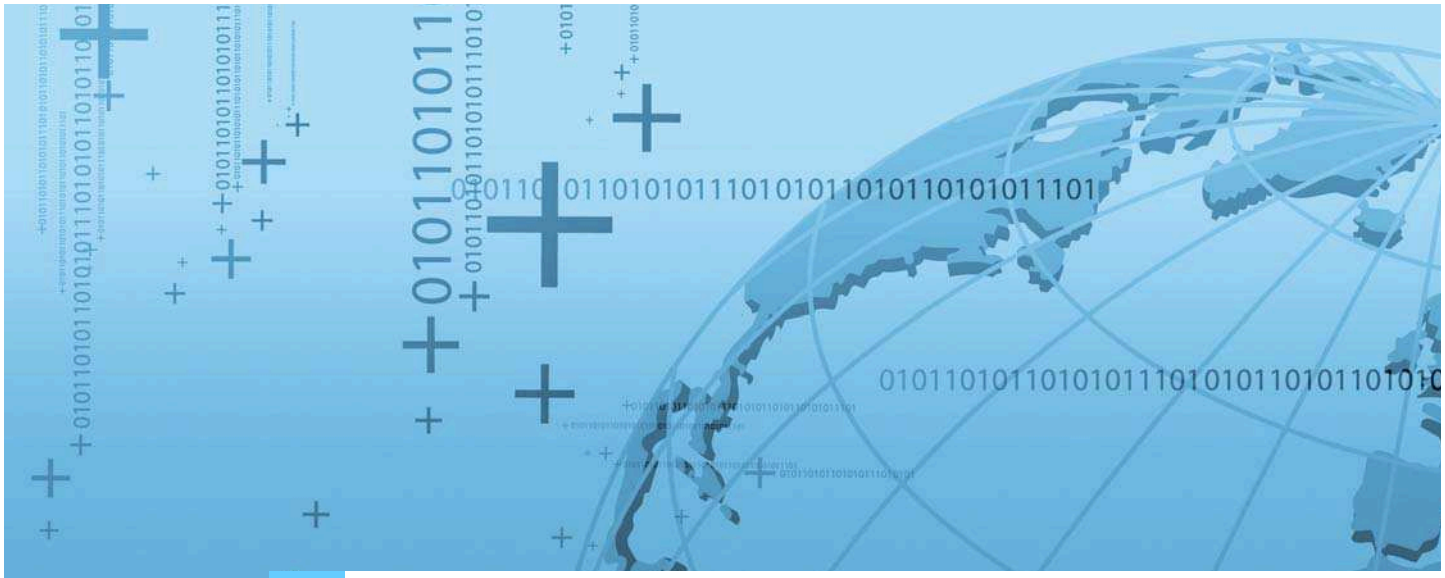


Gross Domestic Product (GDP)

1. Size of GDP
2. GDP growth
3. GDP per capita

GROSS DOMESTIC PRODUCT (GDP)

1. Size of GDP

Gross domestic product (GDP) is the standard measure of the value of final goods and services produced by a country during a period. While GDP is the single most important indicator to capture these economic activities, it is not a good measure of societies' well-being and only a limited measure of people's material living standards. The sections and indicators that follow better address this and other related issues and this is one of the primary purposes of this publication.

Countries calculate GDP in their own currencies. In order to compare across countries these estimates have to be converted into a common currency. Often the conversion is made using current exchange rates but these can give a misleading comparison of the true volumes of final goods and services in GDP. A better approach is to use purchasing power parities (PPPs). PPPs are currency converters that control for differences in the price levels of products between countries and so allow an international comparison of the volumes of GDP and of the size of economies.

Definition

What does gross domestic product mean? "Gross" signifies that no deduction has been made for the depreciation of machinery, buildings and other capital products used in production. "Domestic" means that it is production by the resident institutional units of the country. The products refer to final goods and services, that is, those that are purchased, imputed or otherwise, as: the final consumption of households, non-profit institutions serving households and government; fixed assets; and exports (minus imports).

GDP at market prices can be measured in three different ways:

- as output less intermediate consumption (*i.e.* value added) plus taxes on products (such as VAT) less subsidies on products;
- as the income earned from production, equal to the sum of: employee compensation; the gross operating surplus of enterprises and government; the gross mixed income of unincorporated enterprises; and net taxes on production and imports (VAT, payroll tax, import duties, etc., less subsidies);
- or as the expenditure on final goods and services minus imports: final consumption expenditures, gross capital formation, and exports less imports.

Comparability

All OECD countries now follow the 1993 System of National Accounts, although in some countries, for example in specific areas such as the own account production of software or financial intermediation services (indirectly measured) (FISIM), differences remain, which can impact on comparisons of GDP. The measurement of the non-observed economy (NOE, often referred to as the informal, grey, shadow, economy) can also have an impact on comparability, although for OECD economies, in general, this is not thought to be significant. (See also "Reader's Guide", relating to PPP based comparisons.)

For some countries, the latest year has been estimated by the Secretariat. Historical data have also been estimated for those countries that revise their methodologies but only supply revised data for some years. This estimation process mechanically links the new and old series to preserve growth rates.

Source

- OECD (2009), *National Accounts of OECD Countries 2009, Volume I, Main Aggregates*, OECD Publishing, *http://dx.doi.org/10.1787/na_vol_1-2009-en-fr*.

Online database

- OECD (2009), "Aggregate national accounts: gross domestic product", OECD *National Accounts Statistics* (database), *http://dx.doi.org/10.1787/data-00001-en*.

Further reading

- Lequiller F., N. Ahmad, S. Varjonen, W. Cave and K.-H. Ahn (2003), *Report of the OECD Task Force on Software Measurement in the National Accounts*, OECD Publishing, *http://dx.doi.org/10.1787/334811030426*.
- Ahmad N. (2003), *Measuring Investment in Software*, OECD Publishing, *http://dx.doi.org/10.1787/335303788330*.
- OECD (2002), *Measuring the Non-Observed Economy: A Handbook*, OECD Publishing, *http://dx.doi.org/10.1787/9789264175358-en*.
- Lequiller F. and D. Blades (2007), *Understanding National Accounts*, OECD Publishing, *http://dx.doi.org/10.1787/9789264027657-en*.
- OECD (2000), *System of National Accounts, 1993 – Glossary*, OECD Publishing, *http://dx.doi.org/10.1787/9789264180871-en*.
- UN, OECD, IMF, Eurostat (eds.) (1993), *System of National Accounts 1993*, United Nations, Geneva, *http://unstats.un.org/unsd/sna1993*.

Table 1.1. **Gross domestic product, current PPPs**

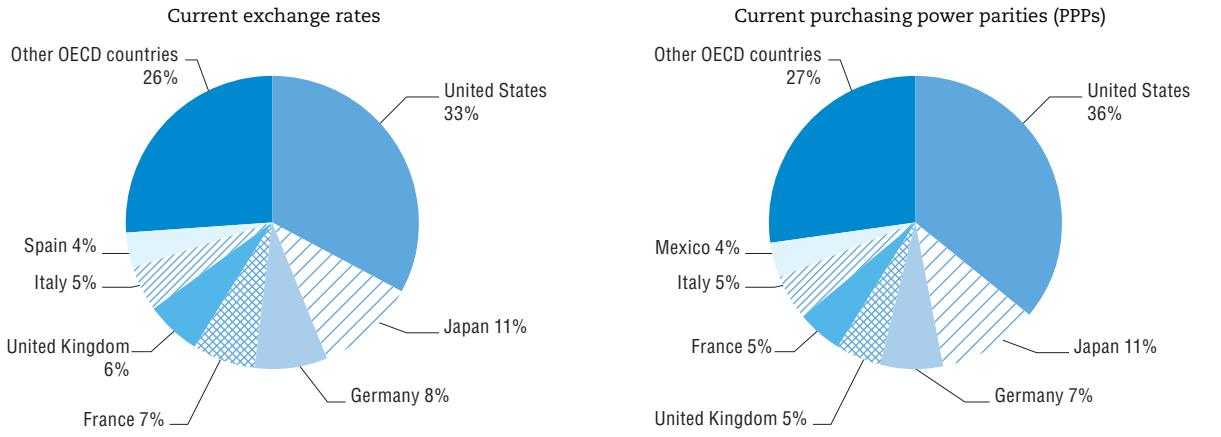
Billion US dollars

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Australia	392	412	437	464	497	525	552	585	622	656	697	743	795	830 e
Austria	187	194	199	208	216	230	232	244	252	266	275	291	309	316
Belgium	228	232	243	248	259	282	292	309	313	323	336	354	376	374
Canada	667	691	732	771	825	873	910	938	990	1 050	1 131	1 204	1 270	1 303 e
Czech Republic	133	141	143	144	147	154	165	172	184	197	208	226	248	257
Denmark	120	127	134	139	143	154	158	165	164	175	180	190	196	200
Finland	96	99	108	117	122	133	138	143	144	156	161	172	184	188
France	1 204	1 243	1 302	1 369	1 425	1 533	1 630	1 711	1 701	1 768	1 869	1 961	2 081	2 115
Germany	1 840	1 892	1 936	1 990	2 064	2 130	2 212	2 275	2 358	2 468	2 587	2 709	2 835	2 928
Greece	156	163	173	179	185	201	218	237	250	267	274	294	316	324
Hungary	93	97	104	111	115	124	138	150	156	165	171	181	189	193
Iceland	6	7	7	8	8	8	9	9	9	10	10	11	11	12
Ireland	65	71	80	89	97	109	118	130	138	148	160	177	195	186
Italy	1 202	1 242	1 285	1 351	1 377	1 456	1 546	1 532	1 564	1 596	1 649	1 737	1 813	1 849
Japan	2 831	2 964	3 061	3 032	3 071	3 246	3 331	3 417	3 510	3 710	3 873	4 081	4 293	4 356 e
Korea	603 e	657 e	699 e	658 e	731 e	806	860	936	964	1 041	1 097	1 195	1 300	1 358
Luxembourg	16	17	17	18	21	23	24	26	27	30	32	36	40	41
Mexico	688 e	737 e	800 e	849 e	894 e	986 e	1 009 e	1 048 e	1 108	1 186	1 294	1 402	1 485	1 537 e
Netherlands	334	352	376	400	426	468	494	516	515	541	573	607	644	681
New Zealand	64	67	70	71	76	80	85	89	93	99	103	109	115	115 e
Norway	103	114	123	122	133	162	167	168	175	194	219	243	252	278
Poland	287	311	340	363	383	404	419	442	458	497	526	566	614	674
Portugal	131	137	146	154	164	175	183	191	196	201	218	229	242	246
Slovak Republic	45	49	52	56	56	59	65	70	73	79	87	97	108	119
Spain	631	661	701	751	792	857	920	994	1 040	1 109	1 188	1 304	1 420	1 447
Sweden	193	200	207	215	229	246	249	259	269	289	292	313	335	344
Switzerland	189	194	203	210	215	228	234	245	246	258	266	290	314	332
Turkey	427 e	467 e	511 e	536	518	589	561	572	588	689	781	881	945	998
United Kingdom	1 146	1 220	1 308	1 363	1 423	1 533	1 631	1 714	1 778	1 900	1 969	2 068	2 167	2 200
United States	7 359	7 784	8 279	8 741	9 301	9 899	10 234	10 590	11 089	11 812	12 580	13 336	14 011	14 369
Euro area	6 178	6 394	6 664	6 978	7 256	7 711	8 131	8 442	8 636	9 021	9 483	10 044	10 644	10 910
OECD-Total	21 438 e	22 541 e	23 775 e	24 724 e	25 914 e	27 671 e	28 784 e	29 879 e	30 977	32 878	34 805	37 004	39 103	40 170 e

StatLink http://dx.doi.org/10.1787/740520054135

Figure 1.1. **Gross domestic product: current exchange rates and current PPPs**

The seven largest economies in the OECD. Percentage of OECD total, 2008



StatLink http://dx.doi.org/10.1787/738543187137

2. GDP growth

Changes in the size of economies are usually measured by changes in the volume (often referred to as real) of GDP. Real reflects the fact that changes in GDP due to inflation are removed. This provides a measure of changes in the volume of production of an economy.

Definition

Converting nominal values of GDP to real values requires a set of detailed price indices, implicit or directly collected. When applied to the nominal value of transactions, the corresponding volume changes can be captured. The detailed volume changes for goods and services – typically several hundred – are then aggregated to yield an overall change in the volume of GDP. In the past, most countries used fixed weights for this aggregation and the base year to which weights related was only modified every five to ten years. It is important to recognise that growth rates are not invariant to the choice of this reference period and measures of growth could turn out to be biased for reporting years that were remote from the base year.

Since the 1993 System of National Accounts it has therefore been recommended that weights should be representative of the periods for which growth rates are calculated. This means that new weights should be introduced every year, giving rise to chain-linked (volume) indices (see Comparability, below).

Comparability

As described in Section 1, comparability of nominal values of GDP across countries is good. There is generally some variability in how countries calculate their volume estimates of GDP, particularly in respect of government consumption, but this doesn't necessarily mean that growth rates are less comparable.

With the exception of Mexico, all OECD countries derive their annual estimates of real GDP using annually chain-linked volume indices (that is the fixed prices/weights are updated every year). Mexico, like many non-OECD countries, revise their fixed weights less frequently – the last revision occurring after ten years. Such practices however tend to lead to biased growth rates, usually upward.

Source

- OECD (2009), *National Accounts of OECD Countries 2009, Volume I, Main Aggregates*, OECD Publishing, *http://dx.doi.org/10.1787/na_vol_1-2009-en-fr*.

Online database

- OECD (2009), "Aggregate national accounts: gross domestic product", OECD *National Accounts Statistics* (database), *http://dx.doi.org/10.1787/data-00001-en*.

Further reading

- Ahmad N., F. Lequiller, P. Marianna, D. Pilat, P. Schreyer, A. Wölfl (2003), *Comparing Labour Productivity Growth in the OECD Area: The Role of Measurement*, OECD Publishing, *http://dx.doi.org/10.1787/838342850485*.
- Eurostat (2001), *Handbook on Price and Volume Measures in National Accounts*, Eurostat, Luxembourg.
- Lequiller, F. and D. Blades (2007), *Understanding National Accounts*, OECD Publishing, *http://dx.doi.org/10.1787/9789264027657-en*.
- OECD (2000), *System of National Accounts, 1993 – Glossary*, OECD Publishing, *http://dx.doi.org/10.1787/9789264180871-en*.
- UN, OECD, IMF, Eurostat (eds.) (1993), *System of National Accounts 1993*, United Nations, Geneva, *http://unstats.un.org/unsd/sna1993*.

Table 2.1. **Gross domestic product, volume**

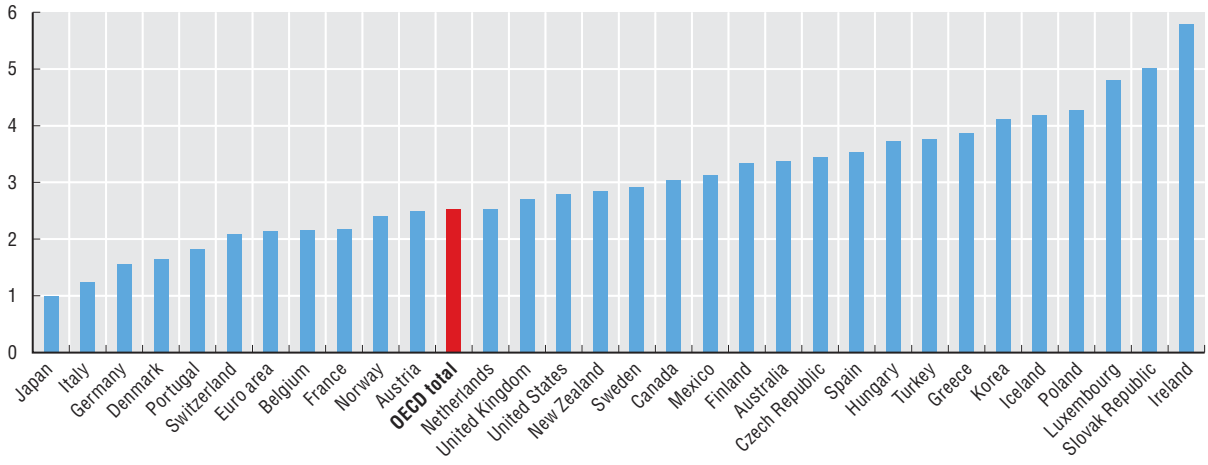
Annual growth rates in percentage

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Australia	4.1	3.9	4.5	5.2	4.0	1.9	3.8	3.2	4.0	2.8	3.0	3.3	3.7	2.3 e
Austria	2.5	2.2	2.1	3.6	3.3	3.7	0.5	1.6	0.8	2.5	2.5	3.5	3.5	2.0
Belgium	2.4	1.2	3.5	1.7	3.4	3.7	0.8	1.5	1.0	3.0	1.8	3.0	2.8	1.1
Canada	2.8	1.6	4.2	4.1	5.5	5.2	1.8	2.9	1.9	3.1	2.9	3.1	2.7	0.4 e
Czech Republic	5.9	4.0	-0.7	-0.8	1.3	3.6	2.5	1.9	3.6	4.5	6.3	6.8	6.1	2.5
Denmark	3.1	2.8	3.2	2.2	2.6	3.5	0.7	0.5	0.4	2.3	2.4	3.3	1.6	-1.2
Finland	3.9	3.7	6.2	5.2	3.9	5.1	2.7	1.6	1.8	3.7	2.8	4.9	4.2	1.0
France	2.1	1.1	2.2	3.5	3.3	3.9	1.9	1.0	1.1	2.5	1.9	2.2	2.3	0.4
Germany	1.9	1.0	1.8	2.0	2.0	3.2	1.2	0.0	-0.2	1.2	0.8	3.2	2.5	1.3
Greece	2.1 e	2.4 e	3.6 e	3.4 e	3.4 e	4.5 e	4.2	3.4	5.9	4.6	2.2	4.5	4.5	2.0
Hungary	1.5	1.0	4.3	5.2	4.2	4.9	4.1	4.4	4.3	4.9	3.5	4.0	1.0	0.6
Iceland	0.1	4.8	4.9	6.3	4.1	4.3	3.9	0.1	2.4	7.7	7.4	4.5	5.5	0.3
Ireland	9.6	8.1	11.5	8.4	10.7	9.4	5.7	6.5	4.4	4.6	6.2	5.4	6.0	-3.0
Italy	2.8	1.1	1.9	1.4	1.5	3.7	1.8	0.5	-0.0	1.5	0.7	2.0	1.6	-1.0
Japan	2.0	2.7	1.6	-2.0	-0.1	2.9	0.2	0.3	1.4	2.7	1.9	2.0	2.4	-0.7 e
Korea	9.2 e	7.0 e	4.7 e	-6.9 e	9.5 e	8.5 e	4.0	7.2	2.8	4.6	4.0	5.2	5.1	2.2
Luxembourg	1.4	1.5	5.9	6.5	8.4	8.4	2.5	4.1	1.5	4.4	5.4	5.6	6.5	0.0
Mexico	-6.2 e	5.2 e	6.8 e	5.0 e	3.8 e	6.6 e	-0.0 e	0.8 e	1.4 e	4.0	3.3	5.0	3.4	1.3 e
Netherlands	3.1	3.4	4.3	3.9	4.7	3.9	1.9	0.1	0.3	2.2	2.0	3.4	3.6	2.0
New Zealand	4.2	3.5	1.7	0.5	5.3	2.4	3.6	4.9	4.3	3.8	3.0	1.8	3.1	-1.1 e
Norway	4.2	5.1	5.4	2.7	2.0	3.3	2.0	1.5	1.0	3.9	2.7	2.3	3.1	2.1
Poland	7.0	6.2	7.1	5.0	4.5	4.3	1.2	1.4	3.9	5.3	3.6	6.2	6.8	5.0
Portugal	4.3	3.6	4.2	4.9	3.8	3.9	2.0	0.8	-0.8	1.5	0.9	1.4	1.9	-0.0
Slovak Republic	5.8	6.9	4.4	4.4	0.0	1.4	3.4	4.8	4.7	5.2	6.5	8.5	10.4	6.4
Spain	2.8	2.4	3.9	4.5	4.7	5.0	3.6	2.7	3.1	3.3	3.6	4.0	3.6	0.9
Sweden	4.0	1.5	2.5	3.8	4.6	4.4	1.1	2.4	1.9	4.1	3.3	4.2	2.6	-0.2
Switzerland	0.4	0.6	2.1	2.6	1.3	3.6	1.2	0.4	-0.2	2.5	2.6	3.6	3.6	1.8
Turkey	7.2 e	7.0 e	7.5 e	3.1	-3.4	6.8	-5.7	6.2	5.3	9.4	8.4	6.9	4.7	0.9
United Kingdom	3.0	2.9	3.3	3.6	3.5	3.9	2.5	2.1	2.8	2.8	2.1	2.8	3.0	0.7
United States	2.5	3.8	4.5	4.4	4.9	4.2	1.1	1.8	2.5	3.6	3.1	2.7	2.1	0.4
Euro area	..	1.6	2.6	2.8	2.9	3.9	1.9	0.9	0.8	2.2	1.7	3.0	2.8	0.7
OECD-Total	2.5 e	3.1 e	3.6 e	2.7 e	3.5 e	4.2 e	1.3 e	1.7 e	2.0 e	3.2	2.7	3.1	2.8	0.6 e

StatLink http://dx.doi.org/10.1787/740522845022

Figure 2.1. **Gross domestic product, volume**

Average annual growth rates between 1998 and 2008



StatLink http://dx.doi.org/10.1787/738560831304

3. GDP per capita

Gross Domestic Product (GDP) per capita is a core indicator of economic performance and commonly used as a broad measure of average living standards or economic well-being; despite some recognised shortcomings.

For example average GDP per capita gives no indication of how GDP is distributed between citizens. Average GDP per capita may rise for example but more people may be worse off if income inequalities also increase.

Equally, in some countries (see Comparability on the right), there may be a significant number of non-resident border or seasonal workers or indeed inflows and outflows of property income and both phenomena imply that the value of production differs from the income of residents, thereby over or understating their living standards.

A full discussion of these issues can be found in the Stiglitz-Sen-Fitoussi report (see references).

Definition

The definition for GDP is described in section 1 and population estimates are described in the Reader Notes section.

A focus on per capita GDP is also useful in decomposing drivers of overall GDP growth. For example real GDP can grow without there being any improvement in real GDP per capita. Decomposing per capita growth into two parts, labour productivity growth (measured as GDP per hour worked) and labour utilisation growth (measured as hours worked per capita) is helpful in this context.

Comparability

Generally, particularly because all countries follow the *1993 System of National Accounts*, the comparability of population and GDP estimates across countries is good (see Section 1). However, some care is needed in interpretation, for example Luxembourg and, to a lesser extent, Switzerland have a relatively large number of frontier workers. Such workers contribute to GDP but are excluded from the population figures, which is one of the reasons why cross-country comparisons of income per capita based on gross or net national income (GDI and NNI) are often preferred, see Chapter 2 on Income. (See also "Reader's Guide", relating to PPP based comparisons.)

Source

- OECD (2009), *National Accounts of OECD Countries 2009, Volume I, Main Aggregates*, OECD Publishing, *http://dx.doi.org/10.1787/na_vol_1-2009-en-fr*.

Online database

- OECD (2009), "Aggregate national accounts: gross domestic product", OECD *National Accounts Statistics* (database), *http://dx.doi.org/10.1787/data-00001-en*.

Further reading

- Report of the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz-Sen-Fitoussi Report), *www.stiglitz-sen-fitoussi.fr*.
- OECD (2002), *Measuring the Non-Observed Economy: A Handbook*, International Labour Office/International Monetary Fund/International Statistical Committee of the Commonwealth of Independent States, OECD Publishing, *http://dx.doi.org/10.1787/9789264175358-en*.
- Lequiller, F. and D. Blades (2007), *Understanding National Accounts*, OECD Publishing, *http://dx.doi.org/10.1787/9789264027657-en*.
- OECD (2000), *System of National Accounts, 1993 – Glossary*, OECD Publishing, *http://dx.doi.org/10.1787/9789264180871-en*.
- UN, OECD, IMF, Eurostat (eds.) (1993), *System of National Accounts 1993*, United Nations, Geneva, *http://unstats.un.org/unsd/sna1993*.

Table 3.1. **Gross domestic product per capita, OECD=100**

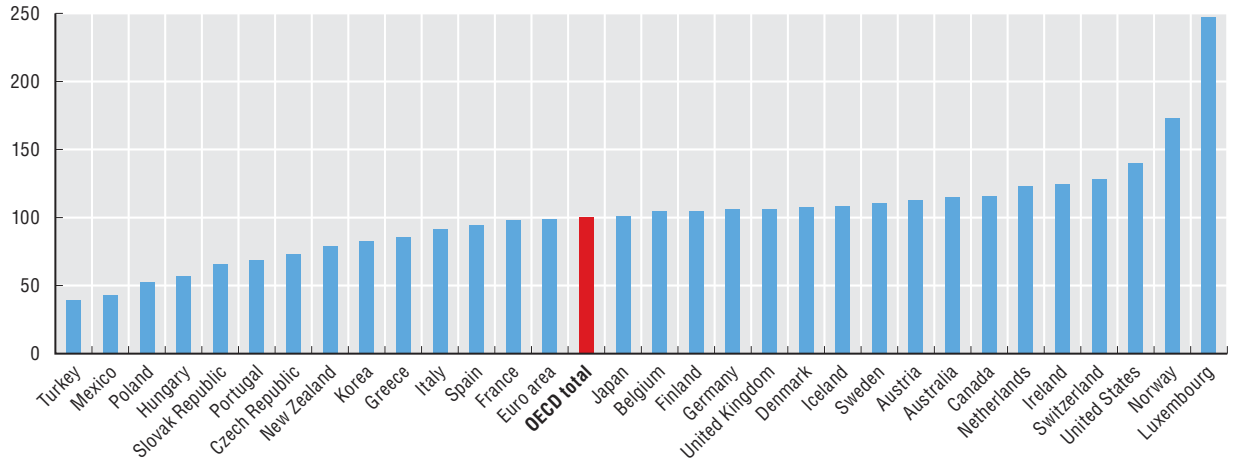
Based on current PPPs

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Australia	110	109	109	111	113	111	112	114	116	115	114	113	114	115 e
Austria	120	119	116	117	117	117	114	116	116	115	112	112	113	113
Belgium	114	111	111	109	109	112	112	115	112	110	108	107	107	105
Canada	116	114	114	115	117	116	116	115	116	116	118	117	117	116 e
Czech Republic	65	67	64	63	62	61	64	65	67	68	68	70	73	73
Denmark	117	117	117	118	116	118	116	118	113	114	112	111	109	108
Finland	96	94	98	102	102	105	105	106	103	106	103	104	105	105
France	103	102	101	103	102	103	105	107	102	100	100	98	99	98
Germany	115	113	110	109	109	106	106	106	107	106	105	105	104	106
Greece	75	74	75	74	74	75	79	83	85	85	83	84	86	86
Hungary	46	46	47	49	49	49	54	57	57	58	57	57	57	57
Iceland	118	118	121	125	124	118	120	119	115	119	118	112	110	108
Ireland	91	95	101	108	112	117	121	127	129	129	130	133	136	125
Italy	108	106	105	107	105	104	107	103	101	97	95	94	93	92
Japan	115	115	113	108	105	104	103	103	102	103	102	102	102	101 e
Korea	68 e	70 e	71 e	64 e	68 e	70	72	75	75	77	77	79	81	83
Luxembourg	198	196	189	194	211	218	213	221	226	230	230	243	250	248
Mexico	38 e	39 e	40 e	40 e	40 e	41 e	40 e	40 e	41	41	42	43	43	43 e
Netherlands	110	110	112	115	116	120	122	123	118	117	118	118	119	123
New Zealand	87	86	85	84	86	84	85	85	85	85	83	82	82	79 e
Norway	120	127	130	123	129	147	147	142	143	149	159	166	162	173
Poland	38	40	41	43	43	43	43	44	45	46	46	47	49	53
Portugal	67	66	67	68	70	70	70	71	70	68	69	69	69	69
Slovak Republic	42	44	45	46	45	45	48	50	51	52	54	57	61	66
Spain	81	82	82	85	86	87	89	92	92	92	92	94	96	94
Sweden	111	110	109	109	111	113	111	111	112	113	109	110	111	111
Switzerland	136	133	132	133	130	129	127	128	124	122	119	122	125	128
Turkey	35 e	36 e	38 e	38	35	36	32	32	31	34	36	38	39	40
United Kingdom	100	102	104	105	105	106	109	111	111	112	110	109	108	107
United States	140	141	141	142	144	143	142	141	142	142	143	142	141	140
Euro area	102	101	100	101	100	100	102	102	101	100	99	99	99	99
OECD-Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100

StatLink http://dx.doi.org/10.1787/740577530321

Figure 3.1. **Gross domestic product per capita, OECD=100**

Based on current PPPs, 2008



StatLink http://dx.doi.org/10.1787/738633261342

Hutchins Center Working Paper #43

August 2018

GDP as a Measure of Economic Well-being

Karen Dynan

Harvard University

Peterson Institute for International Economics

Louise Sheiner

Hutchins Center on Fiscal and Monetary Policy, The Brookings Institution

The authors thank Katharine Abraham, Ana Aizcorbe, Martin Baily, Barry Bosworth, David Byrne, Richard Cooper, Carol Corrado, Diane Coyle, Abe Dunn, Marty Feldstein, Martin Fleming, Ted Gayer, Greg Ip, Billy Jack, Ben Jones, Chad Jones, Dale Jorgenson, Greg Mankiw, Dylan Rassier, Marshall Reinsdorf, Matthew Shapiro, Dan Sichel, Jim Stock, Hal Varian, David Wessel, Cliff Winston, and participants at the Hutchins Center authors' conference for helpful comments and discussion. They are grateful to Sage Belz, Michael Ng, and Finn Schuele for excellent research assistance.

The authors did not receive financial support from any firm or person with a financial or political interest in this article. Neither is currently an officer, director, or board member of any organization with an interest in this article.

ABSTRACT

The sense that recent technological advances have yielded considerable benefits for everyday life, as well as disappointment over measured productivity and output growth in recent years, have spurred widespread concerns about whether our statistical systems are capturing these improvements (see, for example, Feldstein, 2017). While concerns about measurement are not at all new to the statistical community, more people are now entering the discussion and more economists are looking to do research that can help support the statistical agencies.

While this new attention is welcome, economists and others who engage in this conversation do not always start on the same page. Conversations are impeded by a lack of understanding of how the statistics are defined and how they are limited, both in terms of the concept and in terms of how they are calculated given the concept. We explore the basic economics surrounding the measurement of GDP, focusing, in particular, on the question of whether GDP should be viewed as a measure of aggregate economic well-being.

Our exploration suggests that while GDP, as currently defined, is not a comprehensive measure of welfare or even economic well-being, the GDP concept—along with the pieces of GDP available through the national accounts—is useful in and of itself and should provide a great deal of information that is closely related to welfare.

Our finding that changes in real GDP do a reasonable job in capturing changes in economic well-being has one important exception. We argue that the exclusion of non-market activities that bear on economic well-being merits more attention, particularly given the potential for changes in the importance of such activities over time to change the degree to which changes in GDP capture changes in well-being.

Moreover, there are several important areas where measurement falls short of the conceptual ideal. First, the national accounts may mismeasure the nominal GDP arising from the digital economy and the operation of multinationals corporations. Second, the deflators used to separate GDP into nominal GDP and real GDP may produce a biased measure of inflation. Our analysis suggests that, for goods and services that do not change in quality over time, current deflator methods work reasonably well. But, for new goods and services or goods in services that are changing in quality, current methods may not capture consumer surplus well. We believe that efforts to improve price measurement in order to measure consumer welfare should be pursued, as it is clear that such a measure would be very useful for understanding the current state of the economy and for policymaking.

1. Introduction

Published measures of growth in productivity and real gross domestic product (GDP) since the early 2000s have been distressingly slow despite very visible improvements in high-tech equipment (the smart phone), in internet-based services (Facebook and Google), in business models (Uber and Lyft), and in the quality of health care. This has revived interest in how well official measures capture improvements in standards of living (see, for example, Feldstein, 2017). Part of the literature that considers the explanations for recently weak productivity growth explicitly explores measurement issues. Much of this work concludes that measurement is at best a small part of the explanation for slower trend productivity growth (Byrne, Fernald, and Reinsdorf, 2016, Syverson, 2016, and Fernald, Hall, Stock, and Watson, 2017) but a few argue that measurement has played a larger role (Varian, 2016, and Hatzius, 2017.).

Concerns about measurement are, of course, not at all new to experts on economic statistics, including those in government and in academia. For decades, data-producing agencies have been working to improve measurement and to make sure that standards are consistent across countries. These efforts have yielded major methodological advances. Moulton (2018), for example, catalogs key improvements to the U.S. national income and product accounts since the late 1990s.

More people are now entering the discussion and more economists are looking to do research that can help support the statistical agencies. The starting point for these efforts should be a basic understanding of how the statistics are defined and how they are limited, both in terms of the concept and in terms of how they are calculated given the concept. While much of this information can be found in writings by experts on economic statistics, this literature is large in volume and often hard to understand by non-experts, even other economists. The goal of this paper is to supply some basic answers, with a focus on real GDP, the most closely-watched aggregate economic indicator and one which is so often used as a measure of the standard of living. Accurately measuring real GDP is essential to accurately measuring productivity, which is essentially output (real GDP) divided by inputs.

We begin our paper with a discussion of how the established GDP concept relates to welfare, or more specifically to a somewhat narrower concept that we term "aggregate economic well-being" which excludes factors that are very far outside the scope of GDP, such as the quality of the environment. We explain the advantages to GDP as defined and consider the importance of the differences between GDP and economic well-being. We also discuss some alternative and complementary approaches that can help bridge the gap between GDP and economic well-being.

We next turn to how well GDP as conceptualized by data producers is captured in practice. Notwithstanding the important advances in measurement over time, increases in the share of GDP represented by difficult-to-measure sectors (such as health care and the digital economy) may mean that the published GDP figures do not track the conceptual ideal as well as they have done in the past. Moreover, the limited resources of data-producing agencies (which are at risk of future cuts in the current political environment) may constrain these agencies' ability to cope with such challenges.

We consider first whether the nominal (i.e. current dollar) GDP figure adequately captures the size of our economy measured in dollars. We conclude that mostly it does, but there are two important measurement challenges. One challenge is the treatment of so-called "free goods," particularly given the dramatic rise in services provided by the internet for which consumers do not explicitly pay. Another is the understatement of the domestic economic activity of multinational enterprises that arises from tax incentives.

Converting current dollar figures to real GDP (that is, GDP expressed in the dollars from a particular base year) presents even thornier issues. Hence, the second (and much larger) part of our measurement discussion concerns challenges related to the deflators used to calculate real GDP. A central issue here is how to separate changes in prices that reflect quality improvements from those that represent true inflation. Another issue is estimating the value of dollars spent on newly introduced goods and services. The paper offers a discussion of the ideal way to treat these measurement issues and then discusses what the statistical agencies do in practice.

We draw several conclusions. First, GDP, as currently defined, should retain its stature as a major economic statistic. While it is not a comprehensive measure of welfare or even economic well-being, the GDP concept—along with the pieces of GDP available through the national accounts—is useful in and of itself and should provide a great deal of information that is closely related to welfare. Second, there is scope for materially improving specific parts of the GDP calculation to be more closely aligned with the conceptual ideal. Doing so should be a goal for the statistical community and for the broader community of economists. Third, given the limitations of GDP as a measure of welfare (and the potential for those limitations to increase over time), we should continue to develop complementary measures or sets of measures (sometimes termed "dashboards") that more completely capture well-being.

2. The GDP concept

The Bureau of Economic Analysis (BEA) gives a clear definition for GDP:

> Gross domestic product (GDP) is the value of the goods and services produced by the nation's economy less the value of the goods and services used up in production. GDP is also equal to the sum of personal consumption expenditures, gross private domestic investment, net exports of goods and services, and government consumption expenditures and gross investment.[1]

The U.S. Commerce Department began to publish regular estimates of GDP, defined essentially as above, in the early 1940s (Carson, 1975). The Commerce Department framework built on methods that Simon Kuznets used to estimate national income for 1929-32 under the auspices of the National Bureau of Economic Research (NBER). Kuznets's work was preceded by two volumes published by the NBER in the early 1920s that provided estimates of national income over the preceding decade. Others were also engaged in efforts to measure economic activity around this time. For example, the National Industrial Conference Board (which later became just the Conference Board) began publishing a regular estimate of national income in the 1920s. Colin Clark, a British economist and statistician, was doing work similar to Kuznets's, measuring the aggregate economy of the United Kingdom (Coyle, 2014).

GDP is the featured measure of output in the National Income and Product Accounts (NIPAs), a vast set of economic data that captures economic activity in the United States.[2] Some explanation of the NIPAs

. . .

1. See www.bea.gov/newsreleases/national/gdp/gdpnewsrelease.htm.

2. The NIPAs are, in turn, just one part of a broader set of U.S. national accounts that also include the Labor Department's productivity statistics and the Federal Reserve's system of financial accounts. Dale Jorgenson, who has made enormous contributions over his career to a wide array of national accounting practices both in this country and in other countries, describes the national accounts "as a kind of central nervous system for federal statistics" (Jorgenson, 2010).

is needed to understand the text that follows. As described in Bureau of Economic Analysis (2015), there are different approaches to measuring GDP. The "expenditure approach," in which GDP is measured as the sum of consumption, investment, government spending, and net exports, is the most familiar to many people. The expenditure side of the national accounts includes estimates of these pieces as well as their components. GDP can also be measured through the "income approach," which adds up all of the income earned through production, and the income side of the national accounts includes the various types of income that goes into GDP. The income-side measure of GDP is known as Gross Domestic Income (GDI). In theory, GDP measured through the expenditure approach should equal GDI; in practice, of course, GDP does not equal GDI because of measurement error, and BEA publishes a "statistical discrepancy" that captures the gap between the two series.[3]

2.1 The differences between GDP and welfare

As a long literature has emphasized, GDP as conventionally defined differs in many ways from welfare.[4] The economists who developed the modern concept of GDP were well aware of this distinction. For example, in a 1934 report to Congress, Kuznets stated that "the welfare of a nation … can scarcely be inferred from a measure of national income" (Bureau of Foreign and Domestic Commerce and Kuznets, 1934).

Some of the differences between GDP and welfare are outside the scope of this paper. For example, GDP does not include important societal features such as discrimination and crime. In addition, as an economy-wide concept, GDP does not provide information about the distribution of income, which bears importantly on the welfare of individuals within an economy.[5] Nor does GDP capture features of the environment such as climate change and the availability of natural resources.

Much of the discussion of GDP and welfare in this paper will focus on a narrower distinction—the difference between GDP and what we call *aggregate economic well-being*, defined as the consumer welfare derived from market-based activities and selected non-market-based activities such as services provided by governments, certain nonprofit institutions, and homeownership.

The key differences between GDP and aggregate economic well-being are:

1. GDP excludes most home production, and other "non-market" activities such as leisure, even though most such activity effectively increases the true consumption of households and thus enhances welfare (more discussion of this point below).

. . .

3. There is also a "value-added approach" to measuring GDP which involves taking the difference between total sales and the value of intermediate inputs or summing up the "value added" at each stage of the production process. This approach is central to analyzing the economy at the industry level, but it does not figure prominently in the discussion that follows.

4. Coyle (2014) summarizes the historical debate over this issue. Jorgenson (forthcoming) provides an extensive discussion of the relationship between measured GDP and welfare. See also Constanza, Hart, Posner, and Talberth (2009), Wesselink, Bakkes, Best, Hinterberger, and ten Brink (2007), Kassenboehmer and Schmidt (2011), and Boyd (2007) for more on this topic and alternative measures of economic progress.

5. Piketty, Saez, and Zucman (2016) create distributional national accounts for the United States that shed light on how standards of living have evolved at different points in the income distribution.

2. GDP represents domestic production, but some of that production is "owned" by foreigners; furthermore, Americans own some foreign production. The welfare of Americans is more closely correlated with the income they receive from the production they own regardless of where it occurs than simply the production done in this country.
3. GDP includes production that makes up for the depreciation of physical assets. Such production is done to maintain the current capital stock rather than increasing the services consumed by households.
4. GDP includes investment—by businesses, by government, and by households (through housing and consumer durables). While this investment may provide future services to households, it does not represent services enjoyed immediately by households. We note, though, that there is some disagreement as to whether investment should be counted in a measure of well-being. For example, Corrado, Fox, Goodridge, Haskel, Jona-Lasinio, Sichel, and Westlake (2017) point out that one might view well-being as depending not only on current consumption but also on future consumption, which, in turn, is influenced by what firms are investing today.[6]

Despite these well-known differences, GDP is often used—by politicians, reporters, the general public, and even economists—as a proxy for welfare or at least economic well-being. This begs the question of why the economists and statisticians who developed the modern concept of GDP chose the definition they did. Our reading of the literature suggests several factors contributed to their thinking.

One factor is that the modern market-production-based concept of GDP is better aligned with the Keynesian concept of "demand." Although new homes might yield services for consumers that raise welfare by modest increments over a long period of time, the investment associated with the building of those homes or cars use a lot of the economy's productive resources over a short period of time. Policymakers who are trying to use fiscal or monetary tools to stabilize the economy in the face of business-cycle fluctuations need to know how the use of productive resources compares to the economy's supply of such resources.

A second factor might be war-related. In particular, some have argued that it is no coincidence that the modern interest in measuring the aggregate economy arose during World War I and that needs related to the war contributed to the production focus of the modern GDP concept. On the practical side, understanding what the economy could produce presumably greatly facilitated planning for war efforts (Landefeld, 2000). Coyle (2014) also notes the political advantages of a production focus—production-based measures do not show the economy shrinking during wartime even if resources available for private consumption plummet.

A third factor is feasibility. In particular, the literature suggests that home production and many other activities that are not captured by market transactions were left out because they were viewed as difficult to measure. Indeed, there was a vigorous debate about whether it made sense, for example, that the services provided by professional and paid housekeepers were included in the GDP concept but that any personal housekeeping efforts were not included. It was accepted, though, that the latter was more difficult to measure, and, as Carson (1975) describes, the economists involved in the NBER effort

. . .

6. This view echoes longer discussions in in Weitzman (1976) and Weitzman (2003), which argue that net investment belongs in a welfare measure because it captures future consumption opportunities.

"retreat[ed] … to ground more securely buttressed by reliable data" (p. 158). Similarly, trying to put a value on leisure can be quite difficult, particularly given that individuals sometimes freely choose to take leisure but at other times cannot work as many hours as they would like at the prevailing wage, and sometimes may not be able to find work at all.

Regarding this third factor, several additional points are worth noting. First, the precise boundary between market production and non-market production has never been well defined. For example, the current methodological framework for GDP is not conceptually consistent. Services provided by consumer durable goods (like a car) are not included in GDP because they are viewed as non-market production, but services provided by owner-occupied homes are included through imputed rent (with the rent included on both the product and income sides of the account so that the two sides will be aligned). Second, while it may have been extremely difficult to measure non-market production at the time the accounts were originally constructed, new technologies and data sources may offer opportunities to capture components of economic well-being that previously were difficult or impossible to measure. Third, the cost of excluding traditionally defined non-market services may be greater than in the past, given that many of the services that people enjoy from the internet are not "paid for" through traditional market transactions. We return to this issue in our section on "free goods" below.

2.2 Do these conceptual differences matter?

Any assessment of the GDP concept as a measure of aggregate economic well-being needs to recognize that many of the shortcomings are addressed by looking at measures that are already available as part of the standard national income accounts. For example, investment (including that making up for depreciation of assets) can be netted out of GDP. To address the issue that some of the income associated with domestic production belongs to foreigners (and, likewise, that Americans receive some income from production that is done in other countries), gross national product (GNP), which captures the production of assets owned by Americans regardless of where in the world it occurs.[7]

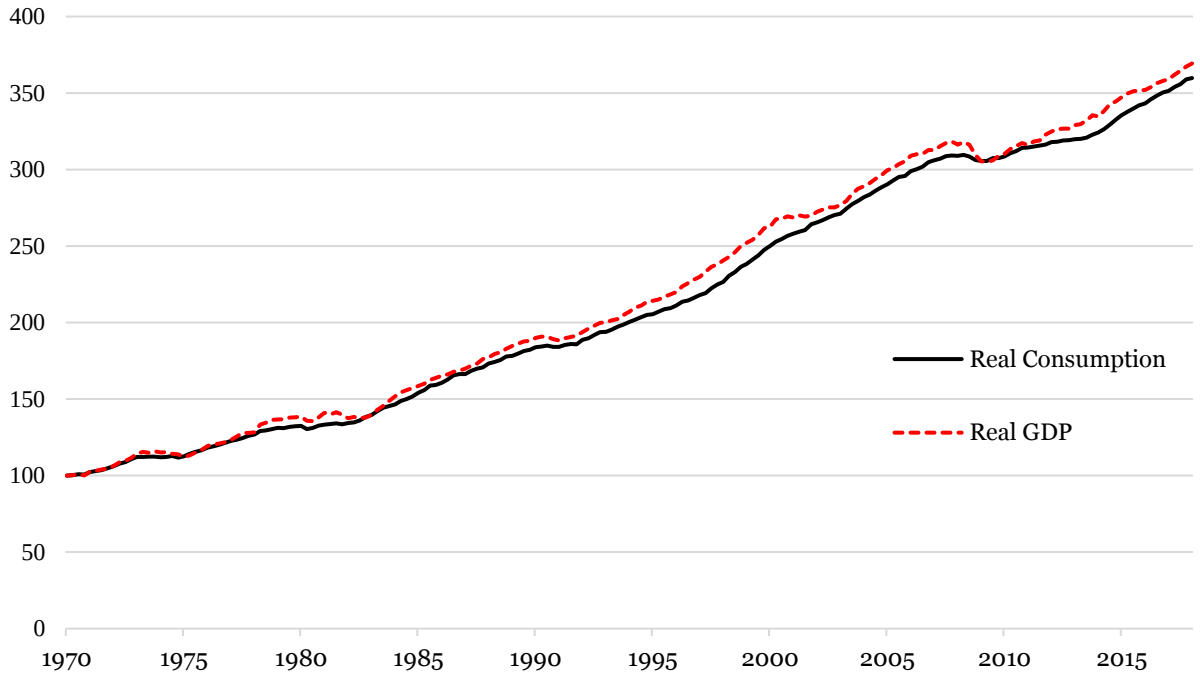
Indeed, one might expect consumption—derived from standard national accounts series and broadly defined to include both the spending done directly by households and the services provided to households by government spending—to align fairly well with economic well-being. (Note that consumption defined in this way overcomes both shortcomings discussed in the previous paragraph: it excludes investment and is funded by income earned by Americans rather than income related to domestic consumption.) The solid black line in Figure 1 shows cumulative growth in real broadly defined consumption (the sum of personal consumption expenditures plus government consumption expenditures) since 1970.[8] The series has risen by roughly three-and-one-half fold over the 48-year period shown. The figure also shows that cumulative growth in real GDP (depicted by the red dashed line) has been about the same over this period—suggesting that GDP, even with its conceptual differences, is not a bad proxy for broadly defined consumption.

. . .

7. BEA treated GNP as the primary measure of U.S. economic activity for many decades but switched its focus in 1991 to conform with practices of statistical agencies in other countries.

8. For this exercise, we ignore the fact that consumption expenditure includes some durable goods, which yield consumption services (i.e. provide utility to the household) over time. Looking at only nondurables and services would not materially change our conclusion.

Figure 1: Cumulative growth in consumption and GDP since 1970



Source: Bureau of Economic analysis (Haver Analytics)

The one important conceptual shortcoming of GDP as a measure of economic well-being that cannot be resolved through series already in the standard national income accounts is GDP's exclusion of (most) non-market activities that create welfare for households. Trends in the importance of non-market activities could lead to a widening gap between household welfare and GDP such that *changes* in measured GDP may not proxy for *changes* in well-being over the longer run. For example, the surge of women into the labor force in the 1970s, 1980s, and 1990s would have boosted GDP even if newly employed women were previously producing the same amount outside of the marketplace—a case in which the increase in GDP would have overstated the increase in welfare. However, some more recent trends would go in the opposite direction. For example, the internet has made it easier for people to arrange for travel directly instead of going through a travel agent—these personal efforts to book travel are not counted in GDP but the services of a travel agent would be counted, leading GDP growth to understate the increase in welfare. (In this case, at least the travel purchased shows up in GDP—in our section on "free goods" we discuss the degree to which services consumed more broadly via the internet are showing up in GDP.)[9]

BEA does periodically publish satellite accounts with values for some types of non-market activities. Recent updates to these accounts (Bridgman, Dugan, Lal, Osborne, and Villones, 2012, and Bridgeman, 2016) include estimates for home production (such as cooking, cleaning, and shopping) and the services provided to households from durable goods (such as cars and appliances). Building upon this work—

. . .

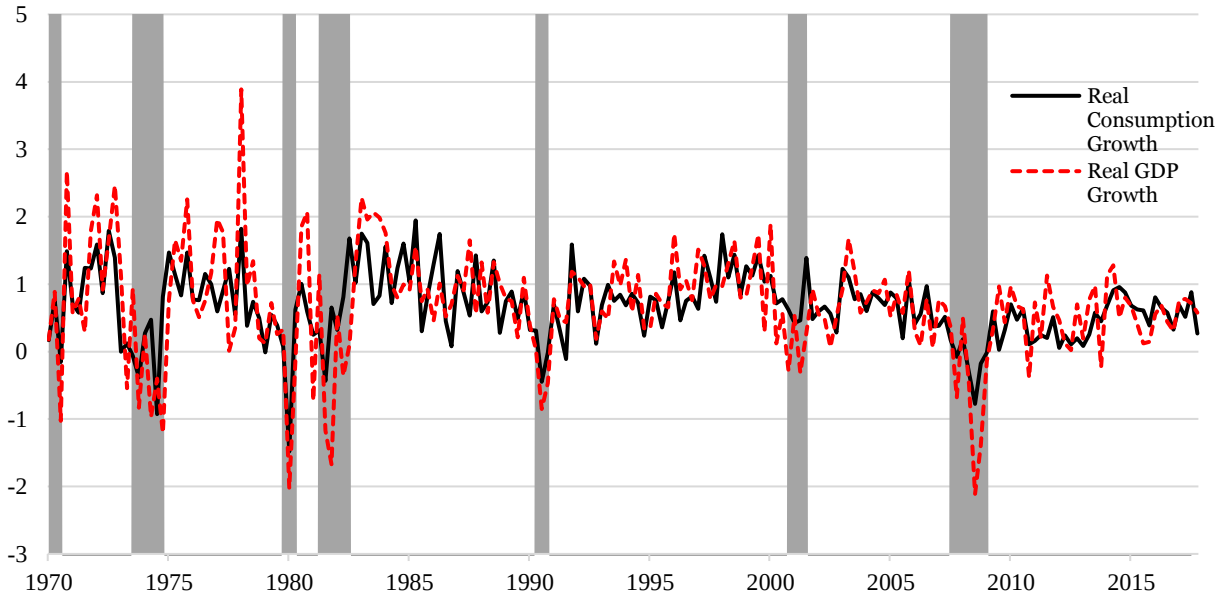
9. Importantly, while these types of trends might distort measured GDP growth, they will *not* generally distort measured growth in productivity (output per hour) because hours get undercounted or overcounted in the same way as GDP.

improving the source data, refining the methods, and releasing them on a regular basis—would allow users to create measures that perhaps better capture trends in economic well-being than GDP. As satellite accounts, however, these data are inherently of lower priority, and thus have received limited scrutiny and are also likely to be subject to resource constraints in the current era of tight statistical agency budgets.

2.3 Advantages of the GDP concept

Figure 1 shows that the level of GDP is fairly well correlated with broadly defined consumption, suggesting that GDP may not be a bad proxy for at least the market-based portion of economic well-being over long periods of time. On a shorter-term basis, the two measures may deviate materially, as can be seen by looking at quarter-to-quarter growth rates. Figure 2 shows these growth rates. GDP is generally more volatile than consumption. Among other things, GDP tends to fall more during recessions (denoted by the shaded bars in the figure) than consumption.

Figure 2: Quarter-to-quarter consumption and GDP growth



Shaded regions are NBER recessions. Source: Bureau of Economic Analysis (Haver Analytics)

From the perspective of policymakers trying to stabilize the economy at the business cycle frequency, such deviations are likely viewed as a feature rather than a bug of the GDP concept. The underlying argument echoes the discussion about the initial designers of GDP preferring a production-based concept in part because it better aligned with the Keynesian notion of aggregate demand. In particular, an important indicator of the health of an economy is whether economic resources are being fully utilized. In periods where economic resources are not fully utilized, unemployment is elevated and incomes are lower. It is thus unsurprising that government officials responsible for countercyclical monetary and fiscal policy would want to focus on an aggregate metric that represents all production done domestically—including that related to investment—even if that production does not immediately enhance consumer welfare. For example, the Federal Open Market Committee considers the deviation of real GDP from "potential GDP,"

in addition to other indicators like the unemployment rate and measures of inflation, when setting monetary policy.

2.4 Conceptual alternatives

Despite recent calls (most prominently, Stiglitz, Sen, and Fioussi, 2009) for shifting the emphasis of government statistical indicators from measuring economic production to measuring overall well-being, trying to broadly capture all the factors that enter well-being would be highly ambitious.[10] In addition to the challenge of accurately measuring all of the many factors that bear on households, one needs to grapple with how to weight different factors in order to produce a single comprehensive measure. (As Corrado, Fox, Goodridge, Haskel, Jona-Lasinio, Sichel, and Westlake, 2017, note, GDP effectively weights the units produced of different goods and services by their prices, which should correspond to the values of these items.) Of course, a single measure is not absolutely necessary—some proposals, such as the OECD's Better Life initiative, merely call for a "dashboard" of factors related to welfare. The strength of dashboards is that they allow users to apply their own weights; however, this is also a weakness when it comes to trying to reach consensus about how different countries compare or how much welfare has increased over time.

Jones and Klenow (2016) (and Bernanke and Olson, 2017) took one concrete step toward creating a broader measure of welfare that draws from economic theory to weight different factors. The authors use a "consumption-equivalent" welfare approach combining data on consumption, leisure, inequality, and mortality into a single summary statistic using an expected utility calculation that applies equal weight to each person. They go on to explore differences over time and across countries between this summary statistic and GDP, finding, for example, that their alternative statistic implies that living standards in Western European countries appear much closer to those in the United States because of longer life spans, greater consumption of leisure, and lower inequality.

An entirely different approach to capturing welfare would be to simply ask people how happy they are. Wolfers (2003), Stevenson and Wolfers (2008), and Sack, Stevenson and Wolfers (2012), for instance, explore measures of so-called "subjective well-being." Stevenson and Wolfers (2008) provide a thorough analysis of subjective well-being over time and across countries and conclude that such measures are fairly well correlated with absolute real income per capita (with some role for relative income). While there is some worry about the biases people exhibit when answering questions about their well-being (see Krueger, 2008) and uncertainties about how to aggregate responses to questions about happiness (Bond and Lang, 2018), these measures are potentially important complements to indicators of well-being based on hard data.

2.5 Summary

We have highlighted some important conceptual differences between GDP and aggregate economic well-being (which itself is much narrower than overall welfare). However, series that are already included in the standard national income accounts can be used to construct a measure that, on a conceptual basis, should be correlated with the bulk of the goods and services that determine economic well-being. Moreover, we show that such a measure—combined private and public consumption—is well correlated

. . .

10. See Coyle and Mitra-Kahn (2017) for one proposal along these lines.

with GDP over long periods of time, suggesting that changes in GDP, at least in principle, could be a fairly good measure of aggregate economic well-being over time.

We believe that one conceptual difference—the exclusion of non-market activities that bear on economic well-being—merits more attention, particularly given the potential for changes in the importance of such activities over time to alter the degree to which changes in GDP capture changes in well-being. However, even if this issue were important, it does not necessarily follow that the definition of GDP should be changed; rather, it suggests that we need to develop good alternative measures to supplement it. Retaining the current definition of GDP has the very significant advantage of keeping GDP comparable across time and across countries.

The discussion in this section has all concerned GDP *as conceptualized*. We now turn to the degree to which actual GDP reflects economic well-being, which depends heavily on how well GDP is measured in practice.

3. Issues related to measuring nominal GDP

Given our focus on economic well-being, we are ultimately interested in how well the official statistics measure *real* GDP—that is, GDP that abstracts from the effects of price inflation. The real GDP concept is, for the most part, estimated by collecting data in current dollars to produce the components of "nominal GDP" and then adjusting these components to remove price inflation, thereby leaving just the "real" activity in the economy. Real GDP measurement problems can thus arise from either errors in the estimation of nominal GDP or errors in the way that price adjustment (or "deflation") is done. In this section, we describe two major challenges in measuring *nominal* GDP.

3.1 "Free" goods

A recent source of discussion and debate is whether and how GDP should account for the vast amount of information, entertainment, and services that consumers obtain through the internet seemingly for free. This problem is not new—households have consumed entertainment and news services via television, for example, for many decades without paying directly for it. But, with internet-provided services an ever-growing part of our regular lives, there are increasing questions about the degree to which these services are already accounted for in GDP and whether they should be counted in GDP.[11]

Before delving into the national accounting issues surrounding free goods, it is useful to think very generally about how these transactions work. The development and maintenance of internet-provided "free" services is often funded by an interest in selling something, whether it be a traditional good or service (think shoes) or a premium product offered by the internet company supplying the service (think Spotify Premium). In the case of shoes, for example, the shoemaker pays dollars to advertising companies to create ads and pays dollars to internet companies (like Facebook) to place those ads, often targeting the ads using data that the internet company has collected from individual users. The internet company uses these dollars to develop content that induces individual users to look at ads and to give over their data. Meanwhile, individual users consume the content, are influenced by the ads, and ultimately use some of their earnings to purchase shoes.

. . .

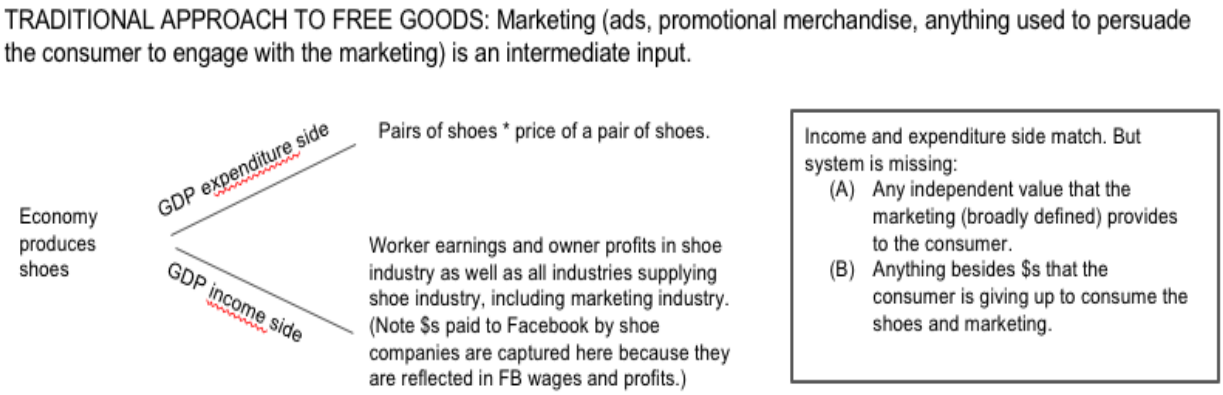
11. The statistical community is engaging in significant discussion of this particular issue and other implications of the "digitalization" of the economy. For example, the IMF held a two-day forum in November 2017 on "Measuring the Digital Economy" and the OECD has done extensive work on measuring the digital economy (see, for example, Ahmad and Schreyer, 2016); likewise, the Fifth World KLEMS Conference, held in June 2018, dedicated a plenary session to the issue.

Clearly much of what going on as part of these transactions is already captured in GDP and related parts of the national accounts. The dollars spent on shoes show up in nominal consumption (the "expenditure side" of GDP). The wages and salaries of the workers at the shoe company, their suppliers, and the internet company all show up in nominal national income (the "income side" of GDP), as do the profits of all of these companies. The key question, though, is whether these items are the only parts of the broader transaction described above that should be counted as part of nominal GDP.

The traditional national accounts approach views marketing broadly defined (including ads, promotional merchandise, and anything else that comes along with the ads such as internet content) as an intermediate input to the final good being promoted. This approach is applied, for example, in the context of "free" network television that is supported by advertisements. It means that the marketing does not independently contribute to GDP; it is captured in nominal GDP as long as we are capturing all of the dollars that consumers spend on the final good.

Figure 3 shows how the national accounting would work for the example above under the traditional approach. The economy is viewed as just producing shoes, and, as described above, the dollars that consumers pay for shoes show up on the expenditure side of nominal GDP and the dollars that that all workers earn and owners make show up on the income side of nominal GDP. The dollar amount is the same on both sides of the accounts so the two sides match (as should be the case with GDP accounting).

Figure 3.

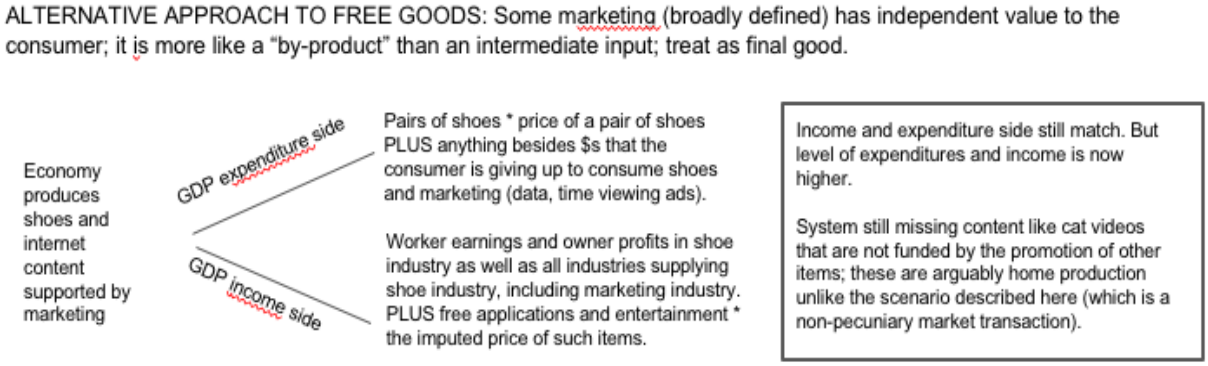


However, as noted in the box on the right side of Figure 3, there are parts of the transaction that are not counted under this approach. In particular, the approach would not count any independent value that the marketing broadly defined (including the internet-provided information, entertainment, and services) provides to the consumer. For many people these days, such items have become a regular part of their lives in highly important ways—they allow people to get home faster, keep up with friends, do their taxes, meet potential romantic partners, manage their finances, cook better, find information at blistering speeds relative to the past, and consume a rich offering of video and audio entertainment. The approach also does not include anything (besides the dollars spent on shoes) that the consumer is giving up to consume this marketing—more specifically, the time they spend viewing ads and the data that they provide that allows for better targeting of ads.

There is an alternative approach to measuring nominal GDP that would include these items.[12] This approach recognizes the independent value to consumers provided by some part of marketing broadly defined—viewing these parts more as a by-product of the final good being sold rather than an intermediate input. Under this approach, consumers are engaging in a transaction where they exchange their time spent viewing ads and data for the "free" internet content they consume. This transaction is very much a market transaction, albeit not a monetary transaction. But the absence of money in the transaction has not stopped statistical agencies from including other non-monetary transactions in GDP such as the "free" services that are provided by banks to account holders and the rental value of owner-occupied housing, both of which are imputed and then included as part of consumption.

Figure 4 shows how this alternative approach would work in the national accounts. In this case, both shoes and internet content supported by marketing are final goods. The value of the time spent viewing ads and the data that the consumer give up by consuming the internet content are akin to the dollars that they pay to consume shoes on the expenditure side of the GDP accounts. The internet content represents what they are "paid" for viewing ads and providing data, much like the earnings they receive for doing traditional work.

Figure 4.



As noted in the box, including these additional market (but non-monetary) transactions would raise nominal GDP, whether measured as expenditures or as income, while still leaving the expenditure and income sides of the accounts matching.

What would still not be counted in nominal GDP under the alternative approach would be internet content that is not created as part of an effort to promote a product—such as cat videos created as someone's hobby. One might view these services as largely home production or leisure, being arguably similar to the value provided to friends when one visits them. As a result, it would be appropriate not to include them in a GDP measure that excludes home production.

Thus far, our discussion has focused on the right way to capture "free" goods in nominal GDP *in principle*—doing so *in practice* raises a number of challenges, as does translating the imputed dollars

. . .

12. Much of our discussion here has been influenced by work done by Leonard Nakamura and his coauthors at the Commerce Department—see, for example, Nakamura, Samuels, and Soloveichik (2017).

involved into real GDP.[13] Nakamura, Samuels, and Soloveichik (2017) propose that the transaction can be captured (in nominal terms) by the money spent on conventional advertising and the marketing-related portion of "free" information or entertainment. While the former may be fairly straightforward to measure, measuring the latter is more difficult—the authors attempt to identify the costs associated with traditional in-house marketing activity at internet companies, but arguably some or all of the costs of software and application development should be included.

Translating nominal GDP for this category into real GDP is also complicated. In the absence of market transactions, the value could be ascertained by finding similar services that do have market prices, recognizing, of course, that the final good in these cases is the content, not the advertising. For example, one could use a deflator for entertainment in the case of videos or software in the case of free applications.

Alternatively, statistical agencies could measure real GDP in this area directly—by asking people directly about willingness to pay for free goods and services, as suggested by Corrado, Fox, Goodridge, Haskel, Jona-Lasinio, Sichel, and Westlake (2017). This approach has the potential to get around many of the challenges just described. Brynjolfsson, Eggers, and Gannamaneni (2017) sample very large numbers of people online about what compensation they would require to forgo use of a digital service like Facebook or email. They find that consumers would require significant compensation to give up a variety of free digital goods, in particular goods that are essential to many professions—the median willingness to accept to give up a service for a year is $17,530 for Search Engines, $8,414 for email, and $3,648 for digital maps. Brynjolfsson, Diewert, Eggers, Fox and Gannamaneni (2018) propose a new metric, "GDP-B," that would use results like these to augment GDP by the value of new goods and free goods; they find that a GDP-B statistic that includes just the value of Facebook would have grown measurably faster than published GDP between 2004 and 2017. This research not only refutes the view that accounting for "free" goods is simply unfeasible but also underscores the urgency for serious consideration of these issues.

Given that the consumption of "free" goods appears to be growing, more accurately capturing this consumption would lead to larger GDP and faster GDP growth than officially reported. The changes would partly flow through to productivity as well (though not entirely, given that consumers' "work" viewing ads and giving up data should, at least in principle, raise inputs as well).

3.2 Understatement of the domestic economic activity of multinational enterprises

As discussed by Varian (2016) and Guvenen, Mataloni, Rassier, and Ruhl (2017), the rise of global supply chains and the legal latitude that companies have in declaring in which countries their economic activity takes place lend material downward bias to estimates of U.S. nominal GDP. In particular, "transfer pricing" and other practices allow multinational enterprises (MNEs) operating in the United States to underprice the sale or lease of intangible assets—such as blueprints, software, or new drug formulas—to affiliates in low-tax jurisdictions so that more of their profits are booked in these countries.

The economic importance of such transactions has been documented in a variety of ways. For instance, in 2012, a Senate subcommittee questioned Microsoft about its agreements to shift some R&D costs and regional royalty rights to affiliates in Singapore and Ireland (U.S. Congress Senate Committee on Homeland Security and Governmental Affairs, 2012). In 2013, the subcommittee found that Apple

. . .

13. This does not need to be done separately for both the expenditure and income sides of the accounts; the two parts of the exchange should be equivalent so the calculation can be done just once and then added to both sides. The national accounts take this approach with the services arising from owner-occupied housing (with the homeowner viewed as both the consumer and the producer).

used favorable transfer pricing agreements to shift billions of dollars of profits from the United States to Ireland (U.S. Congress Senate Committee on Homeland Security and Governmental Affairs, 2013). More generally, Hines (2005) and Lipsey (2006) show that U.S. MNEs register more profits in tax havens than can plausibly be accounted for by economic activity. Jenniges, Mataloni, Stutzman, and Xin (2018) find that U.S. companies that have a cost sharing agreement with a foreign entity appear less productive than similar companies without such an agreement, and foreign companies that have a cost sharing agreement with a parent company in the U.S. appear more productive than similar foreign companies. A 2016 OECD brief described how such transactions drove a 26 percent increase in measured GDP in Ireland in 2015. And, Tørsløv, Wier, and Zucman (2018) estimated that nearly 40 percent of multinational profits are shifted to low-tax countries each year.

Under current methods, transfer pricing and profit shifting have led to an understatement of both nominal GDP and nominal gross domestic income (GDI). Consider the example of a smartphone whose software, blueprints, and branding are developed in the United States. If the phone is assembled in the United States, then the full value of the phone (priced at its market price) is included in GDP. If the phone is assembled abroad, then so long as the contract between the company doing the assembly (e.g. Foxconn) is an arm's length transaction, GDP will still be correctly measured, as it will include the value of the phone less the amount paid to the foreign assembler.

However, if a foreign-affiliate of the U.S. company is introduced in the transaction, GDP could end up understated. Here's one way this could happen: the U.S. company leases the rights to the intangible capital—the software, blueprints, and branding—to an affiliate in a low-tax country (say, Ireland) and it prices that lease at a value that is much less than its market value. Then the *Irish affiliate* contracts with Foxconn to do the assembly. Phones are then exported from Ireland to the United States and from Ireland to the rest of the world. In this case, only the value of the lease from the U.S. company to the Irish company will be included in U.S. GDP, and if this lease is priced at an artificially low level, U.S. GDP will be too low as well.

Under current methods, estimates of imports associated with sales of the phone in the United States will be too high because the economic activity associated with the leased assets is unlikely to be attributed to this country.[14] In particular, imports will be too high (because they will overstate the Irish content of the phone imported from Ireland), and exports will be too low (because they will understate the U.S. content of phones exported from Ireland to the rest of the world). The same bias would occur in GDI because of the understatement of the company's U.S. earnings. Note that this transaction works because there is intangible capital that is hard to value and hard to pin to a location, and because the Irish company is an affiliate of the U.S. company, so that it does not matter to shareholders whether the Irish affiliate or the U.S. headquarters books the profits.

This problem is of increasing concern both because of the evidence discussed above regarding the importance of profit-shifting in today's economy and, more generally, because of the growth in MNE activity in recent decades.[15] MNEs are now a large part of the global economy—according to Guvenon et

. . .

14. Imports are estimated using customs data, not tax data, so it would be possible for a company to report numbers that are inconsistent with their tax data (but a better representation of where the economic activity occurred). However, it seems unlikely that the company would do so.

15. The Tax Cuts and Jobs Act of 2017 sharply lowered the U.S. corporate tax rate, which should reduce the incentives for profit-shifting, but also lowered taxes on the foreign income of U.S. corporations, which should increase it. The net effect is likely a reduction in profit shifting (and an increase in measured GDP), but the magnitude of the effect is unclear. See Harris and Looney (2018) for further details.

al. (2017), they accounted for $4.7 trillion of global value-added in 2017, an amount that was about the size of the fourth largest economy in the world at the time.[16] The statistical community recognizes the issue, and the international statistical guidelines most recently adopted by the United Nations Statistical Commission (System of National Accounts 2008) called for estimates of the production activity of MNEs to reflect the economic ownership of intangible assets rather than the legal ownership (Moulton, van de Ven 2018). There are practical challenges associated with how to do so, and the BEA has yet to change its official methods to follow this guideline.

Guvenon et al. (2017) explore one way in which the guidelines might be at least partially implemented. The authors use confidential MNE survey data collected by the Bureau of Economic Analysis and reapportion the earnings of U.S. MNE foreign affiliates based on labor compensation and sales to unaffiliated parties. The authors' findings suggest that current practices have materially distorted estimated productivity growth at some points in the past—with an average annual understatement of growth of 0.1 percentage point from 1994 through 2004 and 0.25 percentage point from 2004 through 2008 (though no effect between 2008 and 2014). These figures represent a lower bound on the distortion, as foreign MNEs are probably also shifting some of their profits out of United States.[17] Using this method, Bruner, Rassier, and Ruhl (2018) find that accounting for profit sharing would increase the level of U.S. measured GDP by 1.5 percent in 2014.[18]

4. Translating nominal GDP into real GDP

The most complex set of issues relates to how nominal GDP is translated into *real* GDP. Measuring price changes correctly is central to this process. As we will explain, changes in production processes and the broader economy over time introduce significant challenges. Assessing the best way to proceed requires a clear understanding of the different ways to measure the price changes, with a particular focus on how to "correct" changes in observed prices for quality improvements.

. . .

16. The growth in MNE activity has coincided with growth of "factoryless goods producers" (FGPs) in the United States—which develop the intellectual property, manage the production process, and market but contract out the physical production to manufacturers in other countries. Discussions of the MNE problem often cite the rise of FGPs as related. Although FGPs are often associated with the high-tech industry, Varian (2016) emphasizes that production is off-shored in many sectors (for example, a U.S. company may develop the pattern for a sweater or the design for a toy in this country and send it to a foreign affiliate to produce). Prior to 2013, a wide range of intangible capital, particularly intellectual capital, was excluded from most measures of GDP because spending on R&D was counted as an intermediate expense. Corrado, Hulten and Sichel (2009) find that as much as $800 billion of intangible capital was excluded from U.S. published data between 2003 and 2009. As a result more than $3 trillion of business intangible capital stock was excluded, with significant consequences for measured economic growth and productivity. In 2013, the BEA began treating R&D as a fixed asset (part of investment) to address this issue (Moris et al., 2015). Chen, Gouma, Los, and Timmer (2017) suggest using global value chains to decompose the value of a product into the value added at each stage of production in order to measure returns on intangible assets.

17. The Guvenon et al. methods would also miss the contribution of items developed domestically that are used in production but do not generate profits because they are not proprietary, such as open-source software.

18. As some commentators have pointed out, the 2017 Tax Reform Act may increase reduce profit-shifting behavior in a way that reverses some of this bias (see, for example, Ip, 2017). The result would be an increase in measured GDP that does not correspond to an increase in actual GDP. As yet, the magnitude of any such effect is very unclear.

4.1 Some basic intuition

Nominal GDP (that is, GDP measured at current prices) increases over time because of increases in prices (inflation) and increases in real output (real GDP growth). This sub-section offers some intuition on how to split nominal GDP growth into these two components.

It helps to think about what real GDP is. It is an index of the quantity of goods and services produced in a given period of time—a measure that aggregates the number of tomatoes, haircuts, tractors, and so on. The level of real GDP is difficult to interpret on its own—what does an aggregate of tomatoes and haircuts mean?—but it is useful for measuring changes in production over time. A key question, then, is how changes in the production of the different goods and services are aggregated. This would not matter if the number of every good and service produced increased by the same amount from one year to the next—say, the economy produced 3% more of everything in year 2 than in year 1. In that case, growth in real GDP obviously would be 3%. But when the production of different goods increases at different rates—for example, when production increases 10% for tomatoes and 3% for haircuts but falls 2% for tractors—the weights put on each category will determine what number is reported for real GDP growth.

Equivalently, real GDP growth can be viewed as the growth in nominal GDP less inflation. As noted above, real GDP is mainly calculated this way, as most source data capture expenditures in current prices. BEA uses data on spending and producer revenues from a wide variety of sources to calculate nominal GDP, and then uses prices collected (mostly) by the Bureau of Labor Statistics (BLS) to deflate nominal spending to calculate underlying quantities and growth rates. The key question from this perspective is how to create an inflation index when prices of different goods increase by different amounts.

Two different economic frameworks have been used to divide nominal GDP into inflation and real GDP—one that is based on the perspective of the consumer, where the deflator is called a cost-of-living deflator, and one that is based on the perspective of the producer, or a producer price index. The cost-of-living index (COLI) is sometimes called an input price index, because it is measuring the value of the inputs into the consumer's utility function, whereas the producer price index is called an output price index, because it is measuring the value to the producer of the output they produce.

As noted by Moulton (2018), BLS's producer price index (PPI) is conceptually based on an output price index framework, whereas BLS's consumer price index (CPI) is based on a cost-of-living utility-based framework. But, as we show below, the actual practices used by BLS and BEA do not correspond exactly to either one of these approaches, but are, in general, pretty good approximations of both.[19]

4.2 Two theory-based approaches to measuring inflation and real GDP

The discussion that follows assumes that consumers (who are also producers) do no saving such that GDP (as well as aggregate income) equals consumption. We make this assumption to simplify the exposition, not because we think that GDP necessarily ought to capture consumption only. Also, because our aim is to

. . .

19. BEA uses a mix of PPIs and CPIs when deflating GDP, with PPIs used to deflate nominal quantities based on the revenues producers receive and CPIs used to deflate nominal quantities based on outlays made by the consumer. PPIs do not include sales taxes, for example, because producers do not receive them, but CPIs do, because consumers pay them. BEA uses CPIs to deflate most items counted in aggregate consumption (Personal Consumption Expenditures or "PCE"), but not all. For example, because consumption of medical services funded by Medicare and Medicaid is included in PCE, but not paid directly by consumers, it is not captured in the CPI. Medical expenditures in PCE are measured by the revenues received by medical providers, and are deflated by medical PPIs. For a detailed description of the particular deflators used to calculate real PCE, see https://www.bea.gov/help/faq/521

provide basic intuition about GDP measurement, we present a static model in which there is one representative agent, abstracting from the problems of heterogeneous agents and intertemporal considerations (see Aizcorbe (2014) for a discussion of the literature examining these issues).

Consumer's Perspective: The Cost-of-Living Index (a utility-based approach)

Under this approach, increases in real GDP are changes in the economy that make the consumer better off. If an increase in nominal GDP does not make the consumer better off, that increase represents inflation rather than an increase in real GDP.

When an increase in nominal GDP does make consumers better off, how do we know how much better off they are? Economists do not try to answer that question exactly, because we do not have a reliable way of measuring welfare.[20] Instead, we ask: how much has real purchasing power increased? When prices do not change, then the answer is simply: purchasing power has increased by exactly the amount that nominal income has increased, and real GDP growth is equal to nominal GDP growth. Similarly, if prices of all goods and services increase by the same amount, then the increase in purchasing power and real GDP is equal to the growth in nominal GDP less the inflation rate.

But when different goods and services have different inflation rates—that is, when relative prices do change—the answer is less clear. How much has purchasing power increased when income stays the same but the price of one good falls, for example?

Two approaches have been used within the COLI framework to determine how much purchasing power changes when relative prices change. The first asks: "How much income would consumers have needed in period 1 (before the price change) to get the same utility as they get in period 2 (after the price change)?" The difference between this amount of income and actual income is called the "equivalent variation." The second asks: "How much income would consumers need in period 2 so that they are no better off than they were in period 1?" The difference between actual income and that amount of income is another reasonable, but potentially different, measure of how much better off (in $ terms) consumers are in period 2 after a relative price change. This amount is called the "compensating variation." In both cases, the idea is to compare income in two periods under a counterfactual where prices have not changed.

The Technical Details

Figures 5a, 5b, and 5c show how to calculate equivalent and compensating variation. They are based on a model of a simple economy of just two goods, call them drinks (D) and food (F). For simplicity, we define the price of drinks as always equal to 1 (i.e., drinks are the numeraire). With this definition, we can read the amount of income from the intersection of the budget constraint and the Y axis. (For example, if the price of D is 1 and you can buy 100 units of D if you buy no F, your income must be $100.) In period 1, consumers maximize utility by choosing bundle D_1 and F_1, on indifference curve U_1.

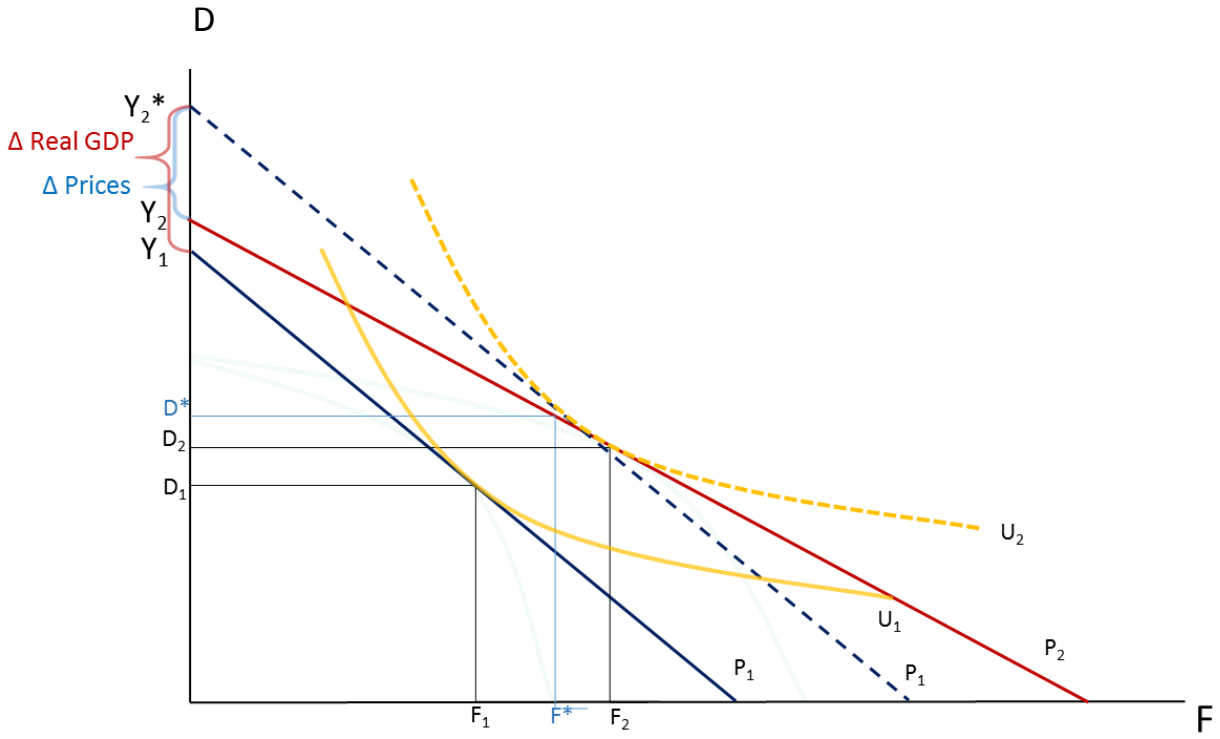
An improvement in the technology of producing F (say the introduction of an improved seed lowers the relative price of F in period 2. The budget constraint shifts out, nominal income increases from Y_1 to Y_2 and consumers buy D_2 units of D and F_2 units of F.

. . .

20. That is, we think welfare depends on real purchasing power, but typically do not quantify the relationship between real purchasing power and welfare. See footnote 25 for more detail.

Figure 5a shows how to calculate real GDP growth between period 1 and period 2 using period 1 prices. It shows that consumers would need to have income equal to Y_2^* in order to get utility U_2 at period 1 prices. With that income and relative price, consumers would choose bundles D_2^* and F_2^* . Thus, if prices had held fixed at P_1, consumers would have needed additional income $(Y_2^* - Y_1)$ in period 2 to make them as well off as they actually were in period 2 given the price change, and thus $(Y_2^* - Y_1)$ is a measure of how much better off they are. In other words, $(Y_2^* - Y_1)$ is the equivalent variation measure of welfare change from the decline in the price of F.

Figure 5a. Cost of living index (period 1 prices)

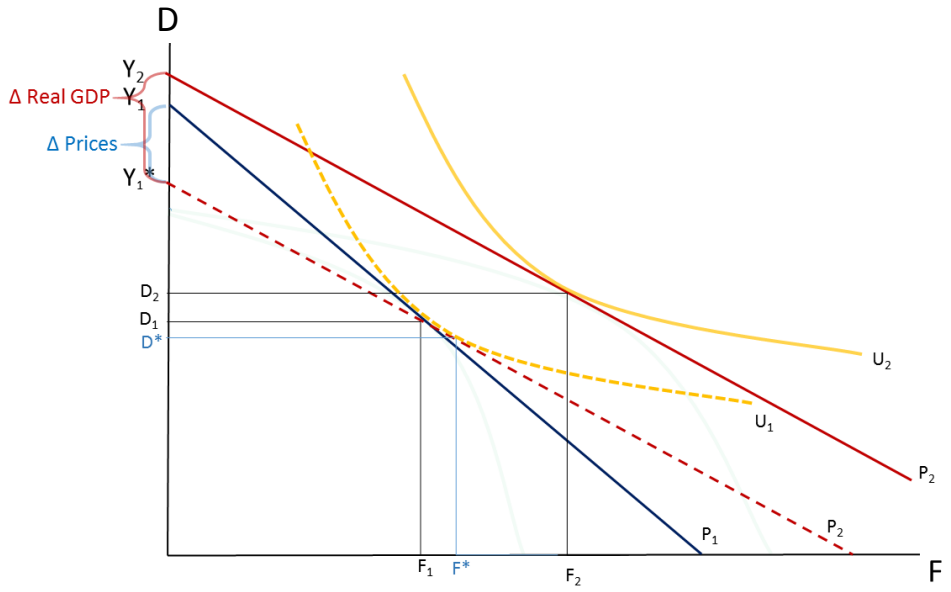


A measure of real GDP growth under this approach would then be the ratio

$$\frac{Y_2^*}{Y_1} = \frac{D_2^* + P_1 F_2^*}{D_1 + P_1 F_1} \qquad (1)$$

Figure 5b shows how to calculate real GDP growth using period 2 prices. It shows that consumers would only need income Y_1^* to get period 1 utility at period 2 prices, and thus $(Y_2^* - Y_1)$ is how much could be taken away from consumers in period 2 so that they are just as well off as in period 1: $(Y_2^* - Y_1)$ is the compensating variation.

Figure 5b. Cost of living index (period 2 prices)

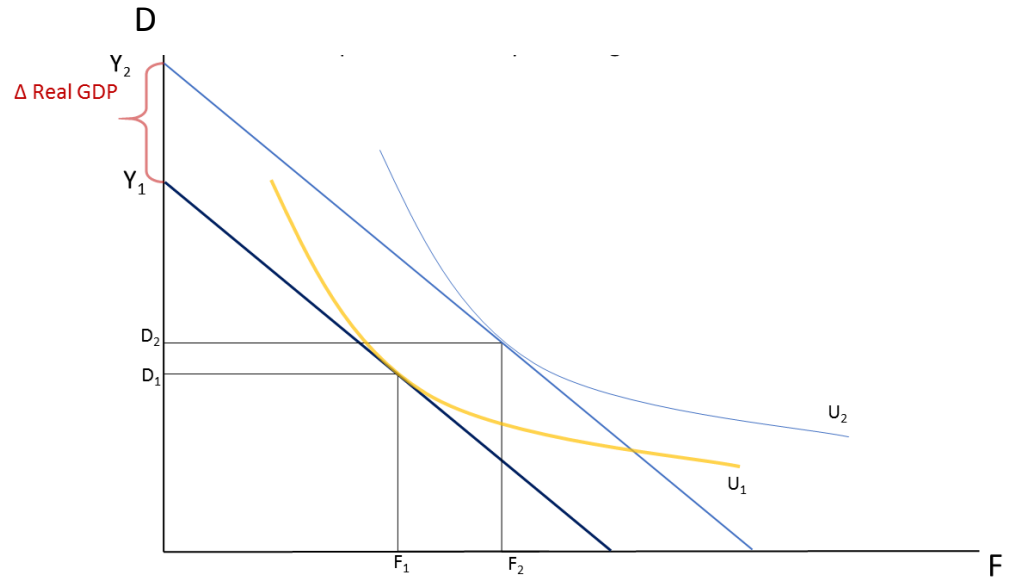


Measured real GDP growth under this approach is then:

$$\frac{Y_2}{Y_1^*} = \frac{D_2 + P_2 F_2}{D_1^* + P_2 F_1^*} \qquad (2)$$

In Figure 5c, there is no change in relative prices, but income increases from Y_1 to Y_2. Consumers would need an additional (Y_2-Y_1) dollars to make them as well off in period 1 as in period 2, and could have (Y_2-Y_1) dollars taken away in period 2 to leave them as well off as in period 1. In this case, the compensating and equivalent variation are the same, and measured real GDP growth is simply $\frac{Y_2}{Y_1}$.

Figure 5c. No change in relative prices; Equivalent = compensating variation



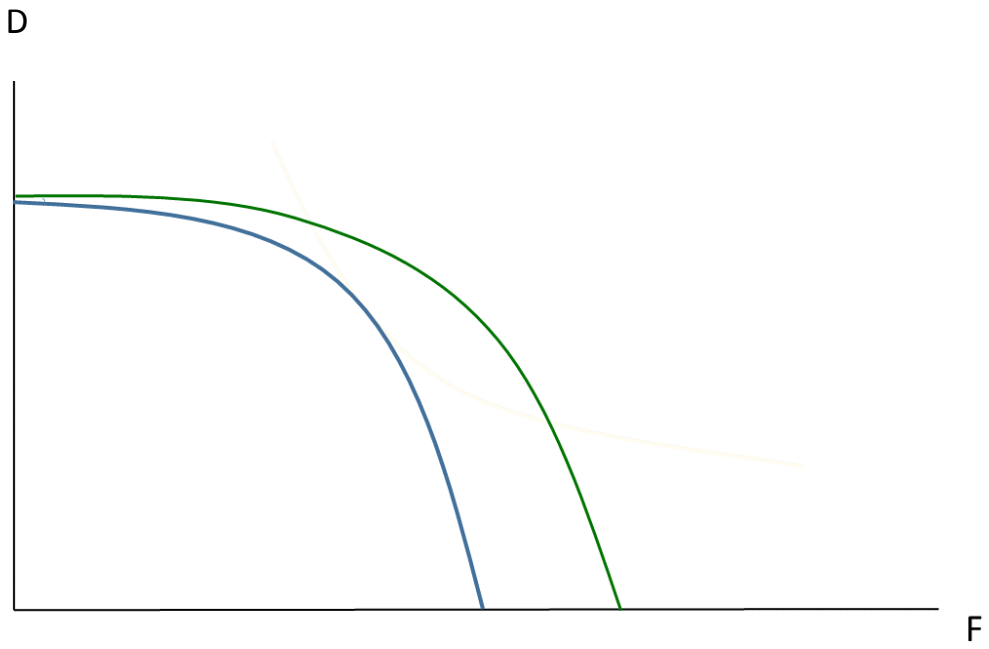
We show below how these COLI-based approaches compare to the method actually used by BEA when it calculates real GDP. But first, we discuss how to think about measuring real GDP growth from the producer's perspective.

Producer's perspective: The Output-Price Index (a production-function-based approach)

Under this approach, the focus of real GDP measurement is on production rather than consumption. Real GDP increases are captured in our stylized economy by the ability of producers of drinks to produce more drinks or the producers of food to produce more food. We use the term "production possibilities frontier" to define all the combinations of output an economy can produce.

Figure 6 shows an example of a production possibilities frontier for our economy. If an increase in nominal GDP is associated with the economy only moving *along* a production possibilities frontier—that is, if a change in the composition of output leads to an increase in nominal GDP, but the output produced this year could have been produced last year—it is counted as inflation. But if an increase in nominal GDP is associated with the economy being able to produce more than it could last year—for example, the change in the technology for producing food that lowers the resources needed to produce any given amount of food that is shown in Figure 6—then real GDP is viewed as having increased.

Figure 6. Production possibilities frontier for food and drinks: Technological improvement in production of food

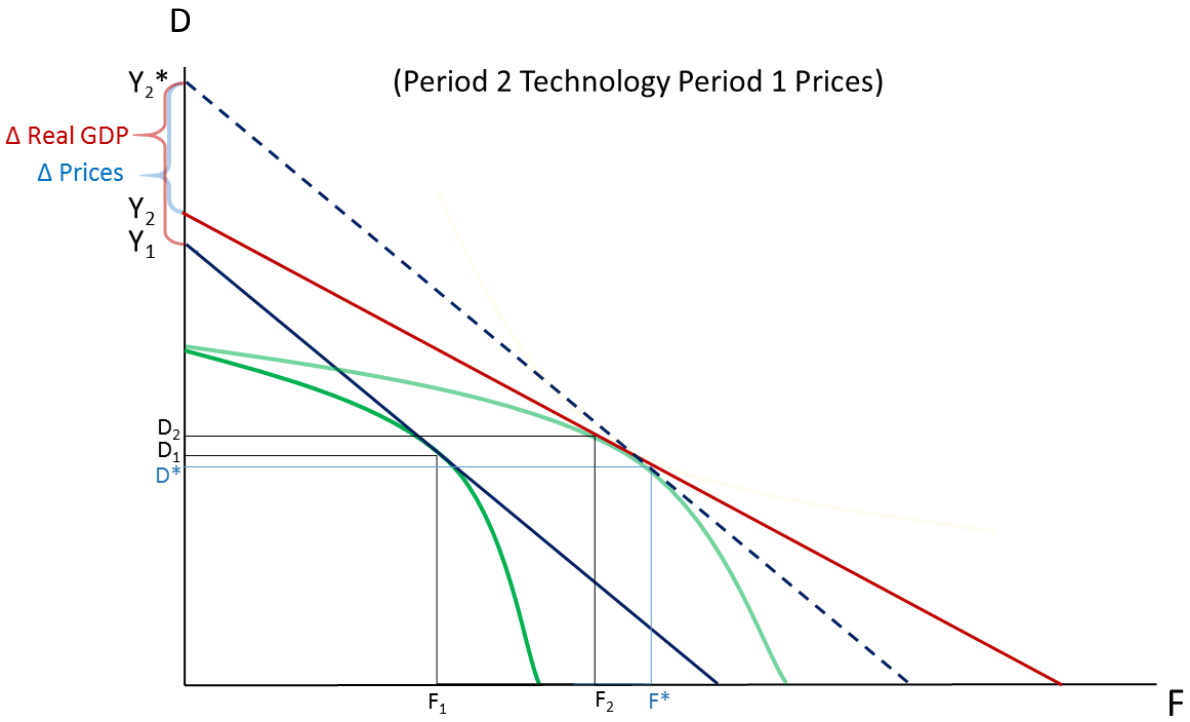


To measure how much real GDP has increased, the producer perspective asks the question: "If prices had been held fixed, how much more could producers have earned in year 2 than in year 1?" Analytically, the exercise is virtually identical to the one for the cost-of-living approach, but the counterfactual is different. Instead of asking how consumers would adjust the composition of their consumption (the *

quantities) under a counterfactual where prices remain fixed, it asks how producers would adjust the composition of their production.[21]

Figure 7a walks through the case described above, where a change in technology lowers the resources needed to produce any given amount of food, and we ask how much revenue producers would earn with period 2 technology but period 1 prices. At period 1 prices but period 2 technology, the producer would choose to produce F* and D*, and would earn income Y_2^*. The difference in producer revenues between period 1 technology and period 2 technology, holding prices fixed at period 1 prices, is a measure of the increase in real GDP, and would be captured by $Y_2^* - Y_1$. Figure 7b shows the calculation when the analysis is based on producer revenue holding prices fixed at period 2 prices. Figure 7c shows that, when relative prices are unchanged, the change in measured real GDP would be equal simply to the change in nominal GDP.

Figure 7a. Producer perspective



. . .

21. It is worth noting that neither counterfactual could have occurred, because there is only one unique equilibrium in each period. Had period 2 prices prevailed in period 1, markets would not have cleared: consumers would have wanted to purchase more of some things than producers were willing to produce, and producers would have wanted to produce more of some things than consumers were willing to buy.

Figure 7b. Producer perspective

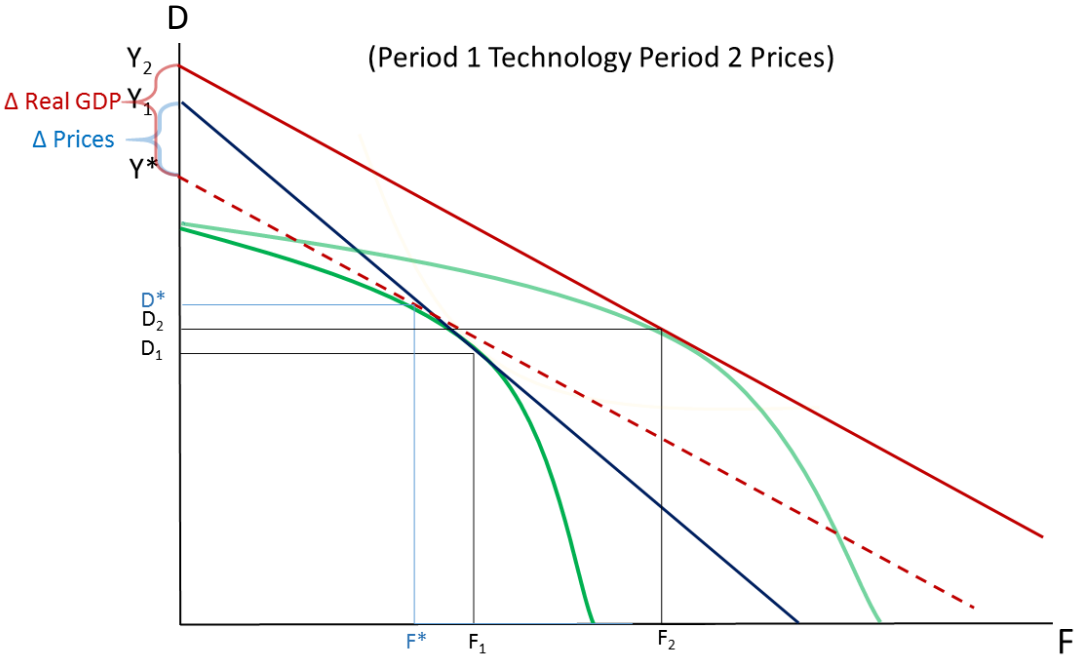


Figure 7c. Producer perspective

4.3 How these two approaches compare with what BEA actually does

Note that under both the consumer approach and the producer approach to real GDP growth measurement, the theoretically-appropriate real GDP index uses counterfactual baskets (D* and F*) that reflect the consumer's or producer's behavioral response to changes in relative prices. These counterfactual consumption and production bundles are unobserved, so cannot be used in the actual calculation of real GDP growth.

Instead of trying to estimate these counterfactuals, the approach used by BEA simply uses the actual observed baskets of what is produced/consumed to calculate real GDP growth. One building block is the Laspeyres Quantity Index, which asks: holding prices constant at period 1 prices, how much more is the period 2 basket worth than the period 1 basket? It is thus calculated as:

$$Q_L = \frac{Y_2^*}{Y_1} = \frac{D_2 + P_1 F_2}{D_1 + P_1 F_1} \qquad (3)$$

which can be rewritten as:[22]

$$Q_L = \frac{D_2}{D_1} S_1^D + \frac{F_2}{F_1} S_1^F \qquad (4)$$

where S_1^D is the share of D in GDP and S_1^F is the share of F in GDP in the first period. That is, for the Laspeyres quantity index, the growth in real GDP is just the growth in D and F weighted by their first period shares in the economy.

Another building block is the Paasche Quantity index, which asks how much more is the second period basket worth than the first period if we assume prices were second-period prices in both periods:

$$Q_P = \frac{D_2 + P_2 F_2}{D_1 + P_2 F_1} \qquad (5)$$

This equation cannot be rewritten quite as simply, but some basic algebra shows that Q_p is still essentially a weighted sum using period 2 shares in the economy as weights:

$$Q_P = \left(\frac{D_1}{D_2} s_2^D + \frac{F_1}{F_2} s_2^F \right)^{-1} \qquad (6)$$

In 1996, the BEA began calculating real GDP according to a chain index formula, which uses the geometric average of these two quantity indexes to create a real GDP index, known as the Fisher Quantity Index.

$$Q_{BEA} = \sqrt{Q_L Q_P}.^{23} \qquad (7)$$

As a comparison of equations (3) and (5) makes clear, when changes in GDP are <u>not</u> associated with relative price changes, the Laspeyres and the Paasche quantity indexes are the same. There is no question

. . .

22. $Q_L = \frac{D_2 + P_1 F_2}{D_1 + P_1 F_1} = \frac{D_2}{D_1}\left(\frac{D_1}{D_1 + P_1 F_1}\right) + \frac{F_2}{F_1}\left(\frac{P_1 F_1}{D_1 + P_1 F_1}\right) = \frac{D_2}{D_1} S_1^D + \frac{F_2}{F_1} S_1^F \qquad (4)$

23. As we see below, and as shown by Diewert (1976), the introduction of chaining made real GDP a better proxy for changes in real living standards.

about what the change in real GDP is in that case. It is only when relative prices change that the two measures yield different answers.[24]

The geometric average of the Laspeyres and Paasche indexes is generally a very good approximation of the average of the theoretically precise quantity indexes, whether from the consumer perspective or the producer's perspective (Diewert, 1976, Triplett, 1993). That means that, given chaining, the index used by BEA to calculate real GDP is *conceptually* a very good approximation of the change in economic welfare, defined in monetary terms, using either the consumer's perspective or the producer's perspective. That is, abstracting from the important questions of scope we discussed in the first section of the paper, real GDP growth as measured by the BEA is a good proxy for the change in real resources available to a society, which is the closest we can get to the change in welfare.[25]

The conclusion that the change in real GDP represents the change in (the monetary value of) welfare from the market economy is sometimes disputed. We attempt to provide intuition and dispel some related misconceptions in several ways. In Boxes 1 and 2, we address the sometimes-heard idea that "GDP is a rectangle but consumer surplus is a triangle so they are not the same." Box 1 provides a numerical example with a production function and a utility function to show how changes in real GDP are a good proxy for the changes in welfare that one would calculate using the theoretically-precise CV and EV. Box 2 provides a graphical analysis to show that the change in real GDP from a price change is essentially equivalent to the change in consumer surplus associated with that price change. (As shown by Willig (1996), consumer surplus, the area above the price line under the demand curve, is itself a reasonable approximation of the theoretically ideal CV and EV measures of the welfare changes arising from a price change.) In Box 3, we discuss why "the paradox of diamonds and water," which explains why the price of a good may not equal its value to the consumer, is not inconsistent with the view that changes in real GDP measure changes in welfare.

A numerical example

To demonstrate the quality of the approximation, consider the following example. Assume that the only input to producing D and F is labor, and that the amount of labor is 100. The consumer's utility function is U = log(F) + log(D). In the first year, the production function is D = $L_D^{1/2}$ and F = $L_F^{1/3}$. There is a technological advance in the production of F in period 2. In the second period, F = $L_F^{2/5}$. With this simple setup, if we assume that the equilibrium is one where consumers and producers are maximizing utility and profits, respectively, we can calculate the BEA's chained price and quantity indexes, as well the theoretically-accurate cost-of-living index and output-price index.

. . .

24. In other words, as noted above, if prices do not change, then the change in real GDP is equal to the change in nominal GDP. If prices all change by the same amount—and relative prices do not change—then the change in real GDP is equal to the change in nominal GDP less the rate of price change.

25. To be more concrete and technically precise: Consider the indirect utility function $V(Y,P_a, P_b\ldots)$. Define a cost-of-living index P as the price index that satisfies: $V(Y/P) = V(Y,P_a, P_b,\ldots)$. That is, a cost-of-living price index is defined as the price index that captures the changes in welfare arising from changes in the prices of the underlying goods.
Define $V(Y/P) = \bar{\lambda}\,(Y/P)$ where $\bar{\lambda}$ is the average utility of real income. Real GDP is Y/P, or nominal GDP deflated by a cost-of-living index. It is a measure of consumers' real purchasing power. $\bar{\lambda}$ is the parameter that translates real resources into utility. If $\bar{\lambda}$ is constant, then, real GDP is proportional to welfare. If the average utility of income changes with real GDP—for example, if $\bar{\lambda}$ declines as real income increases—GDP will not capture this change, and it is not intended to. But holding $\bar{\lambda}$ constant, welfare is directly proportional to real GDP.

These three measures are shown in Table 1. To calculate inflation and real GDP growth in each column, we use the geometric average of the Laspyeres and Paasche measures.[26] As can be seen from the final two columns, the measures are in practice very similar. Why is that? It is because, in equilibrium in this simple economy, market prices represent both the ratio of marginal production costs of the two goods and the ratio of marginal utilities.

We concluded above that the deflators used by BEA yield a measure of real GDP change that is in general a good measure of the change in consumer welfare (measured in monetary terms) and the change in production. This example validates the conclusion quantitatively and, relatedly, shows that the two deflators concepts are very similar in practice, suggesting that the choice of deflator should not matter much to the interpretation of real GDP as a measure of well-being. Below, we explore whether this holds true when it comes to changing quality of goods.

Table 1
Comparing the BEA Deflator, the Cost-of-Living Index, and the Output-Price-Index

	Prices		Quantities		Inflation	Real GDP Growth
	Laspeyres Index	Paasche index	Laspeyres Index	Paasche index		
BEA Deflator	0.89	0.87	1.15	1.12	-11.7%	13.3%
Consumer Perspective	0.88	0.88	1.13	1.13	-11.7%	13.3%
Producer Perspective	0.90	0.86	1.16	1.11	-11.8%	13.4%

Note: The indexes calculated here assume that the economy is in equilibrium, with producers maximizing profits and consumers maximizing utility. D is drinks, F is food, and L is labor. The first year production functions are: $D = L_D^{1/2}$ $F = L_F^{1/3}$. The second year production functions are: $D = L_D^{1/2}$ $F = L_F^{2/5}$. The utility function is U = log(D) + log (F) and L = 100 in both years.

4.4 Quality change and new goods

The discussion above assumed that the quality of goods and services was fixed, so that the only way that production and welfare could increase was for the actual quantities of goods and services to increase. In reality, of course, the quality of goods and services changes over time, and these quality improvements also yield improvements in welfare and GDP. That is, instead of having more goods and services, real GDP and welfare can improve by having better goods and services.

In order to use the formulas described above to compute changes in real GDP when the quality of goods and services improves, two often-unobserved prices are needed: the price of year-1-quality items in year 2 (to compute how much the year 1 basket would have cost with year 2 prices) and the price of year-2-quality items in year 1 (to compute how much the year 2 basket would have cost in year 1). With these prices, everything flows through as above.

Although these prices are unobserved, it is possible to impute them in some cases. The standard approach—the hedonic method— explored in depth by Triplett (1983), is to view goods as combinations of

. . .

26. For the consumer and producer perspective quantity indexes, we use the term Laspeyres when first period prices are used and Paasche when second period prices are used.

their underlying characteristics. People purchase goods and services because they value their underlying characteristics. Under this approach, when a good's quality improves, that good must now embody more of a particular characteristic. If the characteristics can be measured and priced, then it is possible to impute the prices for items of different qualities. For example, if you know in a particular year that each additional 100 square feet in a house raises the selling price by 5%, then it is possible to impute prices of houses with various square footages. If, in year 2, the average square footage of houses has increased, it is possible to know what those houses would have sold for last year; similarly, it is possible to know what smaller houses (year 1 houses) would sell for this year.

In a simple world where *the set of embodied characteristics is fixed,* there would be no issue of changes in quality if characteristics were priced instead of goods. This caveat about a fixed set of characteristics is important. It means that "better" goods and services are just combinations of existing goods and services. For example, an increase in the size of a cereal box might mean the cereal box is "better", but one could have purchased as much cereal before by buying two (smaller) boxes. Similarly, an increase in computer "MIPS" – millions of instructions per second—means that you need fewer computers to accomplish a particular task, but that task was doable the year before as well. When all characteristics exist in two consecutive years, the necessary adjustment for quality change is theoretically straightforward and intuitive, as we show in Case 1 below.

When a product improvement creates something that is actually new—an improvement in a medical treatment that increases survival, a cell phone app that allows users to figure out where their kids are, a printer that can print in 3-D, etc. then market data alone won't be sufficient to impute the prices necessary to calculate the indexes described above (because the item did not exist previously). In that case, a quality improvement has to be treated as a new good. In this section, we provide an overview of how to account for quality when the characteristics approach applies and when a good needs to be treated as new.

Case 1: Quality Improvements Embodying More of Existing Characteristics

Table 2 presents a very simple example to demonstrate this approach. Imagine that instead of buying "food", the consumer buys "boxes of cereal." What goes into consumers' utility function is not the box of cereal, of course, but the cereal itself. Imagine that there is no change in the underlying production technology for the economy—both a unit of cereal and a unit of drinks need the same amount of labor to produce in period 1 as in period 2—but for some reason, producers have decided to sell cereal in larger boxes—let's say doubling the cereal content— and the price of the box of cereal goes up accordingly.

Table 2 shows how this increase in quality can distort measured prices and quantities using Fisher (geometric average of Laspeyres and Paasche) quantity and price indexes for both. The price of a box of cereal doubles, because it contains double the cereal. Nominal GDP—which is equal to total nominal spending (the number of drinks times the price of a drink (1) plus the \number of cereal boxes times the box price) is unchanged. Without making an adjustment for this improved quality, though, it looks like prices increased 41% and real GDP decreased 29% (using Fisher ideal indexes to calculate both). But it is obvious that real GDP and prices would be unchanged with appropriate quality adjustments to the price of cereal boxes.[27]

. . .

27. Triplett (2004) argues that this package size adjustment is too simple, because the relationship between volume and price is generally not linear—that is, bigger boxes usually sell at a lower price per unit. We ignore this issue in order to provide a simple example that gives the basic intuition of quality adjustments.

Table 2

Quality Adjustment: Example 1
Cereal box becomes larger; No change in underlying technologies

	Drinks (price==1)	Cereal Boxes	Box Price	Nominal GDP	Cereal per Box	Cups of Cereal Purchased	Price per Cup
Year 1	8	4	2	16	1 cups	4	2
Year 2	8	2	4	16	2 cups	4	2
Percent Change	0%	-50%	100%	0%	100%	0%	0%

Price and Quantity Changes

	No Quality Adjustment	Correct Quality Adjusted
Inflation:	41%	0%
Real GDP Growth	29%	0%

The calculations are all very simple in this example because there is only one defining characteristic—the quantity of cereal—and it is directly observable. It is intuitively obvious that one would adjust the price of the cereal box for the change in quantity. But the exact same issues arise when the underlying characteristic may be not so easily observed and when products can improve on multiple dimensions.

When the proper quality adjustment is not so obvious, two approaches have been advocated. One is based on cost, which is generally linked to the producer's perspective and the output-price index. The other is based on utility, which is viewed as the appropriate method for the consumer's cost-of-living-index perspective. It is helpful to compare these two methods in this simple case.

Cost-based method. This method of adjusting for quality involves asking the producer how much the change in quality cost and subtracting that change—marked up to a selling price—from the box price in year 2. The box in year 2 contains an additional cup of cereal compared with the box in year 1. The producer's cost plus any markup for this additional cup amounts to $2. To quality-adjust the price, subtract $2 from the year-2-box price to get a quality-adjusted price of $2, or the same as in the first period.

Note that, in this case, one gets the same answer using the period 2 basket instead. How much would the period 2 box have cost the producer in period 1? An additional $2 dollars. Add that to the cost of the period 1 box to get a quality-adjusted period-1 price of $4, the same as the second period. In both cases, the price index adjusted with the cost-based method shows no change in the quality-adjusted price of cereal boxes.[28]

Utility-based method. A second method of adjusting for quality is to subtract from the price of the period-2 product the consumer surplus received as a result of the productivity improvement. Since the

. . .

28. This symmetry will not always occur. In particular, when goods have multiple characteristics, instead of just one, the effect on price of a reduction in price and increase in quantity of just one of those characteristics may depend on the set of characteristics (i.e., the particular basket) being priced.

marginal value of a cup of cereal is worth $2 to the consumer, that extra cup of cereal in the larger cereal box is worth $2. Subtracting that from the price, the quality-adjusted price is $2.

The cost-plus-markup method is a way of directly estimating the prices that period-1 and period-2 boxes would have sold for in the market had they been sold. A common approach to gathering those prices is to use hedonic functions – regressions that relate selling prices to the characteristics of the goods and services sold in a given year.

The basic idea of hedonic regressions is that, if there are enough different models of similar goods with varying amounts of underlying characteristics, a regression analysis, using cross-sectional variation within a given year, can uncover how much having more of certain characteristics contributes to a price. With this regression, any given combination of characteristics can be priced.[29]

The utility-based method does not try to directly measure the prices at which the goods would be sold but the difference in value to the consumer of the different quality items. The utility-based method yields the same answer because, in equilibrium for most goods, people purchase goods and services until their marginal value equals their price.

The equivalence between a cost-plus-markup quality adjustment and a utility-based adjustment breaks down when a quality improvement introduces a characteristic that was not available previously. We go through that case now, and argue that the utility-based approach is the appropriate one when the two approaches differ.

Case 2. Quality improvements that introduce something new

When a quality improvement introduces a characteristic that was not available previously, the good can be viewed as a "new" good. New goods might be, for example, treatments that increase survival time for cancer, the smartphone, or the laptop. The value to society of a new good is the difference between the value people place on the new characteristic embedded in the new model and its cost.

Standard methods that attempt to measure the prices of new goods in the previous year (hedonics or the cost method) are likely to be unsuccessful, since the goods did not exist. For example, when a camera is first added to a phone, there are simply no observations with cell phone cameras in the previous year's data, so no way to infer what such a product would have sold for.

As Hicks first showed (Triplett, 2004), the correct price to use for the previous year's price is the consumer's reservation price[30]— the lowest price at which consumers would not purchase the good.[31] As shown in Figure 8, the decline in price from the reservation price to the introduction price will capture the consumer surplus that consumers derive from the introduction of the good.

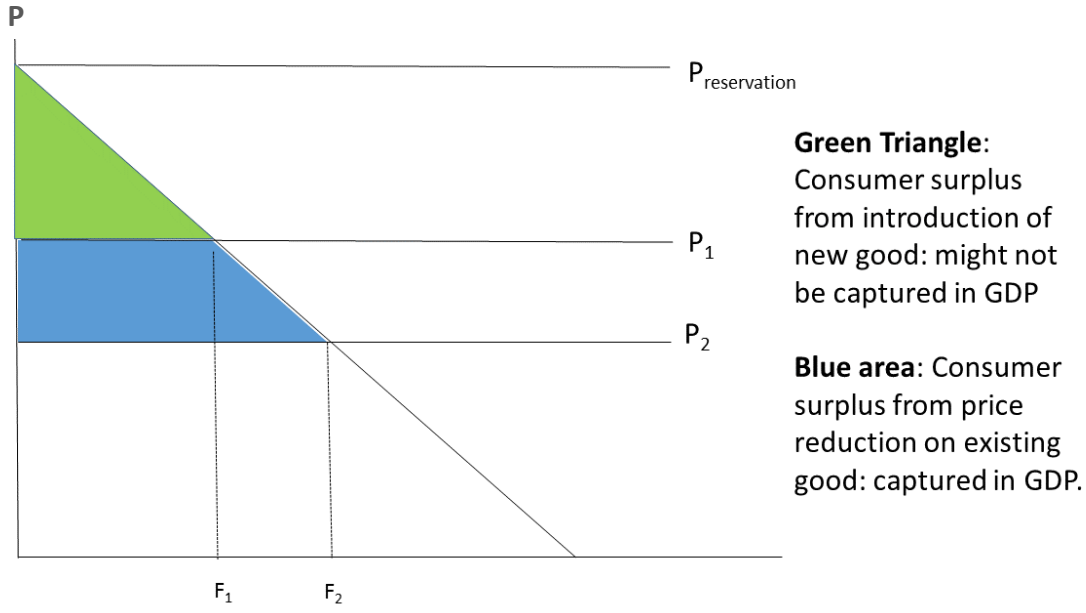
. . .

29. Of course, this is putting a functional form on the relationship between characteristics and prices. Plugging in characteristics that are very different in magnitude from the ones actually observed will likely be quite problematic. For example, a regression of house prices on square footage might not do a very good job predicting the cost of a 100 square foot house, or a 50,000 square foot house, if the range of houses actually observed in the data range from 1000 to 7000 square feet.

30. This is obviously the correct price from the consumer's perspective. From the producer's perspective, the correct price might be: at what price would you have produced this new product last year, but that price might be infinity if the technology simply did not exist.

31. When a new good that is just a combination of existing characteristics is introduced, the reservation price is equal to the price of the characteristics in the previous year, so this method works for any good.

Figure 8. Consumer surplus from new goods



It is not always clear when a good is new, versus when it simply represents a rebundling of existing characteristics. One important case where this situation arises is when there is a technological limit that means that people cannot get as much of a characteristic as they like, holding all else equal. This could be the case for computer chips that increase computing power without increasing the size of the chip, allowing for the production of laptops, smart phones, smart watches, etc., and many other areas where innovation is pushing the technological frontier. In these cases, the hedonic method may appear feasible—because characteristics (like computing power, for example) might not appear new—but something about the improvement is introducing a characteristic that was not available before, and, as Triplett (2004) notes, "the arrival of a new characteristic cannot be evaluated satisfactorily with hedonic methods."

One important area where this issue arises is in medical advances. Consider an example where consumption equals the consumption of everything excluding medical treatments. Assume that people undergo medical treatments not because they get utility directly from them, but because they get increased longevity—that is, years of life is the "characteristic" that goes into the utility function.[32] A

. . .

32. As discussed above, the treatment of consumer durables in the national accounts is a source of mismeasurement. Instead of counting consumer durables purchases in investment, and then also imputing a return on them over their service life (the way the accounts treat owner-occupied housing), they are treated as if they are entirely consumed in the year purchased. The same criticism can be applied to spending on medical care, much of which can be viewed as an investment in human capital that yields returns over many years. Recognizing that medical care is durable should not change the correction for quality—the quality adjusted price of medical care should still include a correction for the present value of the additional years of life. However, that value should then be spread out over many years as an imputed service flow into consumption, just as with owner-occupied housing. Similarly, Corrado and Byrne (2017) argue that spending on many consumer digital services, like programs that are purchased and then used for several years, should be counted as investment.

quality improvement in medical treatment occurs when the number of years of life produced by a treatment increases.

The cost-based approach to valuing this quality improvement would be to adjust the price of the new treatment for the cost incurred, plus markup, for the change in years of life, as in the cereal example above. This is essentially a hedonic approach that attempts to correct improvements in treatment over time by subtracting out the effect of the quality improvement on price. This is basically the approach taken by Romley, Goldman, and Sood (2014) in measuring the productivity of health spending.[33] In contrast, the consumer's cost-of-living utility-based perspective would value these improvements at the marginal value of life. This approach is taken by Lakdawalla et al. (2015) and Cutler et al. (1998, 2001). Of course, to implement this approach, it is necessary to have a reasonable measure of how much consumers value an increase in a year of life, and this measure cannot be gleaned from the data on medical spending itself. In practice, valuations of a year of life are derived from research that assesses how much people need to be compensated in order to take on risk. For instance, Viscusi and Aldi (2003) summarize the findings of over 100 studies that estimate the value of life using mortality and injury risk premiums.[34]

Efforts to compare these two approaches with actual data show that they yield very different approaches. Sheiner and Malinovskaya (2016), for example, use data from 1984 to 1994 from Cutler et al (2001) to show that a price index that subtracts the estimated value of additional life-years for heart attack treatment shows large declines in quality-adjusted prices, whereas a price index that subtracts the estimated market price of the additional years of life shows much smaller declines. They also show that even when the price of the valued characteristic – years of life – increases, the consumer can still be better off.

Dauda et al (2017) use that insight to compare these two types of quality adjustment with more recent data for three different treatments: treatment of heart attacks, heart disease, and pneumonia. They find similar results for all three conditions, with utility-based treatments showing much larger price declines than cost-based quality adjustments. Their findings for heart attack treatment, for example, are reproduced in Figure 9.[35]

. . .

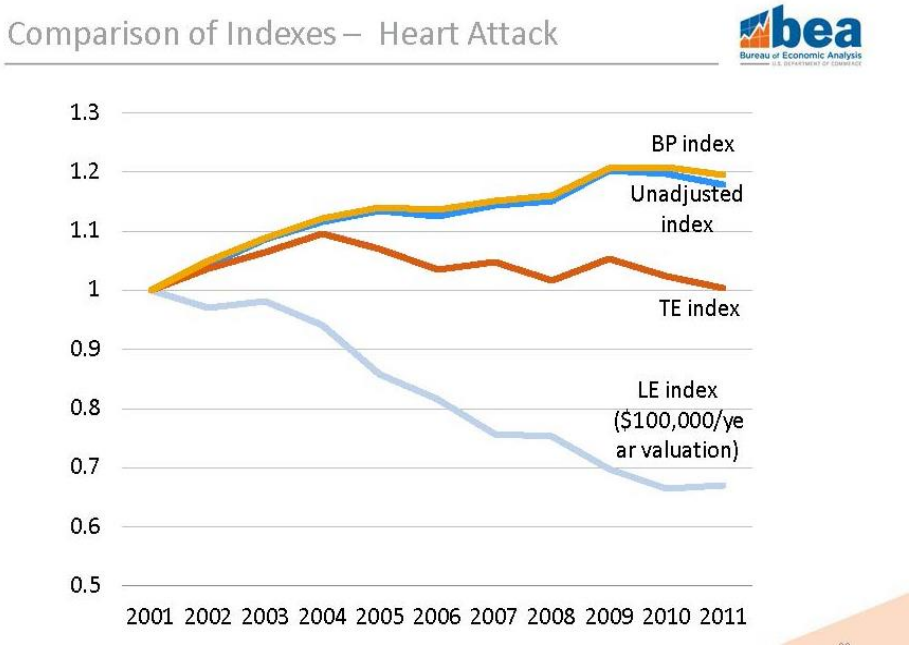
33. Romley et al actually calculate the spending per year of life, like the cereal example above. Sheiner and Malinovskaya (2016) show that this is the same as the cost option assuming that costs are linear in years of life, but won't be the same if there are other attributes of a treatment that affect cost.

34. Federal agencies such as the Department of Transportation, the Food and Drug Administration, and the Environmental Protection Agency depend on these types of studies to value statistical lives when estimating the benefits or costs of policies that affect life expectancy. None of the agencies allows the value of a life to depend on age, and, equivalently, none of them place a value on an additional year of life. (Aldy and Viscusi, 2007).

35. In that figure, the BP and unadjusted index are basically the current price index, the TE measures the change in the price of a successful treatment over time, and the LE index subtracts the marginal value of the quality improvement in heart attack treatment from the price, where the value of an additional year of life is assumed to be $100,000.

Figure 9. Comparing cost and utility-based methods of quality adjusting heart attack treatment prices

Reproduced from Dauda et al (2017)



Notes: The BP/Unadjusted Indexes are basically price indexes with no quality adjustments. The TE index is the cost of an effective treatment over time, where effectiveness is defined as an increase in survival probabilities. The LE index subtracts the value of additional years of life from more effective treatments over time, where it is assumed that a year of life is worth $100,000. See Dauda et al (2017) for details.

Why are the results in this scenario different from those shown above? They are different because, even in equilibrium, the cost of an additional year of life is not the same as the value placed on that additional year of life by consumers.

To illustrate this point concretely, imagine that an additional year of life is worth $50,000.[36] In period 1, the cost of a cancer treatment is $30,000, and it increases survival by one year. In period 2, the treatment has improved and has become more expensive. It now costs $70,000 but provides two additional years of life.

The average price of a year of life has increased from $30,000 to $35,000. If consumers could have purchased as many years of life as they wanted at $30,000 per year in period 1, then they would be worse off in period 2. But because this is not feasible, they are better off in period 2: they pay $40,000 more than in the previous year, but what they get is valued at $50,000. A cost-of-living index would show a *decline* in the price of health.

Note that in both cases, adjusting for quality lowers the inflation rate of medical treatments relative to an unadjusted price. The unadjusted price of treatment increases from $30,000 to $70,000, or 133%. But:

. . .

36. The literature generally assumes that the value of a year of life is given, so that it is the same whether a treatment gives you one additional year or two. This is equivalent to having the demand curve be linear in the relevant area.

A cost-based adjustment would show that the price of treatment was unchanged.

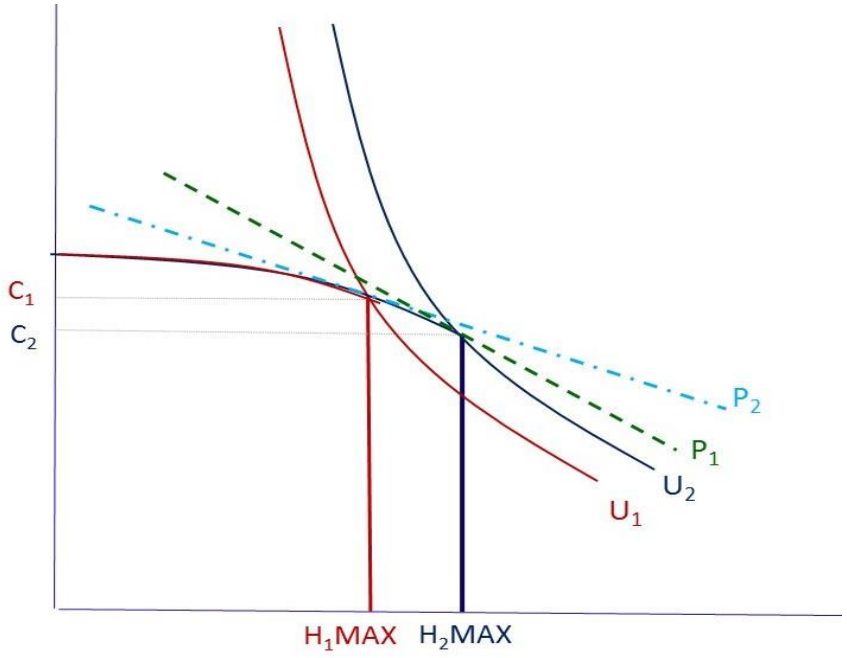
- Subtract $40,000—the incremental cost to produce an additional year of life—from the period 2 treatment price to hold quality at the year 1 level. Then the adjusted price of treatment in period 2 is $30,000, the same as the period 1 treatment price. Note that, in this sense, the cost-based adjustment strips out the effect of the quality improvement from the price.[37]

A utility-based adjustment would show a price decline of 33%.

- Subtract $50,000—the price of an additional year of life in period 2—from the period 2 treatment price to hold quality at the year 1 level. Then the adjusted price of treatment in period 2 is $20,000, 33% lower than the period 1 treatment price.

Figure 10a provides a graphical explanation of this problem. In period 1, the maximum years of life that can be produced is H_1Max, and the production function has a kink. Because of this, there is a large wedge between the value of the medical treatments being produced and the cost—where the cost is read off the production possibilities frontier and represents how much extra C could be produced if one *fewer* unit of H were produced. In period 2, this constraint is relaxed, and the maximum year of life moves to H_2Max. The price per year of life increases from P_1 to P_2, but the consumer's utility increases, moving out from U_1 to U_2.

Figure 10a. Consumer is at a corner solution; marginal utility is less than marginal cost

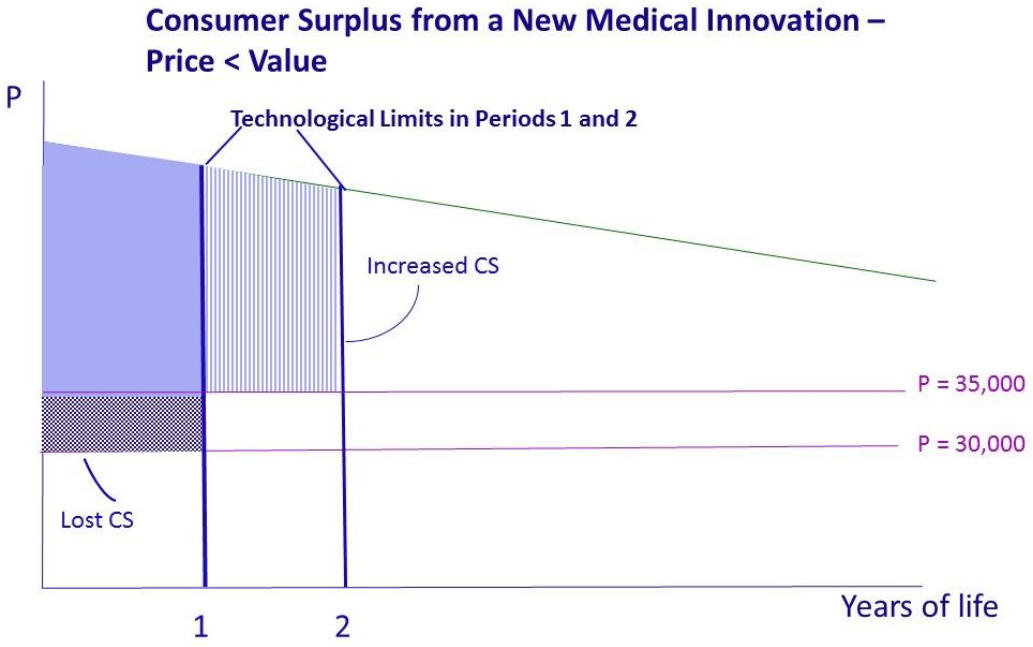


. . .

37. But, as shown above, the correct price to use for last year's treatment price is the price that would prevail in period 2 if it were still sold. For consumers to purchase the old treatment given the availability of the new treatment, the price would likely have to fall. For consumers to be completely indifferent, the price would have to fall $10,000, the consumer surplus that consumers get when they use the new treatment.

Figure 10b shows the same intuition from the perspective of consumer surplus. In the first period, the years of life from the treatment is just 1 year. At a price of $30,000 consumers would wish to purchase much more than 1 year of life, but they can't. In the second year, the quantity limit increases to 2 years, and the price per year increases to $35,000. The loss in consumer surplus from the price increase for the 1st year is more than offset by the gain in consumer surplus from the relaxation of the quantity constraint.

Figure 10b. Consumer is at a corner solution; price increase associated with increase in consumer surplus.



In this case, the framework for the deflator choice matters. Subtracting the cost of the improvement will yield a different result than subtracting the utility value. Furthermore, no matter the perspective—the producers or the consumers—the innovation that allows improved medical care represents an increase in GDP, with the production possibilities frontier shifting out, and consumers on higher indifference curves. Thus, the utility-based framework is the one that best captures the increase in real GDP.

Many researchers, including Groshen, Moyer, Aizcorbe, Bradley, and Friedman (2017) cite Triplett (1983) to argue that a cost-based method is the correct method for output price indexes. However, Triplett's paper only considered cases where there were no new characteristics—he did not try to account for new goods.[38] But, as we show in further detail below, a cost-based method will not capture innovations

. . .

38. This focus on whether an improvement cost anything might have more to do with the question of the scope of GDP vs welfare than the method to measure it. For example, if for some reason having nothing to do with private business people were happier, a pure cost-of-living index might suggest this increase in happiness can be measured as an increase in GDP, whereas a production based approach would say that nobody produced anything more so real GDP hasn't increased. We believe this is correct, which is why we limit our discussion to economic well-being, rather than overall wellbeing.

that shift out the production possibilities frontier through the introduction of new characteristics, which is the underlying economic test for whether real GDP has increased from the producer's perspective.

It is also important to note that a cost-based method can yield different results depending on (1) the path taken to get to a certain innovation/cost point and (2) the producer's pricing power. In Table 3, we compare the cost and utility methods of quality adjustment for a number of different cases. Scenario 1 is the one described above, where the cost method yields no increase in real GDP, and the utility method yields a 5% increase in real GDP.

Now, consider a different scenario, which we call Scenario 2 in the table. When the new treatment is first discovered, it costs an additional $50,000, just equal to the value consumers put on it. There is no increase in consumer surplus or real GDP, regardless of the quality adjustment method used. But in the year after the original discovery (year 3), researchers figure out how to make the new treatment more cheaply, and the cost and price both fall by $10,000. In Scenario 2, the economy in year 3 has a treatment worth $50,000 that costs $40,000 to produce, the same situation as in Scenario 1, but taking a year longer to get there. There is no question that the fall in treatment price from $50,000 to $40,000 from year 2 to year 3 should and would be counted as a price decline, because the quality of the treatment is the same in both years. Taking account of this price decline shows that *using both the utility method and the cost method* real GDP has increased 5%. It seems obvious that the change in real GDP should be the same in both Scenario 1 and Scenario 2 (since consumption is exactly the same by year 3), thus showing that the cost-based method isn't a reliable measure of GDP change when the innovation is a new good.

Finally, imagine that, instead of pricing the drug at cost, the company was able to price it at the consumer value, or $50,000. This is Scenario 3 in the table. Using the cost-method (where cost plus markup is subtracted) and recognizing that profits and nominal income will rise when prices exceed costs, we find that real GDP increases 5%, the same as with the utility-based method. This also shows that the consumer surplus from an innovation is equal to the utility value of it less its production costs. Whether the good is priced at cost or at value should only determine the split between consumers and producers, not the overall effect on GDP.[39]

Summary

The use of chained indexes means that, for goods and services with constant quality, it does not matter much whether you use an output price or a cost-of-living price, and, in any case, the current methodology is a very good approximation of both. When the set of characteristics of goods is fixed, but goods and services improve over time because they contain more of certain characteristics, there is similarly no difference in the quality-adjusted price using either the cost method or the utility method. But when a new characteristic is introduced, either as a quality adjustment to an existing good, or as an entirely new good, prevailing prices in the year before introduction will not yield the correct result for GDP, and a utility-based approach is necessary.

. . .

39. This conclusion is ignoring the inefficiencies associated with monopoly power, which would limit the consumption of the new good and thus lead to a smaller increase in consumer surplus and real GDP.

Table 3: When Cost-based and Utility-based Quality Adjustments Differ, Use Utility-based Adjustment

						Cost-based Quality Adjustment			Utility-based Quality Adjustment		
	Treatment Price	Years of Life	Medical spending	Other spending	Nominal Income	Quality Adjusted Treatment Price	Inflation[a]	Real GDP Growth[a]	Quality Adjusted Treatment Price	Inflation[a]	Real GDP Growth[a]
Scenario 1: Innovation valued at $50,000. Cost of innovation $40,000. Change in Price = Change in Cost.											
Year 1	$30,000	1	$30,000	$170,000	$200,000	$30,000			$30,000		
Year 2	$70,000	2	$70,000	$130,000	$200,000	$30,000	0%	0%	$20,000	-5%	5%
Scenario 2: Innovation valued at $50,000. Cost of innovation $50,000 at first, then falls to $40,000. Change in Price = Change in Cost.											
Year 1	$30,000	1	$30,000	$170,000	$200,000	$30,000			$30,000		
Year 2	$80,000	2	$80,000	$120,000	$200,000	$40,000	0%	0%	$30,000	0%	0%
Year 3	$70,000	2	$70,000	$130,000	$200,000	$35,000	-5%	5%	$26,250	-5%	5%
Scenario 3: Innovation valued at $50,000. Cost of innovation $40,000. Change in Price = Change in Value. Margin between Price and Cost Increases Profits, Increasing National Income.											
Year 1	$30,000	1	$30,000	$170,000	$200,000	$30,000			$30,000		
Year 2	$80,000	2	$80,000	$130,000	$210,000	$30,000	0%	5%	$30,000	0%	5%

a Using Laspeyres price index/Paasche real GDP growth for convenience. Results similar using Fisher ideal.

5. How does BLS actually adjust for quality changes?

The data requirements to actually quality adjust prices in a completely theoretically-appropriate way are substantial and unlikely to be available to BLS, particularly in real time.[40] BLS does not really tackle the issue of new goods or quality improvements that represent new characteristics. Instead, what BLS tries to ensure is that price increases that are directly attributable to quality increases are not counted as inflation. In the cereal example above, say, BLS would want to make sure that the price increase stemming from the introduction of a larger cereal box is stripped out from the observed market price of the larger cereal box.[41] Although BLS makes no attempt to directly measure the utility value from new goods and quality improvements that introduce new characteristics, we show that their methods will sometimes capture the consumer surplus from these innovations indirectly.

BLS collect prices on a fixed set of goods and services sold at a fixed set of outlets, with new outlets and new goods and services rotated in regularly.[42] They use a number of methods to derive inflation indexes from these data.

(1) The matched model: the basic method used by BLS and by far the most common way goods and services are priced. The matched model collects the prices of a specific model (made by the same manufacturer and sold in the same store at the same location) from one month to the next.

The matched model is only feasible when the product in the BLS sample is sold in two adjacent periods.[43] When a match is not available (because a product is not sold anymore or is unavailable), BLS uses other methods of trying to impute the price it would have sold at:

(2) Hedonic regression analysis: a method that specifies that the price of a good is a function of its characteristics.

(3) The direct cost method: a method that adjusts a new good for the cost of the product improvement. That is, when a new model of a good is better than the old model, they subtract the cost (plus markup) of the improvement from the price.

. . .

40. Putting a number on the overall bias in consumer prices is very difficult, particularly when the degree of bias has changed over time given changes in data-construction practices and in the economy. Groshen, Moyer, Aizcorbe, Bradley, and Friedman (2017) summarizes some of the literature that has attempted to do so. For example, Lebow and Rudd (2003) estimated that CPI growth was biased upward by 0.87 percentage points per year as of 2001. Reinsdorf and Schreyer (2018) estimate that adjusting for new goods and quality changes related to the digital economy would reduce growth in the PCE deflator in 2015 by a maximum of 0.58 percentage point per year.

41. As noted above, this discussion abstracts from any disutility consumers might get from having to purchase in bulk.

42. For the CPI, the BLS selects new item and outlet samples, on a rotating basis, for approximately 25 percent of item strata every year (Bureau of Labor Statistics, 2018). For the PPI, the BLS adjusts weights for commodity groupings every 5 years based on the previous economic census. For instance, since January 2012 PPI commodity weights have been derived from the total value of commodities reported in the 2007 economic census (Bureau of Labor Statistics, 2011).

43. Groshen, Moyer, Aizcorbe, Bradley, and Friedman (2017) report that, during the 12 months ending November 2014, CPI data collectors could price the same time 73 percent of the time.

(4) Imputation: when none of the above methods is possible, BLS will impute a price based on price changes of other similar products.

How well do these models do measuring quality relative to the theoretical ideal described above?

5.1 Matched model

The matched model holds quality constant by only pricing the exact same good or service over time. This price change is then used to deflate all the expenditures in a particular category. Because the exact same item is priced over time, quality improvements cannot be mislabeled as price increases.

As a simple example of how the model is used, consider the example of the introduction of the iPhone 8. Assume that iPhones are in the sample of goods and services that BLS tracks. There would be no match for the iPhone 8 in the period of its introduction (though there would be in subsequent periods), and, if the price of the iPhone 7 did not change in this period, then the matched model would imply that there had been no changes in iPhone prices. All increases in spending on IPhones (both 7s and 8s) would be counted as increases in real cellphone spending.[44] For example, if the price of the iPhone 7 was $600 in both years, and the iPhone 8 was introduced at $650, none of the increase in the iPhone price from the 7 to the 8 would be counted as inflation.

As this example shows, the matched model can only handle quality changes if the older model continues to be sold even after the newer model is introduced. If the iPhone 7 had been discontinued when the iPhone 8 was introduced, BLS would not be able to collect any price information for it, and the matched model could not be used.

Assuming that an existing model continues to be sold alongside a new model, so that the matched model is feasible, how well does the matched model do at accounting for quality improvements? That is, using the matched model, will a technological innovation increase real GDP by the difference between the average value of the innovation to consumers and the cost of producing it? Or, in other words, will the change in prices from the matched model be a good proxy for the change in prices for all the models that are sold (where the change in the price for a new model is the change from its reservation price)?

The answer is that it depends on how the innovation is priced and how the prices of the continuing goods that BLS tracks respond. In the case of a quality improvement that is just a rebundling of available characteristics, like the cereal example above, the answer is likely to be yes. In the specific cereal example, where there was only one characteristic, the change in the price of the smaller box of cereal would be a perfect proxy for the change in the price of the larger box, because both were determined by the underlying price for cereal. More generally, so long as the characteristics of the goods that are sold continuously are close to those included in the new model, the price changes should be similar and the matched model will do a reasonable job.[45]

. . .

44. It is important to distinguish between increases in real spending on cellphones and increases in real GDP. If spending on iPhones increases, and nominal income has not changed, then spending on other goods and services must be declining. If the price of iPhones doesn't change when the iPhone 8 is introduced, and if profits from the IPhone don't increase, then the introduction if the iPhone 8 will not have an effect on measured real GDP. If, on the other hand, Apple's profits increase from the introduction of the iPhone 8, then both nominal and real GDP increases.
45. When the prices of different characteristics change at different rates, the actual change in the price of any given model will depend on its exact mix of characteristics. Some upward bias is possible here if new models systematically provide more of the characteristics that have experienced price declines and less of the characteristics whose prices have increased.

Now consider the case of an innovation that introduces a truly new characteristic. The gains from the innovation—the difference between the average value of the innovation to consumers and the average cost of producing it—will be split between producers and consumers. The producers share—the increase in profits—will be counted in nominal and real GDP.[46] In the extreme case of a producer who is able to extract all the consumer surplus from an innovation, all of the benefit of the innovation would end up in profits, and would therefore be in GDP. In this case, there would be no decline in quality-adjusted price from the innovation—because the change in the value would be equal to the change in the price—and the matched model would get the change in quality-adjusted prices and the change in real GDP correct.[47]

However, if there is some consumer surplus from the new innovation, then the degree to which it is captured by the matched model will depend on the price response to the innovation by its competitors. Let's use the iPhone example again. Assume the iPhone 8 is worth $150 more to consumers, on average, than the iPhone 7, but is priced at only $50 more. Consumers get $100 of consumer surplus by buying the iPhone 8. If the iPhone 8 and the iPhone 7 were perfect substitutes, Apple could only continue to sell IPhone 7s if it lowered their price by $100. If the iPhone 7 price fell by $100 when the iPhone 8 was introduced, then the matched model would register this as a decline in the price of cellphones and apply it to all iPhone purchases. iPhone prices would fall, and, even if nominal spending on IPhones is unchanged, real cellphone spending would increase.[48]

In general, one would not expect prices of competing goods to adjust fully. The degree of price adjustment expected would depend on how much a new model substituted for an existing model, the cost structure of the competitor, and the competitors' market power.[49] Consider the introduction of Uber and Lyft, which many believe provide a superior service to taxis at a lower price. It may be that taxis continue to operate (at reduced market shares) and taxi prices do not decline because the costs of providing tax services have not changed or because prices are regulated. In this case, monitoring the price of taxis (the matched service) will not provide a measure of the consumer surplus derived from the introduction of Uber and Lyft, and that consumer surplus will be missing from GDP.

Or consider a situation where it is not the good that changes, but the retailer. For example, consider the introduction of a low-priced supplier into a market—perhaps Walmart or an online retailer—and assume that these suppliers are able to provide lower prices because they have figured out how to lower the costs of operation. The matched model considers the products sold at Walmart and Amazon as different goods than the products sold at Safeway, so the fact that the introduction of Walmart might have

. . .

46. On the income side, profits go directly into the accounts. On the expenditure side, the profits will either be saved and end up as investment purchases or will be used to finance consumption, both of which will also be captured.

47. Reinsdorf and Schreyer (2018) point out that the typical pattern for high-tech goods is for manufacturers to introduce them at high prices and then to let prices fall over time (this may be a form of price discrimination). In this case, getting the new goods into the BLS sample more quickly could help improve GDP measurement, because the price declines from the initially high introductory price would be picked up and more of the consumer surplus from new goods would be captured in GDP.

48. The increase in consumer surplus from the lower price on iPhone 7s would be offset by lower profits on their sale. On net, only the consumer surplus from the iPhone 8s would be added to real GDP (yielding the correct answer).

49. Feenstra (1994) shows that the degree of overstatement of prices from only examining prices of continuing goods depends on both the degree of substitutability of new goods for old goods and the shares of spending allocated to new goods once they are introduced. The introduction of a new good If competitors lower prices, then the shares of spending allocated to new models will be lower.

lowered the average prices paid for a given product would not be counted as a price decline.[50] To the extent the introduction of Walmart into an area forces Safeway to lower its prices, the matched model will capture some of the consumer surplus.[51] But if prices at Safeway do not fall by much, and if many consumers switch from Safeway to Walmart as a result—then the matched model will have understated the change consumer surplus and real GDP associated with the introduction of the lower-priced supplier.[52]

This is the phenomenon known as *outlet substitution bias*. This bias is not limited to the introduction of new retail outlets. It can apply to services—for example, the introduction of a new discount health insurer in a market could result in lower prices for health care that would never be captured as a price decrease if existing insurers don't also lower their prices. It might even apply more broadly. For example, if people move from one city to another to take advantage of a lower cost of living, consumer price indexes will never reflect the decline in the cost of living.

These are all examples where the matched model overstates inflation and understates real GDP. However, the opposite might also occur. Prices of old models that are on their way to being pushed out by newer models might sell for fire sale prices, as retailers try to get rid of their inventory. The matched model might overstate price declines in this case. Similarly, companies might use the introduction of a new model as an opportunity to raise prices, but might leave the price of their old model unchanged. Because the old model price does not change, the matched model will miss this inflation, and will call all additional spending on the new model an increase in real spending.

As we noted above, the matched model is BLS's workhorse, and, in most cases, is used whenever a match is available. When a match is no longer available, either because the product is simply out of stock or has been discontinued, BLS employs a few alternative methods.

5.2 Direct cost method

When a direct match is not available, BLS will sometimes link a new product to an old product by adjusting the price of the new product for the cost of the improvement (plus the producer's markup).[53] For

. . .

50. Because Walmart and Safeway coexist, one could argue that the amenities provided by Walmart must be inferior to those provided by Safeway, otherwise Walmart would drive out Safeway. But the fact that Walmart has gained market share suggests that the introduction of Walmart into a market does provide consumer surplus. Interestingly, BLS decided in 1995 to treat generic drugs (sold by a different manufacturer than brand name drugs) as the same product as brand name drugs, even though the brand name drugs continue to exist (Bureau of Labor Statistics, 2018). The rapid expansion of online retailers may lead to similar problems to the Walmart-Safeway example above. A recent paper by Goolsbee and Klenow (2018) found that inflation for goods sold online is 1.3 percentage points per year lower than for the same categories of goods in the standard CPI.

51. And, as in the IPhone example, if Safeway lowers its prices without managing to lower its costs, the price savings to consumers at Safeway will be offset in GDP by lower Safeway profits.

52. Byrne, Oliner, and Sichel (2017) point out how changes in the pricing behavior of chip manufacturers caused the matched model to fail. They note that, in the past, when a more powerful chip was introduced to the market, the prices of older chips fell, leading the matched model to register large declines in computer chip prices. But the pricing strategy changed, such that listed prices of existing chips no longer declined so much, despite continuing improvements in chips, leading the matched model to miss the continuing productivity in that sector.

53. Although the direct cost method uses "cost" terminology, BLS makes clear that the cost includes the producer's markup. See footnote 2 of BLS (2011) "Hedonic Models in the Producer Price Index" https://www.bls.gov/ppi/ppicomqa.htm.

example, assume that Apple withdrew the iPhone 7 from the market when the iPhone 8 was introduced. If using the direct cost method, the BLS would have asked Apple how much of the increase in the price of the iPhone 8 relative to the iPhone 7 was due to the cost (plus markup) of the improvements, and subtracted that from the price. If Apple says the entire increase in price was due to the improvements, then the quality-adjusted price of the iPhone 8 would be equal to the price of the iPhone 7.

As we saw above, whether this is the correct quality adjustment or not depends on Apple's pricing behavior.[54] If Apple is able to extract the entire increase in consumer surplus from the iPhone 8, then a quality-adjustment that shows no decline in prices would be correct. But if consumers do get consumer surplus, so that the correct quality-adjusted price has declined, the direct cost method will not get this right.

In general, with downward sloping demand functions, we expect the introduction of truly new goods to be associated with consumer surplus, which the direct cost method would not capture.

5.3 Hedonic regression

As noted above, hedonic regression is a useful way of imputing prices for products that are simply bundles of existing characteristics.[55] A well-specified hedonic regression will do a good job of imputing a price that a good would have been sold at had it been available. So, when an existing good disappears, a hedonic regression can be used to impute the price it would have sold at. For example, if the iPhone 7 had disappeared, but other cellphones on the market had very similar characteristics, the price of the iPhone 7 could be imputed. If other older models of phones experienced price drops when the iPhone 8 was introduced, then that could be picked up with the use of hedonics.

The use of hedonics cannot overcome the basic problem identified with the matched model, however. It can only capture the consumer surplus from the introduction of truly new quality improvements (i.e., new characteristics) if the prices of competitors decline. Nothing in the market data on selling prices and characteristics of goods can uncover the price a new good would have sold at in the period before it was introduced—the consumer's reservation price—because there are no observations with that characteristic.

5.4 Imputations

The matched model and hedonics may capture part of the benefits of new innovations, but are likely to leave quite a bit unmeasured. The problem is magnified when a new innovation is good enough that it pushes potential competitors out of business. Then, there will be no close products to price using either the matched model or hedonic regressions. Similarly, a truly new product—like the smartphone—may not really have any competitors.

In both these cases, the price changes used by BLS to deflate will consist of price changes for other items that are being followed. Because it is systematically dropping items that likely would have experienced price declines had they stayed in the basket, the inflation measure will be biased upward.

Imagine in the example of taxis and Uber, for example, that Uber was such a good replacement for taxis that it pushed them completely out of the business. Then the matched model for taxi services could

. . .

54. Note that, regardless of what is assumed about the quality-adjusted price of the IPhone, changes in actual profits will be captured in nominal GDP.

55. See Triplett (2004) for a thorough exploration of the theory and practice of hedonic regressions.

not be used, and hedonic regressions would have no observation on taxi-like services in the year after the introduction of Uber, so could not be used to impute a taxi price. In this example, BLS would likely just use the price changes for other types of services in "intracity mass transit"—like the price change for limousines—as a proxy for the inflation in the category as a whole.

Similarly, imagine if big box stores push mom and pops out of a market, because these stores could not lower their costs enough to be competitive. With the retail outlet missing, BLS would simply drop that outlet from its sample. This upward bias to inflation from the selective exiting of certain goods and services from BLS sample (and from the market) imparts an upward bias to inflation and therefore a downward bias to GDP and productivity. Aghion et al (2017) have labeled this missing productivity from "creative destruction."

5.5 Other quality changes that are not captured

Consumer Services. While the matched model has much appeal for goods, it is harder to think about matching a specific model for services. BLS can price the same medical treatment over time, the cost of having a will written, or the price of man's haircut, but unlike with goods where there are exact replicas based on model numbers, it is difficult to hold the quality constant over time in the case of services. Consider the benefits of requiring pilots or surgeons to use safety checklists, for example. Airline flights or surgeries after the implementation of these checklists would not be viewed as new models under the BLS approach, and thus the benefits of these improvements would be missed. Other costly innovations, including the use of better inputs like better equipment in the plane or better sutures for surgery, might also be missed.

Government and non-profits. Because the services produced by governments and non-profits are not sold in the market place, there is no sense in which the prices over time can be compared. Instead, BLS uses changes in the costs of the inputs (labor, material, supplies, etc.) to deflate changes in nominal spending. For example, when teacher wages increase, the deflator used to deflate spending on public education increases. This method assumes that there is no productivity growth in the provision of public services.[56] But if teacher wages are increasing because teachers are more productive (because of an increase in their education, or because of an improvement in pedagogic methods, for example), a quality adjusted price deflator would not show an increase in prices, but the BLS would nonetheless assume one.

6. Summary and conclusion

Our goal in this paper was to explore the basic economics surrounding the measurement of GDP. We have focused, in particular, on the question of whether GDP should be viewed as a measure of economic well-being. Our ultimate goal is to understand the extent to which productivity growth captures changes in living standards, but we have limited ourselves in this paper to discussing GDP, the numerator in productivity.

We separated the discussion into different parts. We first covered the GDP concept, focusing on how well the scope of GDP *as conceptualized* aligns with aggregate economic well-being. (We recognize, but do not address, the limitations for capturing well-being of *any* aggregate measure that ignores the distribution of resources.) We then considered how well nominal GDP compares with the nominal

. . .

56. Indeed, the deflators for the government sector are typically higher than those for the rest of the economy.

spending that one would expect the GDP concept to capture. Finally, we consider the deflators used to convert nominal GDP to real GDP, with a thorough exploration of the deflators that would be ideal in measuring economic well-being and how they compare with BEA's actual practices.

Our preliminary conclusions are as follows. First, from a theoretical perspective, one would not want to use real GDP per se as a comprehensive measure of aggregate economic well-being. There are some important conceptual differences between the two concepts. For example, as a production-based measure, GDP captures investment (spending that does not immediately translate into higher welfare) in addition to consumption. However, some would argue that investment should be included on the grounds that it eventually translates into consumption. Moreover, a broad measure of private and public consumption is highly correlated with GDP over long periods of time, suggesting that the inclusion of investment does not materially affect the value of GDP as a measure of aggregate economic well-being over long periods of time. At higher frequencies, the growth rate of GDP is more volatile than that of consumption. But, that feature of GDP is useful to government officials responsible for stabilizing the economy in the face of business cycle fluctuations, as they are focused on the degree to which the economy is using all of its productive resources.

That said, we believe that one conceptual difference—the exclusion of non-market activities that bear on economic well-being—merits more attention, particularly given the potential for changes in the importance of such activities over time to change the degree to which changes in GDP capture changes in well-being.

We also flag a number of issues that warrant more attention when it comes to translating GDP as conceptualized into GDP as measured. To start, the national accounts may mismeasure the nominal GDP arising from internet-provided "free goods" and the operation of multinationals corporations. In addition, the deflators used to separate GDP into nominal GDP and real GDP may produce biased measures of inflation in some instances which, in turn, lead to biases in real GDP as measured.

Our analysis showed that there are many cases in which current practices yield measures of prices that serve well as deflators, dispelling some common misconceptions along the way. We show that, by definition, cost-of-living deflators yield changes in real GDP that measure changes in consumer surplus. Furthermore, except in some specific cases, there is little difference in practice between cost-of-living deflators and output-price indexes. In particular, for goods and services that do not change in quality over time, current deflator methods will yield a real GDP measure that captures the change in consumer surplus fairly well.

The important exceptions to these conclusions are for price measures for goods and services that change in quality and for new goods and services. In these cases, current methods may not capture consumer surplus well. We believe that efforts to improve the price measures so that they better capture consumer surplus would be very useful.

REFERENCES

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li. 2017. "Missing Growth from Creative Destruction," Federal Reserve Bank of San Francisco Working Paper 2017-04. https://doi.org/10.24148/wp2017-04

Ahmad, Nadim and Paul Schreyer. 2016. "Are GDP and Productivity Measures Up to the Challenges of a Digital Economy?" *International Productivity Monitor* Number 30 (Spring).

Ahmad, Nadim and Jennifer Ribarsky. 2017. "Issue paper on a proposed framework for a satellite account for measuring the digital economy." OECD Committee on Statistics and Statistical Policy, 14 November, 2017. https://scholar.harvard.edu/files/jorgenson/files/p1b_std-cssp-wpna201710.pdf?m=1527941424

Aizcorbe, Ana. 2014. A Practical Guide to Price Index and Hedonic Techniques. Oxford University Press, Oxford Scholarship Online, August. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198702429.001.0001/acprof-9780198702429

Aldy, Joseph E and W. Kip Viscusi. 2007. "Age Differences in the Value of Statistical Life: Revealed Preference Evidence," Resources for the Future Discussion Paper RFF DP 07-05, April. http://www.rff.org/files/sharepoint/WorkImages/Download/RFF-DP-07-05.pdf.

Bernanke, Ben S. and Peter Olson. 2017. "Are Americans Better Off than They Were a Decade or Two Ago?" in *Brookings Big Ideas for America* edited by Michael E. O'Hanlon. Washington, DC: Brookings Institution.

Bond, Timothy N. and Kevin Lang. 2018. "The Sad Truth about Happiness Scales: Empirical Results." NBER Working Paper 24853 (July). http://www.nber.org/papers/w24853.pdf.

Boyd, James. 2007. "Nonmarket Benefits of Nature: What Should Be Counted in Green GDP?" *Ecological Economics* Vol. 61, Issue 4: 716-723.

Bridgman, Benjamin, Andrew Dugan, Mikhael Lal, Matthew Osborne, and Shaunda Villones. 2012. "Accounting for Household Production in the National Accounts, 1965-2010" *Survey of Current Business* (May).

Bridgman, Benjamin. 2016. "Accounting for Household Production in the National Accounts: An Update, 1965–2014" *Survey of Current Business* (February).

Bruner, Jennifer., Dylan G. Rassier, Kim J. Ruhl. 2018. "Multinational Profit Shifting and Measures throughout Economic Accounts." 2018 Conference on Research in Income and Wealth.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. "Does the United States Have a Productivity Slowdown or a Measurement Problem?" *Brookings Papers on Economic Activity* (Spring).

Byrne, David M., and Carol Corrado. 2017. "Accounting for innovation in consumer digital services." 5th IMF Statistical Forum.

Brynjolfsson, Erik, W. Erwin Diewert, Felix Eggers, Kevin J. Fox and Avinash Gannamaneni. 2018. "The Digital Economy, GDP and Consumer Welfare: Theory and Evidence." Available at: https://www.escoe.ac.uk/wp-content/uploads/2018/06/EM2018-Brynjolfsson-et-al.pdf.

Brynjolfsson, Erik, Felix Eggers, and Avinash Gannamaneni. 2017. "Using Massive Online Choice Experiments to Measure Changes in Well-Being. Massachusetts Institute of Technology.

Bureau of Economic Analysis. 2015. "Measuring the Economy: A Primer on GDP and the National Income and Product Accounts." Available at: https://www.bea.gov/national/pdf/nipa_primer.pdf.

Bureau of Foreign and Domestic Commerce and Simon Kuznets. 1934. *National Income, 1929-1932*. Washington: U.S. Government Printing Office.

Bureau of Labor Statistics. 2018. "The Consumer Price Index." *BLS Handbook of Methods*, Chapter 17. https://www.bls.gov/opub/hom/pdf/cpi-20180214.pdf

Bureau of Labor Statistics. 2011. "Producer Prices." *BLS Handbook of Methods*, Chapter 14. https://www.bls.gov/opub/hom/pdf/ppi-20111028.pdf

Chen, Wen, Reitze Gouma, Bart Los, Marcel P. Timmer. 2017. "Measuring the Returns to Intangibles: a Global Value Change Approach." World Intellectual Property Organization Economic Research Working Paper No. 36.

Constanza, Robert, Maureen Hart, Stephen Posner, and John Talberth. 2009. "Beyond GDP: The Need for New Measures of Progress." The Pardee Papers, No. 4.

Corrado, Carol, Charles Hulten, Daniel Sichel. 2009. "Intangible Capital and U.S. Economic Growth." *The Review of Income and Wealth*. Volume 55, Issue 3.

Corrado, Carol, Kevin Fox, Peter Goodridge, Jonathan Haskel, Cecilia Jona-Lasinio, Dan Sichel, and Stian Westlake. 2017. "Improving GDP: Demolishing, Repointing, or Extending?" Indigo Prize Winning Entry.

Coyle, Diane. 2014. *GDP: A Brief but Affectionate History* Princeton, NJ: Princeton University Press.

Coyle, Diane and Benjamin Mitra-Kahn. 2017. "Making the Future Count." Indigo Prize Winning Entry.

Cutler, David M. and Elizabeth Richardson. 1998. "The Value of Health: 1970-1990." American Economic Review, 88(2): 97-100. http://www.jstor.org/stable/116900

Cutler, David M., Mark B. McClellan, Joseph P. Newhouse, and Dahlia K. Remler. 2001. "Pricing Heart Attack Treatments." In Medical Care Output and Productivity, edited by David M. Cutler and Ernst R. Berndt, 305-362. Studies in Income and Wealth. Chicago, IL: University of Chicago Press. http://www.nber.org/chapters/c7634

Dauda, Seidu, Abe Dunn, and Anne Hall. 2017. "Are Medical Care Prices Still Declining? A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care."

Diewert, Erwin. 1976. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4:114-145.

Feenstra, Robert C. 1994. "New Product Varieties and the Measurement of International Prices", *American Economic Review*, March.

Feldstein, Martin. 2017. "Underestimating the Real Growth of GDP, Personal Income, and Productivity" *Journal of Economic Perspectives* 31 (Spring): 145-64.

Fernald, John, Robert Hall, James Stock, and Mark Watson. 2017. "The Disappointing Recovery of Output after 2009" *Brookings Papers on Economic Activity* (Spring).

Griffith, Rachel, Ephraim Leibtag, Andrew Leicester and Aviv Nevo. 2008. "Consumer Behavior and the Consumer Price Index (Preliminary)". http://www.indexmeasures.ca/dc2008/papers/Measurement%20conference.pdf

Groshen, Erica, Brian Moyer, Ana M. Aizcorbe, Ralph Bradley, and David M. Friedman. 2017. "How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches." *Journal of Economic Perspectives* 31(2) (Spring).

Guvenon, Fatih, Raymond J. Mataloni, Jr., Dylan G. Rassier, and Kim J. Ruhl. 2017. "Offshore Profit Shifting and Domestic Productivity Measurement." National Bureau of Economic Research Working Paper 23324 (April).

Harris, Benjamin H. and Adam Looney. 2018. "The Tax Cuts and Jobs Act: A Missed Opportunity to Establish a Sustainable Tax Code, Tax Policy Center (May 24). https://www.brookings.edu/wp-content/uploads/2018/05/es_20180524_harris-looney_taxreform.pdf

Hatzius, Jan. 2017. "Update on Productivity Measurement" *Goldman Sachs U.S. Daily* (August 15).

Hines Jr., James R. 2005. "Do Tax Havens Flourish?" In Tax Policy and the Economy, Volume 19, edited by James M. Poterba. Cambridge, MA. MIT Press: 65-99.

Ip, Greg. 2017. "This One Weird Tax Trick Could Raise GDP, Shrink the Trade Deficit." *Wall Street Journal* (December 18).

Jenniges, Derrick., Raymond J. Mataloni, Jr., Sarah Strutzman, Yiran Xin. 2018. "Strategic Movement of Intellectual Property within U.S. Multinational Enterprise." 2018 Conference on Research in Income and Wealth.

Jones, Charles I. and Peter J. Klenow. 2016. "Beyond GDP? Welfare across Countries and Time" *American Economic Review* 106(9): 2426-57.

Jorgenson, Dale W. 2010. "Designing a New Architecture for the U.S. National Accounts." *The Annals of the American Academy of Political and Social Science* 631, no. 1 (2010): 63-74.

Jorgenson, Dale. Forthcoming. "Production and Welfare: Progress in Economic Measurement." *Journal of Economic Literature*.

Kassenboehmer, Sonja C., Christoph M. Schmidt. 2011. "Beyond GDP and Back: What is the Value-Added by Additional Components of Welfare Measurement?" IZA Institute of Labor Economics Discussion Paper, No. 5453.

Krueger, Alan. 2008. "Comment on 'Economic Growth and Well-Being: Reassessing the Easterlin Paradox' by Stevenson and Wolfers." *Brookings Papers on Economic Activity*, Spring.

Lakdawalla, Darius, Jason Shafrin, Claudio Lucarelli, Sean Nicholson, Zeba M. Khan, and Tomas J. Philipson. 2015. "Quality-Adjusted Cost Of Care: A Meaningful Way To Measure Growth In Innovation Cost Versus The Value Of Health Gains." Health Affairs, 34(4): 555-61. DOI: 10.1377/hlthaff.2014.0639

Landefeld, Steven J. 2000. "GDP: One of the Great Inventions of the 20th Century," *Survey of Current Business* (January).

Lebow, David, and Jeremy B. Rudd. 2003. "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature* 41(1): 159–201.

Lipsey, Robert E. 2006. "Measuring International Trade in Services." NBER Working Paper 12271. http://www.nber.org/papers/w12271.pdf.

Moris, Francisco, John Jankowski, Mark Boroush, Marissa Crawford, and Jennifer Lee. 2015. "R&D Recognized as Investment in U.S. GDP Statistics: GDP Increase Slightly Lowers R&D-to-GDP Ratio." InfoBrief NCSES, March 30, 2015, NSF 15-315. https://www.nsf.gov/statistics/2015/nsf15315/

Moulton, Brent. 2018. "The Measurement of Output, Prices, and Productivity: What's Changed Since the Boskin Commission." Hutchins Center on Fiscal and Monetary Policy Working Paper. July. https://www.brookings.edu/wp-content/uploads/2018/07/Moulton-report-v2.pdf

Moulton, Brent, van de Ven, Peter. 2018. "Addressing Changes of Globalization in National Accounts." 2018 Conference on Research in Income and Wealth.

Nakamura, Leonard, Jon Samuels, and Rachel Soloveichik. 2017. "Measuring the 'Free' Digital Economy with the GDP and Productivity Accounts." Federal Reserve Bank of Philadelphia Working Paper No. 17-37.

National Research Council Panel to Advance a Research Program on the Design of National Health Accounts. 2010. Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement. Washington DC.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2016. "Distributional National Accounts: Methods and Estimates for the United States. National Bureau of Economic Research Working Paper 22945.

Reinsdorf, Marshall and Paul Schreyer. 2018. "Measuring Consumer Inflation in a Digital Economy." World KLEMS 2018. https://scholar.harvard.edu/files/jorgenson/files/p1a_reinsdorf_schreyer_r2.2ps.pdf?m=1527941394

Remes, Jaana, Jan Mischke and Mekala Krishnan. 2018. "Solving the productivity puzzle: The role of demand and digital." McKinsey Global Institute. Presented at World KLEMS 2018. https://scholar.harvard.edu/files/jorgenson/files/p1d_remes_khrisnan_mischke_klems_productivity_paper_draft.pdf?m=1527941461

Romley, John A., Dana P. Goldman, and Neeraj Sood. 2015. "US Hospitals Experienced Substantial Productivity Growth During 2002–11." Health Affairs, 34(3): 10-1377. DOI: 10.1377/hlthaff.2014.0587

Sack, Daniel W., Betsey Stevenson and Justin Wolfers. 2012. "The New Stylized Facts About Income and Subjective Well-Being." *Emotion*, 12(6): 1181-1187. http://users.nber.org/~jwolfers/papers/NewStylizedFacts(Emotion).pdf

Sheiner, Louise and Anna Malinovskaya. 2016. "Measuring Productivity in Health Care: An Analysis of the Literature." Hutchins Center on Fiscal and Monetary Policy Working Paper. July. https://www.brookings.edu/wp-content/uploads/2016/08/hp-issue-brief_final.pdf

Stevenson, Betsey and Justin Wolfers. 2008. "Economic Growth and Well-Being: Reassessing the Easterlin Paradox." *Brookings Papers on Economic Activity*, Spring.

Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi. 2009. *Report of the Commission on the Measurement of Economic Performance and Social Progress*.

Syverson, Chad. 2016. "Challenges to Mismeasurement Explanations for the U.S. Productivity Slowdown" National Bureau of Economic Research Working Paper 21974 (February).

Tørsløv, Thomas, Ludvig S. Wier, and Gabriel Zucman. 2018. "The Missing Profits of Nations." National Bureau of Economic Research Working Paper 24701.

Triplett, Jack E. 1982. "Concepts of Quality in Input and Output Price Measures: A Resolution of the User-Value Resource-Cost Debate" in *The U.S. National Income and Product Accounts: Selected Topics* edited by Murray F. Foss. Chicago, IL: University of Chicago Press.

Triplett, Jack E. 2004. "Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products" Organisation for Economic Co-operation and Development, STI Working Paper 2004/9.

United Nations Statistical Commission. 2009. "System of National Accounts 2008." https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf.

United States Committee on Homeland Security and Government Affairs. 2012. Offshore Profit Shifting and the U.S. Tax Code—Part 1 (Microsoft and Hewlett Packard, Hearings, September 20, 2012. 112th Congress, 2nd session. Washington: U.S. Government Publishing Office. https://www.gpo.gov/fdsys/pkg/CHRG-112shrg76071/pdf/CHRG-112shrg76071.pdf.

United States Committee on Homeland Security and Government Affairs. 2013. Offshore Profit Shifting and the U.S. Tax Code—Part 2 (Apple). Hearings, May 21, 2013. 113th Congress, 1st session. Washington: U.S. Government Publishing Office. https://www.gpo.gov/fdsys/pkg/CHRG-113shrg81657/pdf/CHRG-113shrg81657.pdf

Varian, Hal. 2016. "A Microeconomist Looks at Productivity: A View from the Valley." Slides Prepared for the Brookings Hutchins Center Conference on "Slow Growth in Productivity: Causes, Consequences, and Policies" (September).

Wesselink, Bart, Jan Bakkes, Aaron Best, Friedrich Hinterberger, and Patrick ten Brink. 2007. "Measurement Beyond GDP." Measuring Progress, True Wealth, and the Well-Being of Nations, International Conference, 19 & 20 November, Brussels.

Weitzman Martin L. 1976. "On the Welfare Significance of National Product in a Dynamic Economy." *Quarterly Journal of Economics* (February).

Weitzman, Martin L. 2003. *Income, Wealth, and the Maximum Principle*. Cambridge, MA: Harvard University Press.

Willig, Robert. 1976. "Consumer Surplus Without Apology." *The American Economic Review* Vol. 66, No. 4.

Wolfers, Justin. 2003. "Is Business Cycle Volatility Costly? Evidence from Surveys of Subjective Wellbeing." *International Finance*, 6(1).http://users.nber.org/~jwolfers/papers/Happiness.pdf

BOX 1. CHANGE IN REAL GDP AND CHANGE IN CONSUMER SURPLUS – A GRAPHICAL ANALYSIS

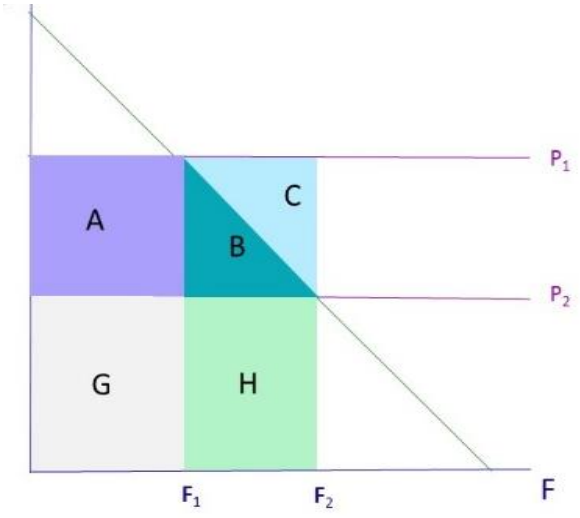
One of the conceptual difficulties with viewing real GDP growth as a measure of the change in welfare is that, to the extent economists think of GDP graphically, we think of it as a rectangle (P times Q) while we think of consumer surplus as having a triangle.

But, as we show in this simple graphical example, chained GDP, which accounts for substitution effects when prices fall, also gives rise to a triangle and the change in real GDP is approximately the change in consumer surplus.

Assume a simple economy with just two goods, F and D. The price of F falls, and nominal income is unchanged. The change in consumer surplus from that price change can be read off the demand curve for good F, as shown below. (Note that this is the entire consumer surplus from the price change, even if the demand for other goods changes as a result of the change in the price of F.) We discuss three cases.

Case 1: Only the demand for good F changes as a result of the price change.

Income is unchanged, so spending on good F is the same in period 1 and period 2, which means that $P_2F_2 = P_1F_1$. Thus, in the diagram below, H + G = A + G, or H = A. In other words, the savings from the price decline (A) is used to finance increased spending on F (H).



Change in consumer surplus is A + B.

Paasche change in GDP (period 2 prices) is $(F_2 - F_1)P_2$, or H.

Laspeyres change in GDP (period 1 prices) is $(F_2 - F_1)P_1$, or H+B+C.

Averaging Paasche and Laspeyres (approximately what chaining does) gives you a change in GDP = H + B.

Because H = A, change in consumer surplus (A+B) = change in real GDP (H+B).

Case 2: Demand for both goods changes as a result of the change in the price of F.

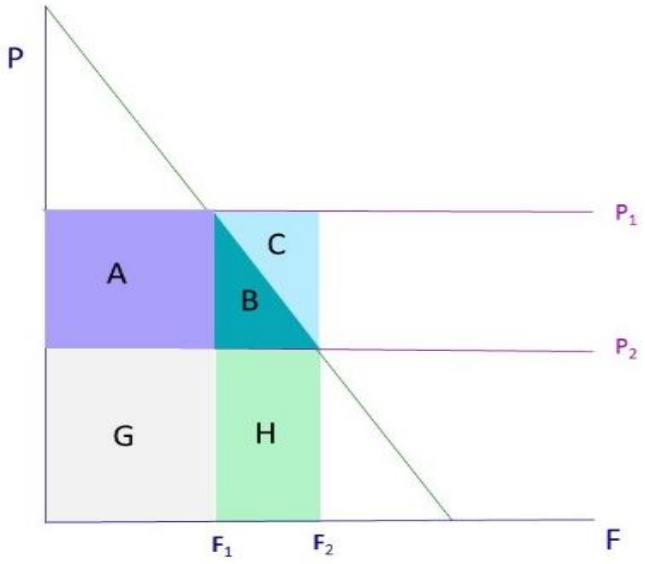
F increases from F_1 to F_2, and D increases from D_1 to D_2. Because income doesn't change between the two periods, neither do total expenditures: $P_1F_1 + D_1 = P_2F_2 + D_2$.

Rewriting, this yields $P_1F_1 = P_2F_2 + (D_2 - D_1)$. Subtracting P_2F_1 from both sides yields:

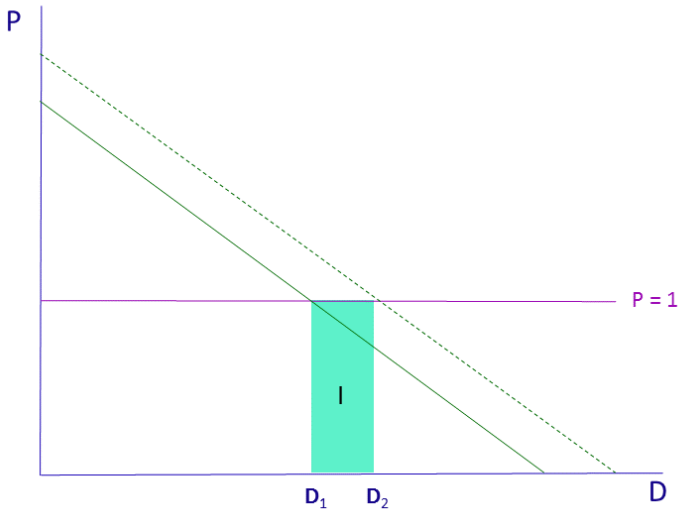
$$(P_1 - P_2)F_1 = P_2(F_2 - F_1) + (D_2 - D_1).$$

Thus, A = H + I. (Savings on F spent on more F and more D)

Market for F



Market for D



Change in consumer surplus: A + B

Change in real GDP

- Paasche: H + I
- Laspeyres: H + B + C + I

Average change in real GDP: H + I + B

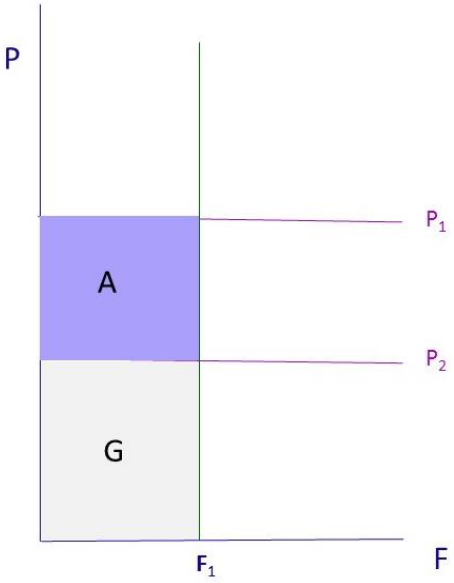
Because A = H + I, change in real GDP (H + I + B) = Change in Consumer Surplus (A + B).

Case 3: Only demand for D changes as a result of the change in the price of F.

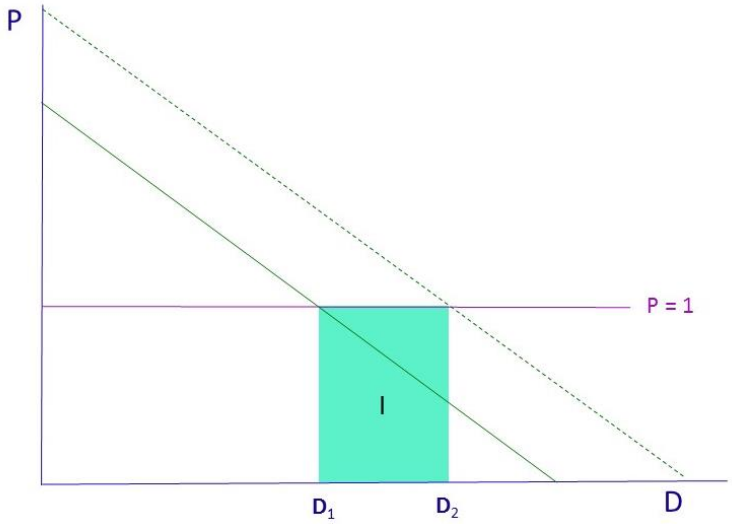
F is unchanged, and D increases from D_1 to D_2. Because income doesn't change between the two periods, neither do total expenditures: $P_1F_1 + D_1 = P_2F_1 + D_2$.

Rewriting, this yields, $(P_1 - P_2)F_1 = (D_2 - D_1)$.
Thus, A = I. (Savings on F spent on more D)

Market for F



Market for D



Change in consumer surplus: A

Change in Real GDP

- Paasche: I
- Laspeyres: I

Average change in real GDP: I

Because A = I, change in real GDP (I) = Change in Consumer Surplus (A).

BOX 2. HOW CAN CHANGES IN REAL GDP MEASURE CHANGES IN CONSUMER SURPLUS GIVEN THE PARADOX OF DIAMONDS AND WATER?

Some people use the intuition from the "paradox of diamonds and water"—which explains why diamonds are more expensive than water even though water is essential to life and hence obviously more valuable—to argue that welfare is much bigger than GDP because many goods (like water) are much more valuable to consumers than what they pay for them. Since nominal GDP is what people pay for the goods they buy, they argue, welfare must be greater than GDP. Another way of stating the argument is that GDP weighs the quantities of goods by prices, which, in equilibrium, represent the marginal utility people get from their consumption, but overall welfare should weigh the quantity of goods by the average utility, which is generally higher. According to this argument, you can choose to measure output, using GDP, or welfare, using consumer surplus, but one should not confuse the two.

This intuition is misleading. First, no particular number captures the *level* of welfare or even the level of real income. How well off are we? It is a concept that only applies in relative terms: Are we better off economically than we were last year, or are we better off than another country? These are the only types of questions we can answer. Similarly, real GDP (as opposed to nominal) is also only defined as a relative concept—that is, we choose a year in which we assume real GDP is equal to nominal GDP, and then create an index of real GDP based on that. So, it doesn't make sense to say that the level of welfare is higher than the level of real GDP.

Furthermore, there is no alternative measure of welfare based on consumer surplus, because the sum of consumer surpluses good by good is not a measure of overall welfare. Consumer surplus (defined as the triangle under a demand function) is an approximation of the **amount of money** one would have to give a consumer if the price of one good got so high that they chose not to buy it at all—or, equivalently, if the good did not exist. It by definition holds everything else constant and answers a well-specified and answerable question: how much money would you need in order to be as well off as you are with your current income but none of the good in question? For example, how much money would you need in a world without iPhones to be as well off as you are with your current income and the ability to buy an iPhone for $700? The increment to income to make you as well off as you are now is the consumer surplus you get from the ability to buy an iPhone for $700.

One cannot add consumer surpluses together good by good to get welfare. How much money would we have to give you if you could not buy anything? No amount of money would make you as well off you are now, since you could not have any goods.

To get more intuition, examine the following utility function and the consumer surplus for each good.

Let $U(C,H) = \sqrt{C} + \sqrt{H}$ and the budget constraint be: $Y = P_C C + P_H H$.

Maximizing U subject to the budget constraint gives you $C = \frac{P_H Y}{P_C(P_C+P_H)}$ $H = \frac{P_C Y}{P_H(P_C+P_H)}$.

Start off with income of $10 and prices of both goods $1, so C = 5, H = 5 and Utility = $2\sqrt{5}$. Spending on C is $5 and spending on H is $5.

Now assume you cannot buy C anymore. How much more money do we have to give you in order for you to be as happy as you were when you could buy C? Let's call the amount of H you would need H* and the income you would need Y*.
With C = 0, $H^* = 20$ and $Y^* = \$20$, and $U = \sqrt{0} + \sqrt{20} = 2\sqrt{5}$.

Thus, you need an additional $10 ($20-$10) to compensate you for losing C, so you might say the value of C is $10—equal to income itself and double the amount that you were spending on C before! You might say that, if C is worth $10 and H is worth $10 (by symmetry), then welfare must be $20 when GDP is $10, but that would be wrong. It would imply that U(5,5) = U(20,20), which is obviously untrue. All it implies is that U(5,5) = U(0,20) = U(20,0).

Thus, there is nothing inconsistent with the fact that the consumer surplus for one good can be extremely large and the fact that the change in real GDP measures the change in welfare, defined in monetary terms.

Hutchins Center
on Fiscal & Monetary Policy
at BROOKINGS

The mission of the Hutchins Center on Fiscal and Monetary Policy is to improve the quality and efficacy of fiscal and monetary policies and public understanding of them.

Questions about the research? Email communications@brookings.edu.
Be sure to include the title of this paper in your inquiry.

Gross domestic product

Introduction

Gross domestic product (GDP) is a key measure of a nation's economic development and growth. This chapter considers economic growth across the regions of the European Union Member States and candidate countries Croatia and the former Yugoslav Republic of Macedonia. It finds that the differences between Member States are quite large, but decreasing.

Economic activity is expressed in national currency, converted by purchasing power parities (PPPs), which take account of different price levels between Member States, allowing for a more accurate comparison. Thanks to PPPs, GDP is converted into an artificial common currency, called purchasing power standards (PPS). This makes it possible to compare purchasing power in countries that use different national currencies.

Finally, the chapter considers the level of economic dynamism in the regions of Member States and candidate countries, and finds that new Member States are continuing to catch up at a relatively strong rate.

Main statistical findings

Regional GDP per inhabitant in 2008

Map 7.1 shows per-inhabitant GDP (as a percentage of the EU-27 average of 25 100 PPS) for the European Union, Croatia, the former Yugoslav Republic of Macedonia and Turkey, which has, after a lengthy interruption, again provided data (for the reference years 2004–06) in line with the European system of accounts (ESA95) Data Transmission Programme.

The regions with the highest per-inhabitant GDP are in southern Germany, the south of the UK, northern Italy and Belgium, Luxembourg, the Netherlands, Austria, Ireland and Scandinavia. The regions around certain capitals, Madrid, Paris, Praha and Bratislava, also fall into this category. The weaker regions are concentrated in the southern, south-western and south-eastern periphery of the Union, in eastern Germany and the new Member States, Croatia, the former Yugoslav Republic of Macedonia and Turkey.

Detailed analysis of the data in this chapter does not cover Turkey, since the data available consists of a time series that only goes up to 2006, i.e. two reference years less than for other countries.

Within the EU, per-inhabitant GDP ranges from 28 % of the EU-27 average (6 500 PPS) in Severozapaden in Bulgaria to 343 % (85 800 PPS) in the capital region of Inner London in the UK.

The factor between the two ends of the distribution is therefore 13.2:1. Luxembourg at 280 % (70 000 PPS) and Brussels at 216 % (54 100 PPS) are in positions two and three, followed by Groningen (Netherlands) at 198 % (49 700 PPS), Hamburg at 188 % (47 100 PPS) and Praha at 173 % (43 200 PPS) in positions four, five and six. Praha (Czech Republic) thus remains the region with the highest per-inhabitant GDP in the new Member States; Bratislavský kraj (Slovakia) follows with 167 % (41 800 PPS) in ninth position among the 275 statistical areas (known as NUTS 2 regions of the countries examined here — 271 regions in the EU plus three regions in Croatia, and the former Yugoslav Republic of Macedonia). However, Praha and Bratislavský kraj must be regarded as exceptions as regards regions in the new Member States that joined in 2004. The next most prosperous regions in the new Member States are a long way behind: Bucureşti - Ilfov in Romania at 113 % (28 300 PPS) in position 74, Zahodna Slovenija (Slovenia) at 109 % (27 300 PPS) in position 87, Közép-Magyarország (Hungary) at 107 % (26 800 PPS) in position 96 and Cyprus at 97 % (24 400 PPS) in position 129.

With the exception of four other regions (Mazowieckie in Poland, Sjeverozapadna Hrvatska in Croatia, Malta and Vzhodna Slovenija in Slovenia), all the other regions of the new Member States, Croatia and the former Yugoslav Republic of Macedonia have a per-inhabitant GDP in PPS of less than 75 % of the EU-27 average.

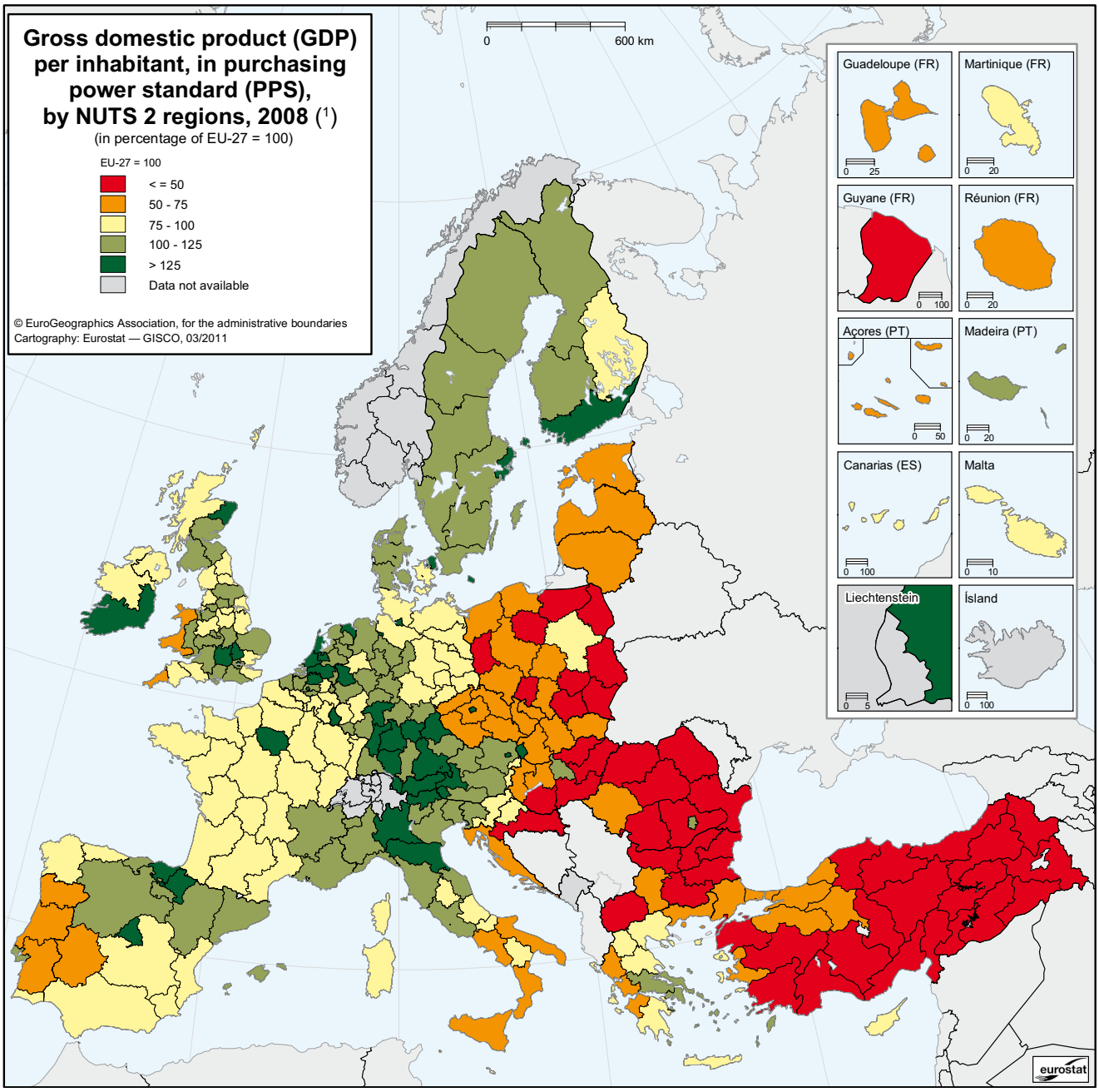
As a result, in 2008, GDP in 67 regions was less than 75 % of the EU-27 average. Some 24.4 % of the population of the EU, Croatia and the former Yugoslav Republic of Macedonia lives in these 67 regions. Only a quarter of these regions are in EU-15 countries, while three quarters are in new Member States, Croatia and the former Yugoslav Republic of Macedonia.

At the upper end of the spectrum, 40 regions have per-inhabitant GDP of more than 125 % of the EU-27 average; these regions are home to 19.4 % of the population. Regions with a per-inhabitant GDP of between 75 % and 125 % of the EU-27 average are home to 56 %, and thus a clear majority of the population of the 29 countries under consideration (EU-27, Croatia and the former Yugoslav Republic of Macedonia). Some 9.3 % of the population live in the 27 regions whose per-inhabitant GDP is less than 50 % of the EU-27 average. With the exception of the French overseas department of Guyane, all these regions are located in the new Member States, Croatia or the former Yugoslav Republic of Macedonia.

Major regional differences even within countries themselves

There are also substantial regional differences within countries themselves, as Figure 7.1 shows. In 2008, the highest per-inhabitant GDP was more than twice the lowest in 13

Map 7.1: Gross domestic product (GDP) per inhabitant, in purchasing power standard (PPS), by NUTS 2 regions, 2008 (1)
(in percentage of EU-27 = 100)



(1) Turkey, 2006.

Source: Eurostat (online data code: nama_r_e2gdp).

of the 23 countries examined here with several NUTS 2 regions. This group includes seven of the nine new Member States/candidate countries, but only six of the 14 EU-15 Member States.

The largest regional differences are in Turkey, where there is a factor of 4.9 between the highest and lowest values, and in the United Kingdom and Romania, with factors of 4.8 and 3.9 respectively. The lowest values are in Slovenia, Ireland and Sweden, with factors of 1.4, 1.6 and 1.6. Moderate regional disparities in per-inhabitant GDP (i.e. factors of less than 2 between the highest and lowest values) are found only in EU-15 Member States, plus Slovenia and Croatia.

In all the new Member States, Croatia and a number of EU-15 Member States, a substantial proportion of economic activity is concentrated in regions that include the capital. Consequently, in 18 of the 23 countries included here in which there are several NUTS 2 regions, these regions are also those with the highest per-inhabitant GDP. For example, Map 7.1 clearly shows the prominent position of the regions of Brussels (Belgium), Sofia (Bulgaria), Praha (Czech Republic), Athina (Greece), Madrid (Spain), Paris (France) and Lisboa (Portugal) as well as Budapest (Hungary), Bratislava (Slovakia), London (United Kingdom), Warszawa (Poland) and Bucureşti (Romania).

A comparison of the extreme values between 2000 and 2008, however, shows that trends in the EU-15 have been quite different from those in new Member States. While the gap between the regional extreme values in the new Member States and Croatia is growing in most cases, it is shrinking in one out of every two EU-15 countries.

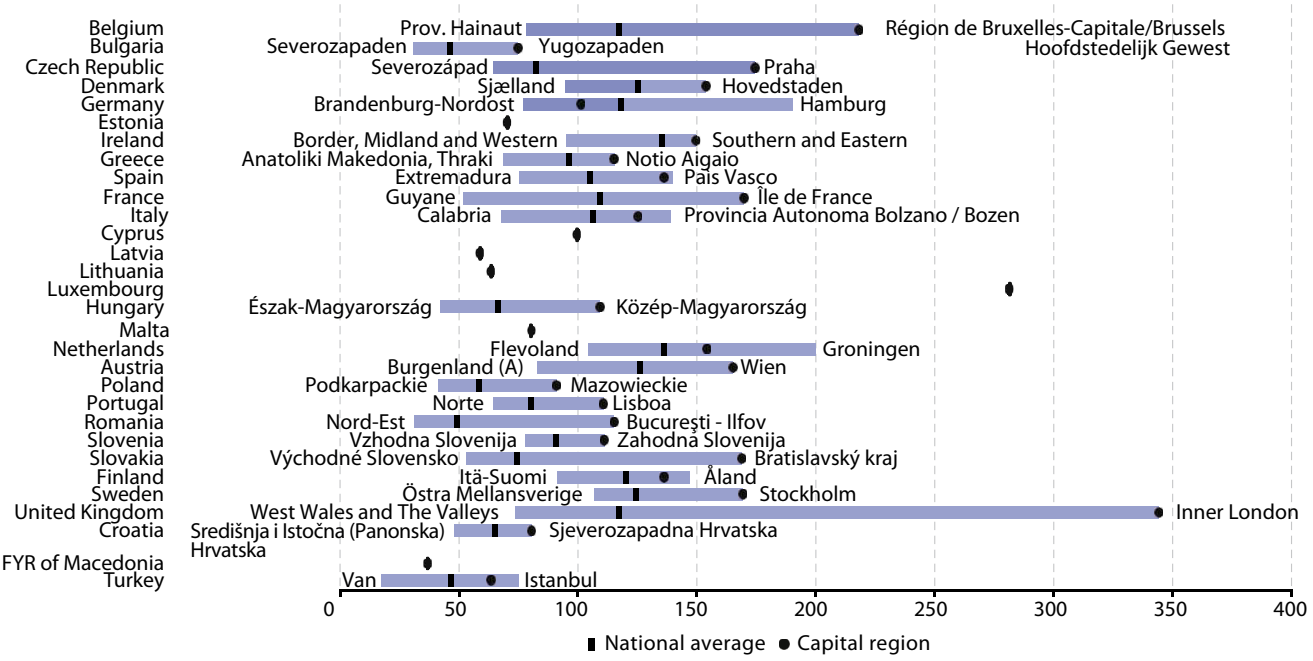
Dynamic catch-up process in the new Member States

Map 7.2 shows the extent to which per-inhabitant GDP changed between 2000 and 2008, compared with the EU-27 average (expressed in percentage points of the EU-27 average). Economically dynamic regions, whose per-inhabitant GDP increased by more than 3 percentage points compared with the EU average, are shown in green. By contrast, less dynamic regions (those with a fall of more than 3 percentage points in per-inhabitant GDP compared with the EU-27 average) are shown in orange and red. The range is from + 58 percentage points for Bratislavský kraj (Slovakia) to – 40 percentage points for Brussels in Belgium.

The map shows that economic dynamism is well above average in the south-western, eastern and northern peripheral areas of the EU, not just in EU-15 countries but particularly in new Member States, Croatia and some regions of Turkey.

Among the EU-15 Member States, strong growth is particularly evident in Spain, parts of the Netherlands and Greece, as well as the north of Finland and Sweden. On the other hand, weak growth that started several years ago is persisting in several EU-15 countries. Italy and France have

Figure 7.1: Gross domestic product (GDP) per inhabitant, in purchasing power standard (PPS), highest and lowest NUTS 2 region within each country, 2008 ([1])
(in % of the EU-27 average, EU-27 = 100)



([1]) Turkey, 2006.

Source: Eurostat (online data code: nama_r_e2gdp).

been particularly badly hit. Not a single region achieved the EU-27 average growth rate during the eight-year period 2000–08. Performance has also been weak in a number of regions of Germany, Portugal, Sweden and the UK. Ireland is a special case. Due to the economic and financial crisis, both NUTS 2 regions fell back to the levels of 2001, i.e. by 15 percentage points, during the year 2008.

Of the new Member States, apart from the very dynamic capital regions, the Baltic States, Romania, Slovakia, the Czech Republic and most regions of Poland have seen growth markedly above the average. Croatia and the former Yugoslav Republic of Macedonia also reveal above-average economic growth for the eight-year period 2000–08.

Closer analysis of the most dynamic regions shows that 41 EU-27 regions have outperformed the EU average by more than 10 percentage points; of these, 24 are in new Member States.

The 10 fastest-growing regions are spread over nine EU Member States. Among these 10, there are five capital regions in new Member States. The three regions in EU-15 countries in this top-10 group (Luxembourg, Groningen in the Netherlands and Inner London) can all be considered special cases.

The non-capital region with the strongest growth in the new Member States was Vest (Romania), where per-inhabitant GDP (in PPS) increased by 23.8 percentage points compared to the EU-27 average between 2000 and 2008.

At the lower end of the distribution curve, there is a clear concentration: of the 34 regions in which per-inhabitant GDP fell by more than 10 percentage points below the EU-27 average, 13 are in Italy, six in France, five in the UK and four in Germany.

Closer examination of the new Member States yields the pleasing result that, between 2000 and 2008, only one region (Malta with– 5.8 percentage points) fell back, compared with the EU-27 average.

The catch-up process in new Member States was of the order of 1.7 percentage points per year between 2000 and 2008, compared to the EU average. Per-inhabitant GDP (in PPS) in these 12 Member States thus rose from 45 % of the EU-27 average in 2000 to almost 59 % in 2008. In 2008, performance was particularly strong, with 2.7 percentage points. This can be explained partly by the fact that the economic and financial crisis struck first in the EU-15 Member States, some of which, like Ireland, Italy and Denmark, were already in recession in 2008. On the other hand, among new Member States, only Estonia and Latvia already had negative volume growth rates in 2008, and the full effects of the crisis became apparent only in 2009. The initial data available on certain Member States for 2009 and 2010 would suggest that the recession affected rural regions and areas lagging behind in terms of economic development less severely than regions with a high per-inhabitant GDP, or with a high level of dependence on exports or tourism.

Different trends within countries themselves

A more detailed analysis of trends within countries between 2000 and 2008 shows that the economic development of regions within a country can be almost as diverse as between regions in different countries.

The largest differences were seen in the Netherlands, Romania, Slovakia and the United Kingdom, where there were performance differences of more than 40 percentage points relative to the EU average for the per-inhabitant GDP of the fastest- and slowest-growing regions. The countries with the smallest differences between regions were Ireland, Slovenia, Denmark and Finland, with regional performance differences of between 2 and 9 percentage points.

In both new Member States and EU-15 countries, significantly diverging regional trends were the result mainly of dynamic growth in capital regions. However, as the values for Slovenia (6 percentage points) and Poland (14 percentage points) show, the data available do not confirm the assumption that major regional growth disparities are a typical feature of new Member States.

The data also show that the regions with the lowest levels of per-inhabitant GDP made significant progress. Between 2000 and 2008, Nord-Est and Sud - Muntenia (both in Romania) caught up by 11 and 18 percentage points and Yuzhen tsentralen (Bulgaria) by 9 percentage points compared to the EU-27 average.

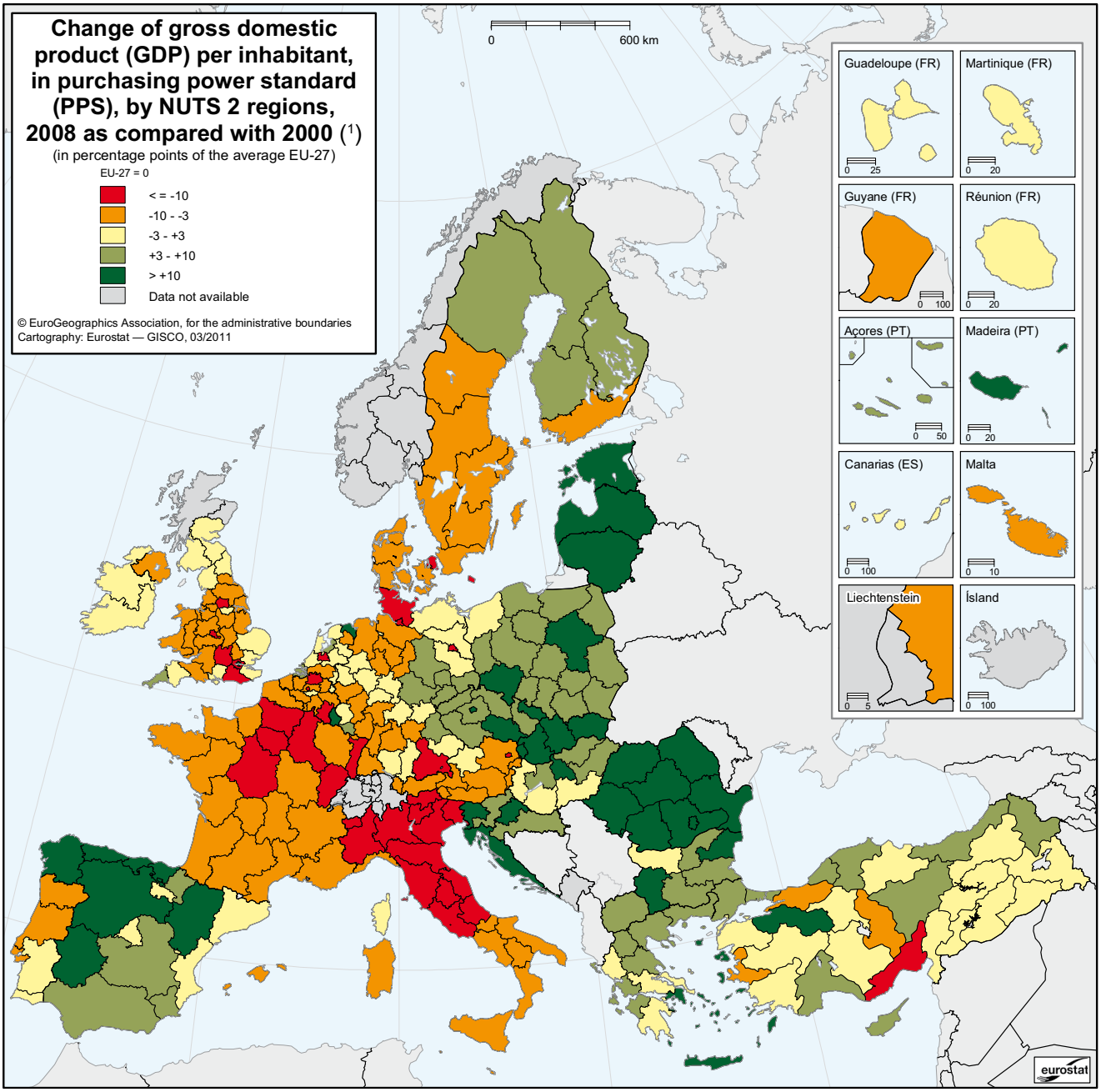
Convergence makes progress

This section addresses the question of whether convergence among the regions of the EU-27 has made progress over the eight-year period 2000–08. Regional convergence of per-inhabitant GDP (in PPS) can be assessed in various ways on the basis of data supplied to Eurostat by national statistical institutes.

The simplest approach is to measure the gap between the highest and lowest values. By this method, the gap closed from a factor of 17.2 in 2000 to 13.2 in 2008. The main reason for this clear convergence was faster economic growth in Bulgaria and Romania. However, as this approach looks only at the extreme values, it is clear that the majority of shifts between regions are not taken into account.

A much more accurate evaluation of regional convergence is afforded by the dispersion of regional GDP calculated by Eurostat for the EU-27 and Croatia since 2007 (for details of the method see below, 'Data sources and availability', 'Dispersion of regional per-inhabitant GDP'). This takes account of divergences from the national average in all NUTS 2 regions for each country in turn, weighted by the regional population. Figure 7.2 compares the values of dispersion at regional level NUTS 2 for 2000 and 2008; the order of countries follows the values ranked in 2008. In the

Map 7.2: Change of gross domestic product (GDP) per inhabitant, in purchasing power standard (PPS), by NUTS 2 regions, 2008 as compared with 2000 (1)
(in percentage points of the average EU-27)



(1) Denmark, Eurostat estimate; Turkey, 2006 as compared with 2000; Croatia, 2008 as compared with 2001.

Source: Eurostat (online data code: nama_r_e2gdp).

first instance, a downward trend is apparent, i.e. a decrease in regional dispersion for the EU-27 as a whole. An examination of the trend in individual countries reveals clear differences between certain groups of Member States. First, most of the EU-15 countries have lower dispersion than the new Member States. In addition, values in the EU-15 countries are generally decreasing, whereas they are increasing considerably in some of the new Member States. It is thus evident that the economic catching-up process in new Member States has so far gone hand-in-hand with increasing regional disparities.

The approach most often used at present involves classifying the regions according to their per-inhabitant GDP (in PPS) in relation to the average of the EU-27. This enables calculation of the proportion of the population living in more or less prosperous regions, and how this proportion has changed over time.

Table 7.1 shows clear progress in economic convergence between regions over the eight-year period 2000–08 for the EU-27, Croatia and the former Yugoslav Republic of Macedonia: the proportion of the population living in regions where per-inhabitant GDP is less than 75 % of the EU-27 average fell from 28.1 % to 24.4 %. At the same time, the proportion of the population living in regions where this value is greater than 125 % fell from 24.3 % to 19.4 %. These shifts at the top and bottom ends of the distribution meant that the proportion of the population in the midrange (per inhabitant GDP of 75–125 %) increased sharply from 47.6 % to 56.2 %. This corresponds to an increase of around 51 million inhabitants.

A comparison between the data for 2000 and 2008 reveals that eight regions managed to pass the 75 % threshold in the course of this period. These were two regions in Greece, as well as one region each in Spain, France, Poland, Romania, Slovenia and Croatia. These regions are home to 19.6 million people, or around 3.9 % of the population of the 29 countries examined here. At the same time, however, GDP in one Italian and one UK region, covering a total of 6 million inhabitants, i.e. approx. 1.2 % of the EU population, again fell below the 75 % threshold. Taking both developments into account, as a result of economic development between the years 2000 and 2008, the population living in regions with a GDP of more than 75 % of the EU-27 average grew by 13.6 million people.

A more detailed analysis shows that, in addition, many regions with a GDP of less than 50 % of the EU-27 average have made quite substantial progress. Between 2000 and 2008, the population living in these regions fell by almost a third, from 14.8 % to 9.3 % of the 29 countries examined here, i.e. by over 25 million. At the same time, only one region (the French overseas department of Guyane) fell back below the 50 % threshold.

Moreover, an examination of the 10 weakest regions as at 2000, where 4.8 % of the population lived at that time, shows that this group made strong progress. Per-inhabitant GDP in these regions rose, from 22.6 % to 36.4 % of the EU-27 average between 2000 and 2008. This shows the strong catch-up process under way in Bulgaria and Romania.

Data sources and availability

What is regional gross domestic product?

The economic development of a region is, as a rule, expressed in terms of its gross domestic product (GDP). This indicator is also frequently used as a basis for comparisons between regions.

But what exactly does it mean, and how can comparability be established between regions of different sizes and with different currencies?

A meaningful comparison can be made only by comparing the regional GDP with the population of the region in question. This is where the distinction between place of work and place of residence becomes significant. GDP measures the economic output achieved within national or regional boundaries, regardless of whether this was attributable to resident or non-resident employed persons. The use of GDP per inhabitant is, therefore, only straightforward if all employed persons involved in generating GDP are also residents of the region in question.

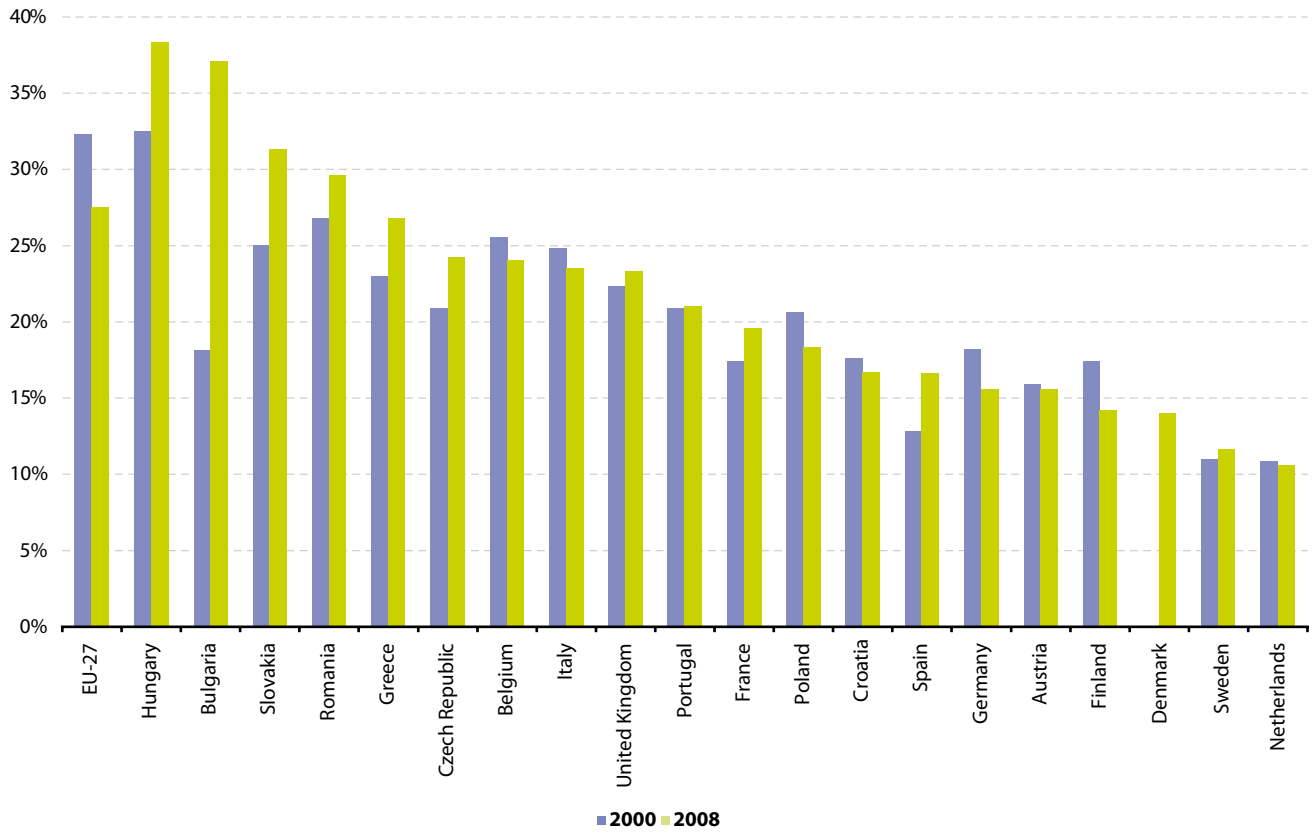
In areas with a high proportion of commuters, regional GDP per inhabitant can be extremely high, particularly in economic centres such as London (United Kingdom) or Wien (Austria), Hamburg (Germany), Praha (Czech Republic) or Luxembourg, and relatively low in the surrounding regions, even if households' primary income in these regions is very high. Regional GDP per inhabitant should, therefore, not be equated with regional primary income.

Regional GDP is calculated in the currency of the country in question. To make GDP comparable between countries, it is converted into euro, using the official average exchange rate for the given calendar year. However, exchange rates do not reflect all the differences in price levels between countries. To compensate for this, GDP is converted using conversion factors, known as purchasing power parities (PPPs), to an artificial common currency, called purchasing power standard (PPS). This makes it possible to compare the purchasing power of different national currencies.

Purchasing power parities and international volume comparisons

International differences in GDP values, even after conversion via exchange rates to a common currency, cannot be attributed solely to differing volumes of goods and services. The 'level of prices' component is also a major contributing factor. Exchange rates reflect many factors

Figure 7.2: Dispersion of regional GDP per inhabitant, in PPS, NUTS level 2, 2000 and 2008 (1)
(%)



(1) Regional dispersion is not applicable for Estonia, Ireland, Cyprus, Latvia, Lithuania, Luxembourg, Malta and Slovenia; Croatia, 2001 and 2008.

Source: Eurostat (online data code: nama_r_e0digdp).

Table 7.1: Proportions of resident population of EU-27, Croatia and former Yugoslav Republic of Macedonia in economically stronger and weaker regions

Percentage of population of EU-27, Croatia and FYR of Macedonia resident in regions with a GDP per inhabitant of	2000	2008
> 125 % of EU-27 = 100	24.3	19.4
> 110 % to 125 % of EU-27 = 100	15.5	16.0
> 90 % to 110 % of EU-27 = 100	21.5	24.7
> 75 % to 90 % of EU-27 = 100	10.5	15.5
less than 75 % of EU-27 = 100	28.1	24.4
of which: less than 50 % of EU-27 = 100	14.8	9.3

Source: Eurostat (nama_r_e2gdp).

relating to supply and demand in the currency markets, such as international trade, inflation forecasts and interest rate differentials. Conversions via exchange rates are, therefore, of only limited use for international comparisons. To obtain a more accurate comparison, it is essential to use special conversion rates which remove the effect of price-level differences between countries. Purchasing power parities are currency conversion rates of this kind, converting economic data expressed in national currencies into an artificial common currency, called purchasing power standard (PPS). PPPs are, therefore, used to convert the GDP and other economic aggregates (e.g. consumption expenditure on certain product groups) of various countries into comparable volumes of expenditure, expressed in PPS.

With the introduction of the euro, prices can now, for the first time, be compared directly between countries in the euro area. However, the euro has different purchasing power in different countries within the euro area, depending on the national price level. PPPs must, therefore, also continue to be used to calculate pure volume aggregates in PPS for Member States within the euro area.

In their simplest form, PPPs are a set of price ratios, which show the relationship between the prices in national currency of the same good or service in different countries (e.g. a loaf of bread costs EUR 1.87 in France, EUR 1.68 in Germany, GBP 1.45 in the UK, etc.). A basket of comparable goods and services is used for price surveys. These are selected so as to represent the whole range of goods and services, taking account of different consumption structures in different countries. The simple price ratios at product level are aggregated to PPPs for product groups, then for overall consumption and, finally, for GDP. To have a reference value for the calculation of the PPPs, a country is usually chosen and used as the reference country, and set to 1. For the European Union, the selection of a single country as a base seemed inappropriate. Therefore, the PPS is the artificial common reference currency unit used in the EU to express the volume of economic aggregates for the purpose of spatial comparisons in real terms.

Unfortunately, for reasons of cost, it will not be possible in the foreseeable future to calculate regional currency conversion rates. If such regional PPPs were available, the GDP in PPS for numerous peripheral or rural regions of the EU would probably be higher than that calculated using the national PPPs.

Calculating in PPS instead of euros can lead to differences in the ranking of regions. For example, in 2008, the Swedish region of Östra Mellansverige was recorded as having a per-inhabitant GDP of EUR 30 800, ranking above the Italian region of Marche, with EUR 26 700. However, in PPS, Marche, at PPS 26 500 per inhabitant, is ahead of Östra Mellansverige, at PPS 26 200.

In terms of distribution, the use of PPS rather than the euro has a levelling effect, as regions with a very high per-inhabitant GDP also generally have relatively high price levels. This reduces the range of per-inhabitant GDP in the NUTS 2 regions in the EU from around EUR 85 300 to around PPS 79 300.

Per-inhabitant GDP in PPS is the key variable for determining the eligibility of NUTS 2 regions under the European Union's structural policy.

Dispersion of regional per-inhabitant GDP

Since 2007, Eurostat has calculated a new, derived indicator which records the differences between regional per-inhabitant GDP and the national average, and makes them comparable between countries. This dispersion indicator is calculated at NUTS 2 and at NUTS 3 levels. The figures used by Eurostat are based on GDP in purchasing power standards (PPS).

For a given country, the dispersion 'D' of the regional GDP of the level 2 regions is defined as the sum of the absolute differences between regional and national GDP per inhabitant, weighted on the basis of the regional share of population and expressed in percent of the national GDP per inhabitant:

$$D = 100 \frac{1}{Y} \sum_{i=1}^{n} \mid (y_i - Y) \mid (p_i / P)$$

In the above equation:

- y_i is the regional per-inhabitant GDP of region *i*;
- Y is the national average per-inhabitant GDP;
- p_i is the population of region *i*;
- P is the population of the country;
- *n* is the number of regions of the country.

The value of the dispersion of GDP per inhabitant is zero if the values of regional GDP per inhabitant are identical in all regions of the country or economic area (such as the EU or the euro area), and it will show, all other things being equal, an increase if the differences in per-inhabitant GDP between the regions grow. A value of 30 % therefore means that the GDP of all regions of a given country, weighted on the basis of the regional population, differs from the national value by an average of 30 %.

Context

GDP is an important indicator of economic activity and growth in a region. It is used to make comparisons between Member States of the EU and is crucial in determining a wide range of policies, such as the extent to which a Member State should contribute to the EU budget.

Three-year averages of GDP, for example, are particularly important, because they are used to decide which regions are eligible to receive support from the European Union's Structural Funds.