

## Measuring economies

# The trouble with GDP

Gross domestic product (GDP) is increasingly a poor measure of prosperity. It is not even a reliable gauge of production

Print edition | Briefing Apr 30th 2016

ONE of Albert Einstein's greatest insights was that no matter how, where, when or by whom it is measured, the speed of light in a vacuum is constant. Measurements of light's price, though, are a different matter: they can tell completely different stories depending on when and how they are made. In the mid 1990s William Nordhaus, an economist at Yale University, looked at two ways of measuring the price of light over the past two centuries. You could do it the way someone calculating GDP would do: by adding up the change over time prices of the things people bought to make light. On this basis, he reckoned, the price of light rose by a factor of between three and five between 1800 and 1992. But each innovation in lighting, from candles to tungsten light bulbs, was far more efficient than the last. If you measured the price of light in the way a cost conscious physicist might, in cents per lumen-hour, it plummeted more than a hundredfold. Mr Nordhaus intended this example to illuminate a general point about how flawed economists' attempts to measure changes in living standards are. Any true reckoning of real incomes must somehow account for the vast changes in the quality of things we consume, he wrote. In the case of light, a measurement of inflation based on the cost of things that generated light and one based on a quality-adjusted measure of light itself would have differed by 3.6% a year. When a first-year undergraduate first encounters the idea of GDP as the value added in an economy, adjusted for inflation, it sounds pretty straightforward, says Sir Charles Bean, the author of a recent review of economic statistics for the British government. Get into the details, though, and it is a highly complex construct—and, as Mr Nordhaus's fable shows, a snare for the unwary. The production boundary Measuring GDP requires adding up the value of what is produced, net of inputs, across a wide variety of business lines, weighting each according to its importance in the economy. Both the output and the materials (if any) used up in making it have to be adjusted for inflation to arrive at a figure that allows for comparison with what has gone before. This is tricky enough to do for an economy of farms, production lines and mass markets—the setting in which GDP was first introduced. For today's rich economies, dominated by made-to-order services and increasingly geared to the quality of experience rather than the production of ever more stuff, the trickiness is raised to a higher level. No wonder GDP statistics are still so prone to constant and substantial revision. The problem is not just that it is hard to make these calculations. It is that what the calculations produce is a

measure put to too many purposes, and, though useful, not truly fit for any of them. And there are worries that things may be getting worse.

As the price of light illustrates, standard measures miss some of the improvements delivered by innovation. But at least new lighting products show up in the figures once people start buying the things in sufficient volume.

These days it seems that a growing fraction of innovation is not measured at all. In a world where houses are Airbnb hotels and private cars are Uber taxis, where a free software upgrade renews old computers, and Facebook and YouTube bring hours of daily entertainment to hundreds of millions at no price at all, many suspect GDP is becoming an ever more misleading measure.

The modern conception of GDP was a creature of the interwar slump and the second world war. In 1932 America's Congress asked Simon Kuznets, a Russianborn economist, to estimate national income over the preceding four years. Until he produced his figures just over a year later, no one knew the full extent of the Depression. In Britain Colin Clark, an enterprising civil servant, had been collecting statistics on national income since the 1920s, and in 1940 John Maynard Keynes made a plea for more detailed figures on Britain's capacity to make guns, tanks and aeroplanes. He went on to establish the modern definition of GDP as the sum of private consumption and investment and government spending (with account taken for foreign trade). Kuznets had treated government spending as a cost to the private sector, but Keynes saw that if wartime procurement by the state was not treated as demand, GDP would fall even as the economy grew. Keynes's idea of GDP won out on both sides of the Atlantic and soon spread further. Countries that wanted to receive post-war aid under America's Marshall plan had to produce an estimate of GDP. In the 1950s Richard Stone, a protégé of Keynes, was asked by the United Nations to prepare a template for GDP accounting that could be used by all member states. To be a nation was, in part, to know your GDP. In wartime, GDP was concerned with managing supply. With peace, the influence of Keynes's ideas on fighting slumps flipped it into a way to manage demand, as Diane Coyle notes in her book, "GDP: A Brief but Affectionate History". Either way it was (and is) a measure of production, not of welfare—which, as GDP growth became a goal for politicians, also became an occasion for criticism. A measure created when survival was at stake took little notice of things such as depreciation of assets, or pollution of the environment, let alone finer human accomplishments. In a famous speech in March 1968, Robert Kennedy took aim at what he saw as idolatrous respect for GDP, which measures advertising and jails but does not capture "the beauty of our poetry or the strength of our marriages". It's a manufacturer's world.

From time to time, such dissatisfactions have brought forth alternatives. In 1972 Mr Nordhaus and James Tobin, a colleague at Yale, came up with a “measure of economic welfare” which counted some bits of state spending, such as defence and education, not as output but as a cost to GDP. It also adjusted for wear-and-tear to capital and the “disamenities” of urban life, such as congestion. The paper was in part a response to environmentalist concerns that GDP treats the plunder of the planet as something that adds to income, rather than as a cost. It was much talked about; it was not much acted on. In 2009 a report commissioned by the French president, Nicolas Sarkozy, and chaired by Joseph Stiglitz, a prominent economist, called for an end to “GDP fetishism” in favour of a “dashboard” of measures to capture human welfare. Kennedy was right. Much that is valuable is neither tangible nor tradable. But much that is tradable is also not tangible. A problem with GDP even when it is being asked to do nothing more than measure production is that it is a relic of a period dominated by manufacturing. In the 1950s, manufacturing made up more than a third of British GDP. Today it makes up a tenth. But the output of factories is still measured much more closely than that of services. Manufacturing output is broken down into 24 separate industries in the national accounts; services, which now make up 80% of the economy, are subdivided into only just over twice that number of categories. A bias toward manufacturing is not the only distortion. By convention GDP measures only output that is bought and sold. There are reasons for this, only some of them sound. First, market transactions are taxable and therefore of interest to the exchequer, an important consumer of GDP statistics. Second, they can be influenced by policies to manage aggregate demand. Third, where there are market prices, it is fairly straightforward to put a value on output. This convention means that so-called “home production”, such as housework or caring for an elderly relative, is excluded from GDP, even though such unpaid services have considerable value. In early editions of his bestselling economics textbook Paul Samuelson joked that GDP falls when a man marries his maid.

Despite convention, a lot of what is included in GDP lies outside the market economy. Many government services are provided free, and for decades the value given to such output was simply the cost of provision. It is only fairly recently that statisticians have started to measure some bits of public-sector output directly by, for instance, counting the number of operations performed by health services or the number of students taught in schools. Some private-sector services are also measured indirectly. Housing services is one.

This is straightforward wherever householders rent the property they live in. Rental payments capture both the value of housing services to tenants as well as the income of landlords from providing them. But in places where most people own the home they live in, a large part of the total value of housing services has to be imputed. Finance is another activity that is mostly measured obliquely (and badly). Typically financial services are not paid for directly in fees: banks make a large part of their income from charging more interest on loans than they pay on deposits. To capture the value being added, statisticians use an imputed figure, the “spread” between a risk-free interest rate and a lending rate, and multiply this by the stock of loans. The problem with this method is that the lending spread is a measure of the risk banks take. For this reason its use in GDP figures can have perverse results. For example, at the turn of 2009 Britain’s financial sector was close to collapse. But because fear of bank defaults was driving spreads up, GDP figures recorded a spike in the sector’s value added, and thus its contribution to GDP (see chart 1). As statisticians try to capture ever more of the economy’s output in their figures, new activities are added to GDP.

The price is wrong

A further complication is that, for all the caution that statisticians offer against seeing GDP as a measure of welfare, the two are intertwined in perhaps the trickiest part of their calculations: adjusting for inflation. Inflation is a measure of how much more you have to pay this year than you did last year to achieve the same level of well-being. It is at least as challenging to measure as output. For a start, a change in the price of a product will influence how much of it people buy. If red apples rise in price, people buy more green apples; if the price of beef shoots up, they buy more pork. There are tricks that capture this sort of substitution when compiling price measures. One is the “geometric-mean aggregation” of price quotes. Multiplying together the prices of  $n$  goods and then taking the  $n$ th root of the product allows price aggregations to take into account a degree of switching proportionate to the change in relative prices. This sounds abstruse: but getting it right has an effect of lowering inflation by half a percentage point or so. Broader shifts in consumer preferences are picked up by updating the weights attached to each category of goods in the overall price index. Then come adjustments for changes in quality. This year’s smartphone might cost more than last year’s, but if so it will also do more. If statisticians focus only on changes in price, they will overstate the true inflation rate by missing improvements in performance. An

advisory committee of leading economists set up by America's Senate in the mid-1990s and headed by Michael Boskin, of Stanford University, reckoned that failure to adjust for quality and new products meant true inflation was overstated by at least 0.6% a year. It called for greater use of "hedonic" estimation, a technique that captures the implicit value of each particular attribute of a product by measuring how variation in those traits affects the product's price: for example, how much more do people pay for a brighter light bulb? Once an implicit price for each attribute is established—processor speed, or memory, say, for a phone—prices are tweaked accordingly. Hedonic estimation helps. But it is a labour-intensive business, because the implicit prices have to be updated frequently to ensure accuracy; in practice only a small fraction of prices are adjusted in this manner. It also runs into problems when quantitative changes get so large as to become qualitative. A modern flatscreen television is simply a different beast from the squat little cathode-ray tube numbers of the 1980s. Such adjustments are even harder to do for services, which tend to be bespoke, than for goods, which are still for the most part standardised. The value of a meal, for instance, depends on the cooking and ingredients but also on the speed of service, the background noise, how close together the tables are, and so on. Each of these factors can change from one period to the next. The true value of public-sector services is even harder to measure comparably over time. The number of operations can be counted quarter by quarter. Their effects on health and longevity may not be seen for years or decades.

As the Boskin commission pointed out, new products are a particular headache. In theory their value to consumers is the gap between the reservation price (what consumers are willing to pay) and the actual price, known as "consumer surplus". In practice, new products enter the consumer-price index without any such adjustment. Then there is the sort of novelty that broadens choice. The number of TV channels or over-the-counter painkillers available in America, for instance, is overwhelming. Yet in 1970 there were just five of each. Though people may complain about too much choice, this greater variety is to a great extent a boon. But it is invisible to GDP measures. For GDP, the output of a million of shoes in one size and colour is the same as a million shoes in every size and colour. The benefits of many new products are simply not picked up at all. The upfront costs of providing services on a digital platform, such as Facebook or Twitter, are hefty. But the marginal cost is close to zero, and the explicit price to users is normally nothing. By global convention, zero-priced goods are excluded from GDP.

So are all voluntary forms of digital production, such as Wikipedia and open source computer programs. Some of this unpaid-for activity can be picked up in the accounting; although there is no charge for a Google search, consumers pay a shadow price by supplying information and attention, for which advertisers pay. But the advertising revenue is likely to be well below the benefits that consumers get.

The review chaired by Sir Charles Bean outlined two other possible approaches to valuing free digital services. One is to estimate the value of the time spent on the internet. The Bureau of Economic Analysis, America's main statistical body, has used market wage rates to estimate the value of home-production activities, such as cooking, cleaning and ironing. Following a similar approach, Erik Brynjolfsson and Joo Hee Oh of MIT estimated that the welfare gain of free internet products added 0.74% a year to America's GDP between 2007 and 2011 (other studies reach somewhat lower estimates). The other approach uses rising internet traffic as a proxy (see chart 2). The review cites research which found consumer internet traffic in western Europe growing at 35% a year from 2006 to 2014. If the output of IT services had grown at a similar clip, official GDP growth rates in Britain would have been 0.7 percentage points higher each year.

It is not just that many new services are now given away free; so are some that used to be paid for, such as long-distance phone calls. Some physical products have become digital services, the value of which is harder to track. It seems likely, for instance, that more recorded music is being listened to than ever before, but music-industry revenue has shrunk by a third from its peak. Consumers once bought newspapers and maps. They paid middlemen to book them holidays. Now they do much more themselves, an effort which doesn't show up in GDP. As commerce goes online, less is spent on bricks-and-mortar shops, which again means less GDP. Just as rebuilding after an earthquake (which boosts GDP) does not make people wealthier than they were before, building fewer shops does not make them poorer.

These problems do not invalidate the use of GDP. But given the direction of technological change in an ever-more digital world they seem likely to grow more serious, and solutions to them are both hard and imperfect. Measuring the consumer surplus from new or free products relies on brave assumptions; estimates vary widely depending on which ones are used. To be consistent over time would require measuring the consumer surplus of goods and services that are well established in the consumer basket.

A sense of the scale of the task can be gained from looking at estimates of how fast the economy grew during a previous time of headlong technological change—the Industrial Revolution. Around the time that GDP was first being used to measure contemporary economies, some economic historians ventured to apply it to the past, too. They concluded that there had been a sudden take-off in economic growth after 1750; a landmark post-war study reckoned that GDP per worker rose by 1.4% a year, an unprecedented rate, in the first half of the 19th century.

You say you measured a revolution

In the 1980s, research by Nicholas Crafts of Warwick University found that the 18<sup>th</sup> century's glut of industrially transformative inventions had been applied rather narrowly, with madcap growth seen only in a few sectors of the economy. He put productivity growth at a less revolutionary 0.5% a year. A generation further on colleagues of Mr Crafts, led by Steve Broadberry, published research which nudged the figures back up a bit. Even centuries on, it is hard to settle on GDP estimates in times of upheaval. And they still miss many of the changes wrought—the consumer surplus due to railways, say. “It is a big mistake to think that one number serves for all purposes,” says Sir Charles. The problem is that, as things stand, GDP risks serving all its purposes ever-less well. The Bank of England has become so chary of GDP figures that it publishes a range of numbers both for its forecasts of growth and for its history. Its latest projections put recent GDP growth in Britain somewhere between zero and 4%. Such hyper-scepticism might seem a bit silly. But is it really no more absurd than proclaiming, with great certainty, that GDP growth in China fell from 6.8% to 6.7% in the year to the first quarter, when it almost certainly didn't?

If comparisons of GDP from one quarter to the next are dodgy, those from decade to decade are perilous. America's Census Bureau calculates that median household income, adjusted for inflation, was barely higher in 2014 than it was 25 years earlier. Measured living standards for a typical American have stagnated for a quarter-century, in other words. This finding undoubtedly reflects something real. But would a typical American really be indifferent between 1989 medical care at 1989 prices and today's medical services at current prices, asks Ken Rogoff of Harvard University? If GDP figures really measured what they try to measure, that would be the rational stance.

The challenge, said Mr Nordhaus in his paper on light, is to construct measures that “account for the vast changes in the quality and range of goods and services that we consume.” But that means finding ways to more readily compare hand-held e-mail with fax machine, self-driving car with jalopy, vinyl records with music streaming services and custom-made prosthesis with

health-service crutches. Perhaps an Einstein could do it. Odds are, though, that he'd take one look and stick with the simplicities of physics instead.