aggregated somewhat across the cross-sectional dimension, could be of the Tall type. For example, total daily cash withdrawals from ATM machines, second by second phone traffic in a given geographical location, hourly cloud coverage of a set of locations resulting from satellite images, or second by second volume of internet searches for a specific keyword. In all these cases T is indeed very large in the original time scale, say seconds, but it should be considered whether it is also large enough in the time scale of the target macroeconomic variable of the nowcasting exercise, say quarters. In other words, for nowcasting applications, it is not enough to have a huge T in terms of seconds if T is instead small when measured in months or quarters, as the evaluation will be typically conducted in the low frequency of the target macroeconomic variable ([39]).

Tall datasets at very high frequency are not easily obtained, as they are generally owned by private companies. An interesting exception is represented by textual data. For example, using a web-scraper, it is possible to download all the financial news articles appearing on the Reuters terminal over the past 10 years on a daily basis. Next, using proper software for textual analysis, it is possible to transform the unstructured textual data into daily numerical sentiment indexes, which can then later be used as coincident or leading indicators of economic conditions.

Tall datasets generally require substantial pre-treatment, as indicators typically present particular temporal structures (related, e.g. to market micro-structure in the case of financial indicators) and other types of irregularities, such as outliers, jumps and missing observations. Apart from these issues, when N is small and T is large, classical time series econometric models and methods can be largely used, even though specific features such as changing volatility and possibly parameter time variation should be carefully assessed.

The possible frequency mismatch between (low frequency) target and (high frequency) indicators should be also properly handled, and MIDAS type approaches are particularly promising in a nowcasting context. Moreover, the choice of the proper high frequency scale

([39]) Also, the classification of variables such as financial indicators or commodity prices depends on the sampling frequency. If tick by tick data are collected, then the T dimension dominates but with daily, and even more with monthly, data the N dimension can become dominant, even more so when indicators for many countries are grouped together.

(extent of temporal aggregation) should be also considered, as in general there can be a trade-off between timeliness and precision of the signal.

4.3. Huge datasets

Huge datasets are characterised by very large *N* and *T*. This is perhaps the most interesting type of data in a nowcasting context, and all the data types surveyed in Section 2 are potentially available in Huge format. For example, all the POS financial transactions in a country over a given temporal period, the activity records of all the subscribers to a mobile operator, or all scanned prices for all products in a 24/7 supermarket chain.

In practice, however, unfortunately Huge datasets are not so often available because, as we mentioned, big data collection started only recently, while the collection of the target economic indicators started long ago, generally as far back as the 1950s or 1960s for many developed countries. Moreover, even when available, public access to the Huge data is often not granted, only some cross-sectionally and temporally aggregated measures are made public. Google Trends, publicly available weekly summaries of a huge number of specific search queries in Google, are perhaps the best example in this category, and not by chance the most commonly used indicators in economic nowcasting exercises (see Sections 2 and 3).

Contrary to basic econometrics and statistics, in Huge datasets both *T* and *N* diverge, and proper techniques must be adopted to take this feature into account at the level of model specification, estimation and evaluation. For example, in principle it is still possible to consider information criteria such as BIC or AIC for model specification (indicator selection for the target variable in the nowcasting equation), although it is the case that modifications may be needed to account for the fact that *N* is comparable or larger than *T*, as opposed to much smaller as assumed in the derivations of information criteria. Further, in the case of a linear regression model with *N* regressors, 2^N alternative models should be compared, which is not computationally feasible when *N* is very large, so that efficient algorithms that only search specific subsets of the 2^N possible models have been developed. Moreover, the standard properties of the OLS estimator in the regression model are derived assuming that *N* is fixed and (much) smaller than *T*. Some properties are preserved, under certain conditions, also when *N* diverges but empirically the OLS estimator does not perform well due to collinearity problems that require a proper regularization of the second moment matrix of the regressors. This is in turn relevant for nowcasting, as the parameter estimators are used to construct the nowcast (or forecast). As a result of these problems for OLS, a number of regularisation or penalisation methods have been suggested.

An early approach, referred to as Ridge regression, uses shrinkage to ensure a well behaved regressor sample second moment matrix. More recently, other penalisation methods have been developed. A prominent example is LASSO where a penalty is added to the OLS objective function in the form of the sum of the absolute values of the coefficients. Many related penalisation methods have since been proposed and analysed.

As an alternative to variable selection, the indicators could be summarized by means of principal components (or estimated factors) or related methods such as dynamic principal components or partial least squares. However, standard principal component analysis is also problematic when *N* gets very large, but fixes are available, such as the use of sparse principal component analysis.

Also, rather than selecting or summarizing the indicators, they could be all inserted in the nowcasting regression but imposing tight priors on their associated coefficients, which leads to specific Bayesian estimators.

Finally, in the case of both Fat and Huge datasets, it could be interesting to split them into smaller ones with a limited *N* dimension, apply standard techniques to each sub-dataset, and then re-group the results. For example, one could extract principal components from each sub-dataset, group all of them, and extract their principal components, which provide a summary of the entire big dataset.

Some of these methods can be also of interest for the analysis of Fat datasets, in particular those related to indicator selection or Bayesian estimation, and can therefore be used also outside the nowcasting context.

4.4. Overall conclusions

Huge numerical datasets, possibly coming from the conversion of even larger but unstructured big data, pose substantial challenges for proper econometric analysis, but also offer potentially the largest informational gains for nowcasting. Fat or Tall datasets can be also relevant in specific applications, for example for microeconomic or marketing studies in the case of Fat data or for financial studies in the case of Tall data. Tall data resulting from targeted textual analysis could be also relevant for macroeconomic nowcasting. Table 3 attempts to classify — generally — the ten big data categories as fat, tall or huge. However, the researcher must notice that the classification depends heavily on the application needs for big data.

Table 3: Doornik and Hendry (2015) classification

Type	Fat	Tall	Huge
Financial Markets			X
Electronic Payments			X
Mobile Phones	X		X
Sensor Data / IoT	X		X
Satellite Images		X	
Scanner Prices	X		X
Online Prices	X		X
Online Search		X	

Before drawing a definite conclusion, it is however important to consider more generally the pros and cons of big data for macroeconomic nowcasting or flash estimation, as this can provide additional details for an informed decision, which is particularly relevant in an institutional context. We tackle this issue in the next section.

5. Big data: pros and cons for macroeconomic nowcasting

In this Section, as mentioned above, we evaluate the general advantages and disadvantages of big data for macroeconomic nowcasting or flash estimation.

It must be noted that the classification depends heavily on the application needs for big data. Also, the last three types could be 'Huge' if disaggregated data is available.

For the sake of exposition, it is convenient to distinguish data related issues and more methodological potential problems with the use of big data. We consider them, in turn, in the following two subsections, both in general terms and specifically for nowcasting.

5.1. Data issues

A first issue concerns data availability. As it is clear from the data categorization described in Section 2, most data pass through private providers and are related to personal aspects. Hence, continuity of data provision could not be guaranteed. For example, Google could stop providing Google Trends, or at least no longer make them available for free. Or online retail stores could forbid access to their websites to crawlers for automatic price collection. Or individuals could extend the use of softwares that prevent tracking their internet activities, or tracking could be more tightly regulated by law for privacy reasons.

Continuity of data availability is more an issue for the use of internet data in official statistics than for a pure nowcasting purpose, as it often happens in nowcasting that indicators become unavailable or no longer useful and must be replaced by alternative variables. That said, continuity and reliability of provision are important elements for the selection of a big data source.

Another concern related to data availability is the start date, which is often quite recent for big data, or the overall number of temporal observations in low frequency (months/quarters), which is generally low, even if in high frequency or cross-sectionally there can be thousands of observations. A short temporal sample is problematic as the big data based indicators need to be related to the target low frequency macroeconomic indicators and, without a long enough sample, the parameter estimators can be noisy and the ex-post evaluation sample for the nowcasting performance too short. On the other hand, several informative indicators, such as surveys and financial condition indexes, are also only available over short samples, starting after 2000, and this feature does not prevent their use.

A second issue for internet based big data is related to the 'digital divide', the fact that a sizable fraction of the population still has no or limited internet access. This implies that the available data are subject to a sample selection bias, and this can matter for their use. Suppose, for example, that we want to nowcast unemployment at a disaggregate level, either by age or by regions. Internet data relative to older people or people resident in poorer regions could lead to underestimation of their unemployment level, as they have relatively little access to internet based search tools.

For nowcasting, the suggestion is to carefully evaluate the presence of a possible (keyword) selection bias, but this is likely less relevant when internet based data are combined with more traditional indicators and are therefore used to provide additional marginal rather than basic information. There can also be other less standard big data sources for which the digital divide can be less relevant. For example, use of mobile phones is quite widespread and mobility of their users, as emerging from calls and text messages, could be used to measure the extent of commuting, which is in turn typically related to the employment condition.

A third issue is that both the size and the quality of internet data keeps changing over time, in general much faster than for standard data collection. For example, applications such as Twitter or WhatsApp were not available just a few years ago, and the number of their users increased exponentially, in particular in the first period after their introduction. Similarly, other applications can be gradually dismissed or used for different uses. For example, the fraction of goods sold by Ebay through proper auctions is progressively declining over time, being replaced by other price formation mechanisms.

This point suggests that the relationship between the target variable and the big data (as well as that among the elements of the big data) could be varying over time, and this is a feature that should be properly checked and, in case, taken into consideration at the modelling stage.

A fourth issue, again more relevant for digital than standard data collection, is that individuals or businesses could not report truthfully their experiences, assessments and opinions. For example, some newspapers and other sites conduct online surveys about the feelings of their readers (happy, tired, angry, etc.) and one could think of using them, for example, to predict election outcomes, as a large fraction of happy people should be good for the ruling political party. But, if respondents are biased, the prediction could be also biased, and a large fraction of non-respondents could lead to substantial uncertainty.

As for the case of the digital divide, this is less of a problem when the internet data are complementary to more traditional information, such as phone or direct interviews, or indicators of economic performance.

A fifth issue is that data could not be available in a numerical format, or not in a directly usable numerical format. A similar issue emerges with standard surveys, for example on economic conditions, where discrete answers from a large number of respondents have to be somewhat summarized and transformed into a continuous index. However, the problem is more common and relevant with internet data.

A related issue is that the way in which the big data measure a given phenomenon is not necessarily the same as in official statistics, given that the data are typically the by-product of different activities. A similar issue arises with the use of proxy variables in econometric studies, e.g. measures of potential output or inflation expectations. The associated measurement error can bias the estimators of structural parameters but is less of a problem in a forecasting context, unless the difference with the relevant official indicator is substantial.

Clearly, the collection and preparation of big data based indicators is far more complex than that for standard coincident and leading indicators, which are often directly downloadable in ready to use format from the web through statistical agencies or data providers. The question is whether the additional costs also lead to additional gains, and to what extent, and this is mainly an empirical issue. The literature review we have presented suggests that there seem to be cases where the effort is worthwhile.

A final issue, again common also with standard data but more pervasive in internet data due to their high sampling frequency and broad collection set, relates to data irregularities (outliers, working days effects, missing observations, etc.) and presence of seasonal / periodic patterns, which require properly de-noising and smoothing the data.

As for the previous point, proper techniques can be developed and the main issue is to assess their cost and effectiveness. The initial selection of keywords or a universe of variables is a qualitative issue and must be addressed by the researcher before the empirical investigation. Then, selecting the number of variables or keywords in the underlying universe is an empirical issue.

We have so far focused on the possible cons of big data, as the pros have been widely emphasized both in the general press and in more specialized journals and meetings. Among the main advantages in a nowcasting context, we believe that big data provide potentially relevant complementary information with respect to standard data, being based on rather different information sets.

Moreover, big data are timely available and, generally, they are not subject to subsequent revisions, all relevant features for potential coincident and leading indicators of economic activity.

Finally, big data could be helpful to provide a more granular perspective on the indicator of interest, both in the temporal and in the cross-sectional dimensions. In the temporal dimension, they can be used to update nowcasts at a given frequency, such as weekly or even daily, so that the policy and decision makers can promptly update their actions according to the new and more precise estimates. In the cross-sectional dimension, big data could provide relevant information on units, such as regions or sectors, not fully covered by traditional coincident and leading indicators.

Overall, our suggestion is to take a pragmatic approach that balances potential gains and costs from the use of big data for nowcasting. Hence, for a specific target variable of interest, such as GDP growth or unemployment, it is worth assessing the marginal gains of big data based indicators that are rather promptly available (such as Google Trends or other variables used in previous studies and made publicly available) with respect to more standard indicators based on soft and hard data.

5.2. Methodological issues

As Hartford (2014) put it: 'big data' has arrived, but big insights have not. The challenge now is to solve new problems and gain new answers — without making the same old statistical mistakes on a grander scale than ever.'

The statistical mistakes he refers to are well summarized by Doornik and Hendry (2015): '… an excess of false positives, mistaking correlations for causes, ignoring sampling biases and selecting by inappropriate methods.'

An additional critic is the 'big data hubris', formulated by Lazer *et al* (2014): 'big data hubris' is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis'. They also identify 'Algorithm Dynamics' as an additional potential problem, where algorithm dynamics are the changes made by engineers to improve the commercial service and by consumers in using that service. Specifically, they write: 'All empirical research stands on a foundation of measurement. Is the instrumentation actually capturing the theoretical construct of interest? Is measurement stable and comparable across cases and over time? Are measurement errors systematic?'

Yet another caveat comes from one of the biggest fans of big data. Hal Varian, Google's chief economist, in a 2014 survey wrote: 'In this period of big data it seems strange to focus on sampling uncertainty, which tends to be small with large datasets, while completely ignoring model uncertainty which may be quite large. One way to address this is to be explicit about examining how parameter estimates vary with respect to choices of control variables and instruments.'

In our nowcasting context, we can therefore summarize the potential methodological issues about the use of big data as follows.

First, do we get any relevant insights? In other words, can we improve nowcast precision by using big data? As we mentioned in the previous subsection, this is mainly an empirical issue and, from the studies reviewed in the previous sections, it seems that for some big data and target variables this is indeed the case.

Second, do we get a big data hubris? Again as anticipated, we think of big data based indicators as complements to existing soft and hard data-based indicators, and therefore we do not get a big data hubris (though this is indeed the case some-times even in some of the nowcasting studies, for example those trying to anticipate unemployment using only Google Trends).

Third, do we risk false positives? Namely, can we get some big data based indicators that nowcast well just due to data snooping? This risk is always present in empirical analysis and is magnified in our case by the size of the dataset since this requires the consideration of many indicators with the attendant risk that, by pure chance, some of them will perform well in sample. Only a careful and honest statistical analysis can attenuate this risk. In particular, as mentioned, we suggest comparing alternative indicators and methods over a training sample, selecting the preferred approach or combine a few of them, and then test if they remain valid in a genuine (not previously used) sample.

Fourth, do we mistake correlations for causes? Again, this is a common problem in empirical analysis and we will not be immune for it. For example, a large number of internet searches for 'filing for unemployment' can predict future unemployment without, naturally, causing it. This is less of a problem in our nowcasting context, except perhaps at the level of economic interpretation of the results.

Fifth, do we use the proper econometric methods? Here things are more complex because when the number of variables N is large we can no longer use standard methods and we have to resort to more complex procedures. Some of these were developed in the statistical or machine learning literatures, often under the assumption of i.i.d. observations. As this assumption is likely violated when nowcasting macroeconomic variables, we have to be careful in properly comparing and selecting methods that can also handle correlated and possibly heteroskedastic data. This is especially the case since these methods are designed to provide a good control of false positives, but this control depends crucially on data being i.i.d. To give an example, it is well known that exponential probability inequalities, that form the basis of most methods that control for false positives, have very different and weaker bounds for serially correlated data leading to the need for different choices for matters like tuning parameters used in the design of the methods. Overall, as we will see, a variety of methods are available, and they can be expected to perform differently in different situations, so that also the selection of the most promising approach is mainly application dependent.

Sixth, do we have instability due to Algorithm Dynamics or other causes (e.g. the financial crisis, more general institutional changes, the increasing use of internet, discontinuity in data provision, etc.)? Instability is indeed often ignored in the current big data literature, while it is potentially relevant, as we know well from the economic forecasting literature. Unfortunately, detecting and curing instability is complex, even more so in a big data context. However, some fixes can be tried mostly borrowing from the recent econometric literature on handling structural breaks.

Finally, do we allow for variable and model uncertainty? As we will see, it is indeed important to allow for both variable uncertainty, by considering various big data based indicators rather than a single one, and for model uncertainty, by comparing alternative procedures and then either selecting or combining the best performing ones. Again, all issues associated with model selection and uncertainty are likely magnified due to the fact that large data also allow for bigger classes of models to be considered and model selection methods, such as information criteria, may need modifications in many respects.

6. Conclusions

In this paper we have provided an overview of the types of available big data potentially useful for nowcasting macroeconomic variables, focusing on: (i) financial markets data, (ii) electronic payments data, (iii) mobile phones data, (iv) sensor data, (v) satellite images data, (vi) scanner prices data, (vii) online prices data, (viii) online search data, (ix) textual data, and (x) social media data. We have also discussed the main empirical applications in a nowcasting/forecasting context, either based on a specific type of big data or with a specific macroeconomic indicator as a target. And we have classified big data based on the relative size of their cross-sectional and temporal dimensions, which is also informative for the proper econometric techniques to be used. Finally, we have discussed the a priori pros and cons of the use of big data, focusing on both data and methodological issues. While there are many pros, and these have been emphasized in many contexts, there are also a few cons that should not be ignored.

Table 4 proposes a summary classification of big data, based on 13 key features that we have identified according to the previous considerations: source, provider, availability, continuity, type, size and sample, meta data, feature, frequency, pre-treatment, link with target, previous use, and required econometrics. For example, online search data are:

- available from private providers (Google, Microsoft, etc.);
- easily available in aggregated form (freely downloadable from internet);
- continuity of provision cannot be guaranteed;
- their type is already numerical (time series);
- the size is limited and user dependent (dependent on the number of inserted search queries or selected number of correlates);
- the sample period (since 2004) is long enough to guarantee a rather reliable econometric analysis and evaluation;
- meta data are not easily available;
- raw data and details on the aggregation process are generally not publicly available;
- the frequency is generally weekly or higher (with daily data possibly available upon request), the release is timely, and the data is not further revised;
- there is not much information about pre-treatment but further treatment to remove outliers and temporal patterns is likely needed;

- the link with specific macroeconomic indicators has been established in previous empirical analyses
- the required econometrics is standard;

Table 4: Taxonomy of big data

Source	**(Social Networks/Traditional Business Systems/Internet of Things)**
Provider	(Private/Public, National/International,…)
Availability	(Open/Restricted, Free/Fee Based, Standard/Customized, …)
Continuity	(Yes/No/Uncertain)
Type	(Pictures/Binary/Text/Numerical…)
Size and Sample	(Gb and/or N/T if numerical)
Meta Data	(Available or not)
Features	(Raw/Transformed/Selected/Aggregated/…)
Frequency	(Low/high, continuous or undetermined, …)
Pre-Treatment	(Outliers, missing observations, seasonal adjustment, measurement errors, …)
Link with Target	(Definition/Economic theory/Empirical/…)
Previous use	(No/Yes, Where/What)
Econometrics	(Standard Methods/big data Specific Methods)

Table 5: Evaluation grid for use of big data for macroeconomic forecasting

A priori assessment of potential usefulness of Big data	Are available nowcasts/forecasts for the target variable biased or inefficient? Is this due to missing information?
	Could timeliness, frequency of release and extent of revision be improved with additional information?
	Is the required missing or additional information available from some type of big data? Is it not available from traditional sources?
	Are there any studies showing the usefulness of big data for forecasting a similar target? How large are the gains?
	Is the temporal dimension of the big data long and homogeneous enough to allow proper evaluation of the resulting nowcasts/forecasts?
Big data Sources	Is big data directly available, perhaps with some investment in data collection, or is a provider needed?
	If a provider is needed, is it public or private? Is it national or international?
	Is access free or fee based? Can it be customized?
	Can data production and access be expected to continue in the future?
	Are there any confidentiality and/or legal issues?
Big data Features	How big is the relevant big data? Does it require specific hardware and software for storing and handling?
	Is it in numerical or non-numerical (text / pictures / binary / etc) format?
	If non-numerical, can it be transformed into numerical format? Does the transformation require specific software? How expensive in terms of time and resources is it expected to be?
	If numerical, is it accessible in clean format or does it requires pre-treatment? How expensive in terms of time and resources is pre-treatment expected to be?
	Is the corresponding big data available for all the units of interest (e.g. countries, sectors, disaggregation levels, various target variables, etc.)?

Table 5: Evaluation grid for use of big data for macroeconomic forecasting

Big data Quality	Does data production and collection satisfy requirements for use from an official institution?
	Is the underlying sampling scheme sufficient to be representative of the entire relevant population?
	Is Meta Data available? Could data collection be replicated?
	If raw data is not available, is the process generating the available summarized or transformed big data clearly explained and reliable?
Big data Econometrics	If big data is in non-numerical format, is there a reliable mapping into numerical format?
	Can data pre-treatment (e.g. outliers removal, missing values imputation, seasonal and other types of adjustment, filtering etc.) be conducted with standard methods?
	Can econometric methods for large datasets (e.g. principal components, shrinkage regressions, large VARs) be adopted or are big data specific methods required (e.g. sparse principal components, special Bayesian priors, genetic algorithms, methods for infinite dimensional matrices, machine learning, etc.)?
	Is it better to work with the entire big data, if available, and related big data econometrics, or with a proper summary of the big data and standard econometric methods?
	How large and robust are, in the end, the forecasting gains when adding big data to traditional indicators? And how large are the gains in terms of timeliness, frequency of release and extent of revision?

It must be noted that a particular caveat of the above is that a dataset may move across these thirteen dimensions depending how a modelling step is done. For example, online search data requires standard econometrics, however this may or may not be needed depending on the choices of the researchers.

The classification in Table 4, as applied to specific big data types as in the example we have provided, can be then also used to create a grid for assessing the need and potential usefulness of big data for macroeconomic forecasting, in general and in official institutions such as Eurostat, according to the outcome of a careful cost-benefit analysis. The proposed grid is presented in Table 5.

As an example, let us assume that we are considering the use of Google trends (online search data) to nowcast unemployment and, according to the analysis based in Table 4, we have concluded that they are indeed potentially useful. We should then consider issues such as:

- the quality of the current unemployment nowcasts and the need for improvement (which depends on the specific country under analysis);
- the possibility of improving the timeliness, frequency of release and extent of revision of the nowcasts when using the online search data (all features can be potentially improved, as online search data are timely available, at least at weekly frequency and not revised);
- the availability of other traditional sources of information (e.g. business surveys with also questions related to future employment perspectives);
- the availability of previous studies showing the usefulness of the specific selected big data type for unemployment nowcasting and the extent of the reported gains;
- the availability of a long enough sample span for an econometric analysis;
- the availability of a sufficient level of disaggregation, e.g. sectors or regions, or age groups, or qualifications (which is not the case, at least not for the freely available data);
- the representativeness of the underlying sample (which only covers internet users);

- the resources required for the big data collection, storage, treatment and analysis (which are limited in this example);
- the possibility that access to the big data could be discontinued or no longer be free (which could indeed happen);
- the presence of confidentiality or legal issues (which could emerge, in particular with private data used by a public institution);
- the need of big data specific econometric techniques (which is not the case in this example as the big data have been pre-aggregated by the provider);
- the type and extent of the marginal gains emerging from the use of the big data from a proper econometric evaluation (for example, the out of sample increase in the unemployment nowcasts precision but also in their timeliness, frequency, and extent of revision).

Overall, a careful selection of the relevant big data, combined with the use of proper econometric methods to formally handle and analyse their relationship with the target macroeconomic indicator, has substantial potential to sharpen predictions and make them more timely and frequent.

In terms of specific big data types, electronic payments data, scanner price and online price data, online search data, textual data and social media data are all promising for nowcasting.

Acknowledgements

The authors thank two anonymous referees for their helpful and constructive comments. The authors would like to thank their Eurostat colleagues Emanuele Baldacci and Martina Hahn for their support, Fernando Reis and Dan Rieser for their feedback. They thank also Diego Kuonen for the fruitful discussion around the use of Big Data in Official Statistics. They are also grateful to the members of the Working Group Methodology for their constructive comments on earlier version of the paper. A special thanks is also extended to Sabine Bankwitz. The views expressed are those of the authors and do not necessarily reflect the views of their current or former employer.

References

Abramovic, G. (2014), '15 Mind-Blowing Stats About Online Shopping', CMO.com, available at: https://goo.gl/xNZvoE.

Acemoglu, D., T.A. Hassan and A. Tahoun (2014), 'The Power of the Street: Evidence from Egypt's Arab Spring', *NBER Working Paper* No. 20665.

Alesina, A., S. Michalopoulos and E. Papaioannou (2016), 'Ethnic Inequality', *Journal of Political Economy*, 124(2), pp. 428–488.

Andreou, E., E. Ghysels and A. Kourtellos (2015), 'Should Macroeconomic Forecasters Use Daily Financial Data and How?', Journal of Business & Economic Statistics, 31(2), pp. 240–251.

Angelini, E., G. Camba-Mendez, D. Giannone and L. Reichlin (2011), 'Short-Term Forecasts of Euro Area GDP Growth', *The Econometrics Journal*, 14(1), C25–C44.

Antweiler, W. and M.Z. Frank (2004), 'Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards', *The Journal of Finance*, 59(3), pp. 1259–1294.

Aprigliano, V., G. Ardizzi and L. Monteforte (2016), 'Using the payment system data to forecast the Italian GDP', Bank of Italy, Working Paper.

Arias, M., A. Arratia and R. Xuriguera (2013), 'Forecasting with Twitter Data', *ACM Transactions on Intelligent Systems and Technology*, 5(1), p. 8.

Bacchini, F., M.E. Bontempi, R. Golinelli and C. Jona-Lasinio (2017), 'Short- and long-run heterogeneous investment dynamics', *Empirical Economics*, DOI: 10.1007/s00181-016-1211-4.

Baker, S.R., N. Bloom and S.J. Davis (2015), 'Measuring Economic Policy Uncertainty', *NBER Working Paper Series*, Working Paper 21633.

Banbura M,. D. Giannone and L. Reichlin (2011), 'Nowcasting', In *Oxford Handbook on Economic Forecasting*, Clements MP, Hendry DF (eds). Oxford University Press: Oxford.

Banbura, M. and G. Runstler (2011), 'A Look into the Factor Model Black Box: Publication Lags and the Role of Hard and Soft data in Forecasting GDP', *International Journal of Forecasting*, 27, pp. 333–346.

Bangwayo-Skeete, P.F. and R.W. Skeete (2015), 'Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach', *Tourism Management*, 46, pp. 454–464.

Barnett, W.A., M. Chauvet, D. Leiva-Leon and L. Su (2016) 'Nowcasting nominal GDP with the credit-card augmented Divisia monetary aggregates', *MPRA Paper* No. 73246.

Bazzani, A., L. Giovannini, R. Gallotti and S. Rambaldi (2011), 'Now casting of traffic state by GPS data. The metropolitan area of Rome', Conference paper, January 2011, available at https://goo.gl/kcpPhm.

Beiler, J. (2013), 'Noise in GDELT', Personal Blog, October 28, 2013, available at: https://goo.gl/KC9iaJ.

Bendler, J., S. Wagner, T. Brandt and D. Neumann (2014), 'Taming Uncertainty in Big Data: Evidence from Social Media in Urban Areas', *Business & Information Systems Engineering,* 05-2014, pp. 279–288.

Bholat, D., S. Hansen, P. Santos and C. Schonhardt-Bailey (2015), 'Text mining for central banks', *Centre for Central Banking Studies*, (33), pp. 1–19.

Boettcher, I. (2015), 'Automatic Data Collection on the Internet (Web Scraping)', New Techniques and Technologies for Statistics, Eurostat Conference, 9–13 March 2015.

Bollen, J., H. Mao and X. Zeng (2011), 'Twitter mood predicts the stock market', *Journal of Computational Science*, 2(1), pp. 1–8.

Bontempi, M.E., R. Golinelli and M. Squadrani (2016), 'A new index of uncertainty based on internet searches: a friend or a foe of other indicators?', *Quaderni*, Working Paper DSE No 1062.

Breton, R., N. Swiel and R. O'Neil (2015), 'Using Web Scraped Data to Construct Consumer Price Indices', New Techniques and Technologies for Statistics, Eurostat Conference, 9–13 March 2015.

Campbell, J.R. and B. Eden (2014), 'Rigid Prices: Evidence from U.S. Scanner Data' *International Economic Review*, 55(2), pp. 423–442.

Capgemini and BNP Paribas (2016), 'World Payments Report', Available online at https://goo.gl/niAkll.

Carlsen, M. and P.E. Storgaard (2010), 'Dankort payments as a timely indicator of retail sales in Denmark', Danmarks National bank, Working Paper 2010-66.

Carriere-Swallow, Y. and F. Labbe (2013), 'Nowcasting with Google Trends in an Emerging Market', *Journal of Forecasting*, 32, pp. 289–298.

Cavallo, A. (2013), 'Online and official price indexes: Measuring Argentina's inflation', *Journal of Monetary Economics*, 60, pp. 152–165.

Cavallo, A. (2016), 'Scraped Data and Sticky Prices', *Review of Economics & Statistics*, forthcoming.

Cavallo, A. (2017), 'Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers', *The American Economic Review*, 107(1), pp. 283–303.

Cavallo, A. and R. Rigobon (2016), 'The Billion Prices Project: Using Online Prices for Measurement and Research', *The Journal of Economic Perspectives*, 30(2), pp. 151–178.

Cerchiello, P. and P. Giudici (2014), 'How to Measure the Quality of Financial Tweets', Working Paper, ECB Workshop on using big data for forecasting and statistics, 07-08/04/2014, Frankfurt.

Chamberlain, G. (2010), 'Googling the present', *Economic and Labour Market Review*, Office for National Statistics, 4(12).

Chan, K.Y., S. Khadem, T.S. Dillon, V. Palade, J. Singh and E. Chang (2012), 'Selection of Significant On-Road Sensor Data for Short-Term Traffic Flow Forecasting Using the Taguchi Method', IEEE *Transactions on Industrial Informatics*, 8(2), pp. 255–266.

Chen, X. and W.D. Nordhaus (2011), 'Using luminosity data as a proxy for economic statistics', *PNAS*, 108(21), pp. 8589–8594.

Chevalier, J.A., A.K. Kashyap and P.E. Rossi (2003), 'Why don't prices rise during periods of peak demand? Evidence from scanner data', *American Economic Review*, 93(1), pp. 15–37.

Choi, H. and H. Varian (2009), 'Predicting initial claims for unemployment benefits', *Google Working Paper.*

Choi, H. and H. Varian (2012), 'Predicting the Present with Google Trends', *Economic Record*, 88(1), pp. 2–9.

Conover, M.D., B. Goncalves, J. Ratkiewicz, A. Flammini and F. Menczer (2011), 'Predicting the Political Alignment of Twitter Users', 2011 IEEE International Conference on Privacy, Security, Risk ,and Trust, and IEEE International Conference on Social Computing.

Dal Pozzolo, A., O. Caelen, R.A. Johnson and G. Bontempi (2015), 'Calibrating Probability with Undersampling for Unbalanced Classification', In Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2015.

D'Amuri, F. and J. Marcucci (2012), 'The Predictive Power of Google Searches in Predicting Unemployment', Banca d'Italia Working Paper, 891.

de Haan, J. (2015), 'A Framework for Large Scale Use of Scanner Data in the Dutch CPI', Report from Ottawa Group 14th meeting, International Working Group on Price Indices, Tokyo (Japan), May, 2015.

De Meersman, F., G. Seynaeve, M. Debusschere, P. Lusyne, P. Dewitte, Y. Baeyens, A. Wirthmann, C. Demunter, F. Reis and H.I. Reuter (2016), 'Assessing the Quality of Mobile Phone Data as a Source of Statistics', European Conference on Quality in Official Statistics, Madrid, 31 May–3 June 2016.

Deloitte (2012), 'What is the impact of mobile telephony on economic growth? A report for the GSM association', November 2012.

Deville, P., C. Linard, S. Martin, M. Gilbert, F.R. Stevens, A.E. Gaughan, V.D. Blondel and A.J. Tatem (2014), 'Dynamic population mapping using mobile phone data', PNAS, 111(45), pp. 15888–15893.

Donaldson, D. and A. Storeygard (2016), 'The View from Above: Applications of Satellite Data in Economics', *Journal of Economic Perspectives*, 30(4), pp. 171–198.

Doornik, J. A. and D.F. Hendry (2015), 'Statistical Model Selection with Big Data', *Cogent Economics & Finance*, 3(1), 2015.

Duarte, C., P.M.M. Rodrigues and A. Rua (2016), 'A mixed frequency approach to forecast private consumption with ATM/POS data', Banco de Portugal, Working Paper 1-2016.

Ebner, J. (2015), 'How Sensors will Shape Big Data and the Changing Economy', Dataconomy, published online: January 27, 2015, available at https://goo.gl/9h3Zf8.

Eckley, P. (2015), 'Measuring economic uncertainty using news-media textual data', MPRA Paper No. 69784.

Elvidge, C.D., P.C. Sutton, T. Ghosh, B.J. Tuttle, K.E. Baugh, B. Bhaduri and E. Bright (2009), 'A global poverty map derived from satellite data', Computers & Geosciences, 35, pp. 1652–1660.

Ericsson, N.R. (2015), 'Eliciting GDP Forecasts from the FOMC's Minutes Around the Financial Crisis', International Finance Discussion Papers, Board of Governors of the Federal Reserve System, Working Paper 1152.

Ericsson, N.R. (2016), 'Predicting Fed Forecasts', IFDP Notes, Board of Governors of the Federal Reserve System, February 12, 2016, available at: https://goo.gl/nOl77h.

Esteves, P.S. (2009), 'Are ATM/POS data relevant when nowcasting private consumption?', Banco de Portugal, Working Paper 25-2009.

Fernandez-Ares, A.J., A.M. Mora, M.G. Arenas, P. de las Cuevas, P. Garcia-Sanchez, G. Romero, V. Rivas, P.A. Castillo and J.J. Merelo (2016), 'Nowcasting Traffic', Working Paper.

Ferreira, P. (2015), 'Improving Prediction of Unemployment Statistics with Google Trends: Part 2', Eurostat Working Paper.

Florida, R. (2014), 'The Economic Data Hidden in Nighttime Views of City Lights', Citylab, published online: May 29, 2014, available at https://goo.gl/uaOYdV.

Galbraith, J.W and G. Tkacz (2007), 'Analyzing Economic Effects of Extreme Events using Debit and Payments System Data', *CIRANO Scientific Series*, Working Paper 2011s-70.

Galbraith, J.W and G. Tkacz (2015), 'Nowcasting GDP with electronic payments data', European Central Bank, Working Paper No 10 / August 2015.

Giannone, D., L. Reichlin and D. Small (2008), 'Nowcasting: The Real-Time Informational Content of Macroeconomic Data', *Journal of Monetary Economics*, 55, pp. 665–676.

Ginsberg, J., M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski and L. Brilliant (2009), 'Detecting influenza epidemics using search engine query data', Nature, 457, pp. 1012–1014.

Graham, D. (2016), 'How the Internet of Things Changes Big Data Analytics', DataInformed, published online: August 9, 2016, available at https://goo.gl/M22uje.

Griffioen, R., J. de Haan and L. Willenborg (2014), 'Collecting Clothing Data from the Internet', Statistics Netherlands Technical Report.

Griswold, A. (2014), 'Can Twitter Predict Economic Data?', Slate, Moneybox: A blog about business and economics, April 4, 2014, Available at: https://goo.gl/DHXOJR.

Heaven, D. (2013), 'World's largest events database could predict conflict', *New Scientist*, May 8, 2013, available at: https://goo.gl/U0VOLI.

Henderson, J.V., A. Storeygard and D.N. Weil (2011), 'A Bright Idea for Measuring Economic Growth', *American Economic Review: Papers & Proceedings*, 101(3), pp. 194–199.

Henderson, J.V., A. Storeygard and D.N. Weil (2012), 'Measuring Economic Growth from Outer Space', *American Economic Review*, 102(2), pp. 994–1028.

Heston, S.L. and N.R. Sinha (2014), 'News versus Sentiment: Comparing Textual Processing Approaches for Predicting Stock Returns', Working Paper.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), 'Scanner data, time aggregation and the construction of price indexes', *Journal of Econometrics*, 111, pp. 24–35.

Jones, J. (2010), 'Nowcasting Applications of GPS Water Vapour Networks', UK Met Office, presentation.

Kapetanios, G., M. Marcellino and F. Papailias (2016), 'Big Data and Macroeconomic Nowcasting', Eurostat Working Paper, ESTAT No 11111.2013.001-2015.278.

Kearns, J. (2015), 'Cheap orbiting cameras and big-data software reveal hidden secrets of life down below', Bloomberg Business Week, published online: July 9, 2015, Available at https://goo.gl/fjhPnm.

Keola, S., M. Andersson and O. Hall (2015), 'Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth', *World Development*, 66, pp. 322–334.

Koop, G. and L. Onorante (2013), 'Macroeconomic Nowcasting Using Google Probabilities', European Central Bank Presentation.

Lansdall-Welfare, T., V. Lampos and N. Cristianini (2012), 'Nowcasting the mood of nation', Significance, *The Royal Statistical Society*, August 2012.

Levenberg, A., S. Pulman, K. Moilanen, E. Simpson and S. Roberts (2014), 'Predicting Economic Indicators from Web Text Using Sentiment Composition', *International Journal of Computer and Communication Engineering*, 3(2), pp. 109–115.

Lloyd, T.A., S. McCorriston, C.W. Morgan, E. Poen and E., Zgovu (2012), 'Retailer Heterogeneity and Price Dynamics: Scanner Data Evidence from UK Food Retailing', Working Paper No. 8, *Transparency of Food Pricing*, TRANSFOP.

Lowe, M. (2014), 'Night Lights and ArcGIS: A Brief Guide', *MIT Economics Working Paper.*

Lunnemann, P. and L. Wintr (2011), 'Price Stickiness in the US and Europe Revisited: Evidence from Internet Prices', *Oxford Bulletin of Economics & Statistics*, 73(5), pp. 0305–9049.

Makhazanov, A. and D. Rafiei (2013), 'Predicting Political Preference of Twitter Users', 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, August 25–29, 2013.

Makrehchi, M., S. Shah and W. Liao (2013), 'Stock Prediction Using Eventbased Sentiment Analysis', 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT).

Manyika, J., M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin and D. Aharon (2015), 'The Internet of Things: Mapping the Value beyond the Hype', McKinsey Global Institute, June 2015.

Mao, H., X. Shuai, Y.-Y. Ahn and J. Bollen (2015), 'Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Cote d'Ivoire', EPJ Data Science, 4(15).

Mass, C. (2012), 'Nowcasting: The Promise of New Technologies of Communication, Modeling and Observation', *Bulletin of the American Meteorological Society*, 93(6), pp. 797–809.

Mellander, C., J. Lobo, K. Stolarick and Z. Matheson (2015), 'Night-Time Light Data: A Good Proxy Measure for Economic Activity?' PLoS ONE, 10(10).

Metcalfe, E., T. Flower, T. Lewis, M. Mayhew and E. Rowland (2016), 'Research indices using web scraped price data: clustering large datasets into price indices (CLIP)', Office for National Statistics, Release date: 30 November 2016.

Mittal, A. and A. Goel (2012), 'Stock Prediction Using Twitter Sentiment Analysis', Standford University Working Paper.

Moat, H.S., C. Curme, A. Avakian, D.Y Kenett, H.E. Stanley and T Preis (2013), 'Quantifying Wikipedia Usage Patterns Before Stock Market Moves', *Scientific Reports*, 3:1801, pp. 1–5.

Modugno, M. (2013), 'Now-casting Inflation using High-Frequency Data', *International Journal of Forecasting*, 29, pp. 664–675.

Nyman, R., D. Gregory, S. Kapadia, R. Smith and D. Tuckett (2014a), 'Exploiting Big Data for Systemic Risk Assessment: News and Narratives in Financial Systems', Working Paper, ECB Workshop on using big data for forecasting and statistics, 07-08/04/2014, Frankfurt.

Nyman, R., P. Ormerod, R. Smith and D. Tuckett (2014b), 'Big Data and Economic Forecasting: A Top-Down Approach Using Directed Algorithmic Text Analysis', Working Paper, ECB Workshop on using big data for forecasting and statistics, 07-08/04/2014, Frankfurt.

Pandya, S.S. and R. Venkatesan (2015), 'French Roast: Consumer Response to International Conflict — Evidence from Supermarket Scanner Data', *Review of Economics & Statistics*, 98(1), pp. 42–56.

Papadimitriou, S., J. Sun, C. Faloutos and P.S. Yu (2013), 'Dimensionality Reduction and Filtering on Time Series Sensor Streams', In: *Managing and Mining Sensor Data*, Aggarwal, C.C. (Eds.), Springer, 2013.

Paul, M.J., M. Dredze and D. Broniatwoski (2014), 'Twitter Improves Influenza Forecasting', *PLOS Currents Outbreaks*, October 28, 2014.

Pineiro-Chousa, J., M. Vizcaino-Gonzalez and A.M. Perez-Pico (2016), 'Influence of Social Media over the Stock Market', *Psychology & Marketing*, 34(1), pp. 101–108.

Radzikowski, B. and A. Smietanka (2016), 'Online CASE CPI', First International Conference on Advanced Research Methods and Analytics, CARMA2016.

Ramli, D. (2012), 'Treasury to mine Twitter for economic forecasts', Financial Review, October, 30, 2012, Available at: https://goo.gl/xHUhE2.

Reis, F., P. Ferreira and V. Perduca (2015), 'The Use of Web Activity Evidence to Increase the Timeliness of Official Statistics Indicators', Eurostat Working Paper.

Ricciato, F., P. Widhalm, M. Craglia and F. Pantisano (2015), 'Estimating population density distribution from network-based mobile phone data', European Commission, JRC technical report, EUR 27361 EN.

Rigobon, R. (2015), 'Presidential Address: Macroeconomics and Online Prices', *Economia*, 15(2), pp. 199–213.

Rönnqvist, S. and P. Sarlin (2015), 'Bank Networks from Text: Interrelations, Centrality and Determinants', *Quantitative Finance*, 15(10), pp. 1619–1635.

Ross, A. (2013), 'Nowcasting with Google Trends: a keyword selection method', Fraser of Allander Economic Commentary, 37(2), pp. 54–64.

Schumaker, R.P. and H. Chen (2006), 'Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System', Working Paper.

Silver, M. and S. Heravi (2001), 'Scanner Data and the Measurement of Inflation', *The Economic Journal*, 111, F383-F404.

Smith, N.A. (2010), 'Text-Driven Forecasting', Carnegie Mellon University, Working Paper.

Smith, P. (2016), 'Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data', *Journal of Forecasting*, 35, pp. 263–284.

Smith-Clarke, C., A. Mashhadi and L. Capra (2014), 'Poverty on the cheap: estimating poverty maps using aggregated mobile communication networks', CHI 2014 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 511–520.

Sprenger, T.O., A. Tumasjan, P.G. Sandner and I.M. Welpe (2014a), 'Tweets and Trades: the Information Content of Stock Microblogs', European Financial Management, 20(5), pp. 926–957.

Statistics New Zealand (2014), M*easuring price change for consumer electronics using scanner data*, ISBN 978-0-478-42940-4.

Stock, J. and M. Watson (2002a), 'Forecasting Using Principal Components from a Large Number of Predictors', *Journal of the American Statistical Association*, 297, pp. 1167–1179.

Stock, J. and M. Watson (2002b), 'Macroeconomic Forecasting using Diffusion Indexes', *Journal of Business & Economics Statistics*, 20, pp. 147–162.

Suryadevara, N.K., S.C. Mukhopadhyay, R. Wang and R.K. Rayudu (2013), 'Forecasting the behavior of an elderly using wireless sensors data in a smart home', *Engineering Applications of Artificial Intelligence*, 26, pp. 2641–2652.

Tefft, N. (2011), 'Insights on unemployment, unemployment insurance, and mental health', *Journal of Health Economics*, 30(2), pp. 258–264.

Thorsrud, L.A. (2016), 'Words are the new numbers: A newsy coincident index of business cycles', Norges Bank Working Paper Series, Working Paper 21-2016.

Tumasjan, A., T.O. Sprenger, P.G. Sandner and I.M. Welpe (2010), 'Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment', Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Varian, H. and S. Stephen-Davidowitz (2014), 'Google Trends: A primer for social scientists', Google Working Paper.

Vosen, S. and T. Schmidt (2011), 'Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends', *Journal of Forecasting*, 30, pp. 565-578.

Wang, H., D. Can, A. Kazemzadeh, F. Bar and S. Narayanan (2012), 'A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle', Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 115-120.

Yuan, Q., E.O. Nsoessie, B. Lv, G. Peng, R. Chunara and J.S. Brownstein (2013), 'Monitoring Influenza Epidemics in China with Search Query from Baidu', *PLoS ONE* 8(5), e64323.

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: http://europa.eu/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 electronic mail via: http://europa.eu/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: http://europa.eu

EU publications

You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/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: http://eur-lex.europa.eu

Open data from the EU

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

PDF: KS-GP-17-001-EN-N
PRINT: KS-GP-17-001-EN-C

Publications Office