GDP and the System of National Accounts: Past, Present and Future.

Nicholas Oulton

Centre for Macroeconomics/
London School of Economics,
National Institute of Economic and Social Research
and
Economic Statistics Centre of Excellence

January 2018

1. Introduction[1]

"China became the world's largest economy in 2014". "UK GDP grew by 0.2% in the first quarter of 2017". "In the eurozone, inflation as measured by the Consumer Price Index was 1.4% in the year to May 2017". Any scanner of websites that cover business news can read statements like these on any day of the week. Each statement relies on modern economic statistics whose basis is the System of National Accounts (SNA). In this chapter I briefly outline how the SNA came to have such a powerful (if background) role. I then go on to discuss some of the many criticisms levelled at the SNA, and in particular at GDP, its centrepiece. These criticisms fall into two groups. The first group raises doubts about how accurately GDP is measured. The second is more about the relevance of GDP (and the SNA) as a guide to policy. Even if GDP is measured accurately, is it measuring anything which thoughtful people should be interested in?

2. GDP and the SNA: a brief history[2]

Simon Kuznets was one of the founders of national income accounting. (He was awarded the Nobel Prize in Economics in 1971.) In 1959 he published a study which revealed perhaps the most important empirical finding in the whole of economics (Kuznets 1959).[3] His discovery was that economic growth, i.e. the growth rate of GDP per capita, has been much higher after the industrial revolution than it had been at any earlier time. So the countries fortunate enough to have passed through the industrial revolution experienced a dramatic acceleration in economic growth and (eventually at least) in living standards. Therefore the industrial revolution marks a new epoch in human history. To non-economists the industrial revolution is usually characterized by the great inventions which accompanied it such as steam power and railways. But the advent of great inventions does not necessarily lead to faster growth of per capita GDP on a sustained basis. For a counter-example, consider the 15th and 16th centuries in Europe which saw the invention of printing and improvements in shipbuilding and navigation such as the magnetic compass which in turn led to the conquest and settlement of the Americas. But we now know that these great discoveries did not lead to an appreciable increase in the European growth rate.

---

[2] See Coyle (2014) and her contribution to this volume for a more in-depth treatment.
[3] He developed the argument further in Kuznets (1966).

How did Kuznets reach his dramatic conclusion? After all, in 1959 he only had data for 19 countries and these data only stretched back in most cases for about 80 years. He had no data for any country before the industrial revolution. The answer is that he employed a thought experiment. He took the growth rates of GDP per capita which he had measured in his sample of countries (mostly in the range 1-2 % per year) and then asked: suppose these growth rates had prevailed in earlier centuries, how low would the standard of living have been two hundred years ago or five hundred years ago? He calculated that the standard of living would have been so low that no-one could have survived. But if they couldn't have survived, then we wouldn't be around to do these calculations today! Therefore growth rates must have been lower before the industrial revolution than after it. One can easily convince oneself of Kuznets' point by calculating what sum would grow to say $1000 (roughly equal to the World Bank's global poverty standard for annual income today). if compounded at 1% over two hundred years. The answer is $135, less than a dollar a day. Compounding over 500 years, the answer is less than 7 dollars a year. Clearly these income levels are impossible. Kuznets' conclusions have subsequently been amply confirmed by direct estimates of income levels and growth rates in pre-modern economies e.g. Broadberry et al. (2016) for Britain.

Uses of GDP

During the second world war pioneering estimates of GDP were used by the UK and US governments for planning the war. Estimates of GDP in current prices sufficed for this purpose since the main question was how much could be spent on the armed forces without reducing household expenditure to an unacceptable level. After the end of the second world war the national income accounting revolution spread rapidly across the world. The United Nations under the guidance of other pioneers like Richard Stone (awarded the Nobel Prize in Economics in 1984) took up the challenge of producing an internationally accepted System of National Accounts (SNA). The first version, all of 48 pages long, appeared in 1953. Subsequent versions have appeared in 1968, 1993 and 2008 and further updates are planned. The latest version (European Commission et al. 2009) has grown to 662 pages. For a time the Soviet Union employed and enforced on its satellites a rival system, the Material Product System, based on Marxist principles. The disappearance of the Soviet Union has meant the disappearance of the MPS too, even in countries like China run by communist parties.

The post-war development of the SNA met the needs of Keynesian macroeconomic management, support for which was spreading rapidly. For this purpose GDP in constant as well as current prices is necessary. Quarterly as well as annual estimates of GDP started to appear. Keynesian notions of macroeconomic management are now less popular than they once were but central banks with a remit to target inflation are just as keen to receive high quality and frequent estimates of GDP and the other main components of national income.

In parallel with the needs of monetary and fiscal policy a new market for GDP and the SNA has arisen because of increasing interest in the problems of long run growth and development both in developing and developed countries. And this has sparked innovations in official statistics too, such as the OECD's capital and productivity manuals (OECD 2001 and 2009). The first of these manuals, on measuring capital, enshrined the fundamental distinction between capital stocks and capital services, originally introduced by Jorgenson (1989), and showed how it could be incorporated into the SNA. The second manual, on productivity, building on the pioneering contributions of Jorgenson and Griliches (1967) and Jorgenson et al. (1987), employed the concept of capital services to show how theoretically consistent measures of total factor productivity growth could be derived within the framework of the SNA.

With the rise of major new economic powers like China and more recently India there has also been increasing interest in international comparisons of the size of different economies (GDP) and their relative standards of living (GDP per capita). The crucial institution here is the International Comparison Program (ICP) run by the World Bank in conjunction with the OECD. The 2005 round of the ICP included 146 countries, covering 95% of the world's population (World Bank 2008). The latest round in 2011 included 199 countries, virtually all the countries in the world, though full results are available for only 177 (World Bank 2015). Just as national statistical agencies (NSAs) track prices over time for their domestic price programs such as the Consumer Price Index, so the ICP tracks prices across countries at a moment in time, e.g. mid-2011, via a collaborative and coordinated network of NSAs. The prices of the individual products and the overall averages for aggregates like household consumption or GDP, all measured relative to US dollar prices in the US, are known as Purchasing Power Parities (PPPs). In both the national and international programmes broadly the same methodology is used: "matched models", under which the agencies try to track the prices of *identical* models either over time or across space. The results can be controversial in

some cases. China (and Asia generally) turned out to be considerably poorer under the 2005 comparison than many observers had expected. Following methodological changes China's and Asia's ranking rose substantially in the 2011 ICP (Deaton and Aten 2017).

3. How accurately is GDP measured?

At least in countries with well-developed statistical systems, GDP in current prices (nominal GDP) is considered to be measured reasonably well. (It may be a different matter in poor countries (Jerven 2013)). There is much more concern about GDP in volume terms, real GDP. This is because moving from nominal to real GDP requires deflating each component by an appropriate price index. There are two major issues with price indices. First, they may not make adequate allowance for quality change and for new goods. Second for some components of GDP price indices often do not exist and so are replaced by proxies or by conventions.

Bias in price indices

There has long been concern that price indices may understate quality change and not make adequate allowance for the appearance of new goods, so leading to an overstatement of inflation and an understatement of real economic growth. (Very few researchers have advocated the opposite position though it may be true of individual products.[4]) Most of this evidence is for the United States but there is no reason to think that other developed countries are any better. Perhaps the strongest advocate of this view is Robert Gordon. His earlier work uncovered a huge underestimate of quality change in durable goods prices in the US in the nineteenth and twentieth centuries up to the early 1980s (Gordon 1990). For example over the period 1947-1983 he found that the rate of growth of the official producers' durable equipment deflator was 3 per cent per annum too high and the official deflator for consumer's durable expenditure was 1.5 per cent per annum too high. He reached this result by replicating the methods used by the US Bureau of Labor Statistics (BLS) but using extensive

---

[4] For example, when the US Bureau of Labor Statistics introduced hedonic methods to measure commercial rents it found that the new index rose more rapidly than the "matched models" index which it was replacing. The reason was that the old method used "matched apartments" to measure rents. But over time the apartments being matched were getting older and so less desirable and this was reflected in the market by declining rents. So the old index was understating inflation in commercial rents and hence overstating growth in the real volume of housing services (Wasshausen and Moulton 2006).

non-official data, mostly successive issues of the Sears mail order catalogues. These gave prices together with descriptions of items like lawnmowers, sometimes accompanied by photographs, so he was able to apply the "matched models" method of statistical agencies. The "matched models" method involves tracking the price of the *same* model over time, thus holding quality constant.

In his more recent book (Gordon 2016) he has argued strongly that the growth in the US standard of living since the Civil War and up to the 1970s is severely understated by official statistics because the revolutionary new products which became available to the typical family over these decades — flush toilets, cars, radio, films, TV, air travel, etc —are not given full credit in the national accounts.

The Advisory Commission to Study the Consumer Price Index (1996), commonly known as the Boskin Commission (of which Gordon was a member), argued that the CPI had been overstating US inflation by over 1% per year for the years leading up to 1996 due to a combination of factors including inadequate allowance for new goods and quality change. Other factors were substitution bias, outlet bias and formula effects.

There are two problems with incorporating the effect of new goods on inflation. First, by virtue of its newness, it may be some time before it is introduced into the price index. Second even when it has been introduced into the index, its effect on the standard of living is understated since only price changes after it has been introduced affect the index. Everyone may agree that the new good represents a significant increase in welfare but this is not captured in the price index and so does not lead to an impact on real income. The first problem is an administrative and budgetary one. The second is more conceptual.

In fact, economic theory has long known how to cope with new goods (or vanished old goods) in calculating a price index. In the case of a new consumer good (or a new input), we should treat it as if it had always existed but at a price which just reduced demand for it to zero, that is, its reservation price, also called its virtual price. More precisely, the reservation price is the minimum price which just reduces demand to zero. Prior to its appearance, the new good's reservation price should be included in the price index, and the good's actual price should be included after its appearance (Hicks 1940). This makes it clear why ignoring new goods leads to an overestimation of price rises and a consequent underestimation of real

growth. For the price of the new good has in fact fallen, from its reservation level to its observed level, which is necessarily lower.

The problem is how to estimate the reservation price. Researchers have done this for individual products, most notably Hausman (1997) for a new brand of breakfast cereal (Apple Cinnamon Cheerios), see also Hausman (2003), but the results are controversial since they are dependent on particular assumptions about demand and on econometric methods (Groshen et al. 2017). A more easily implementable approach, based on a CES demand system (Feenstra 1994; Redding and Weinstein 2016), may be appropriate in some contexts but also suffers from restrictive assumptions about the pattern of demand and has the unpalatable property that the reservation price is infinite. More to the point, no statistical agency currently uses the reservation price approach to measure the impact of new goods. So the problem has been parked and we must wait for further research to see whether a practical method can be developed ("practical" meaning in part, within the budget that governments are willing to allot to statistical agencies). How much difference improved methods would make is hard to judge though ballpark figures like an additional 0.5% per annum on GDP are sometimes mentioned. It is probably no accident that there has been renewed interest recently in possible understatement of GDP growth since GDP and productivity growth seem to have slowed down at least since the Great Recession began at the end of 2007 (or perhaps earlier). But there seems little reason to ascribe the slowdown to mismeasurement since the latter was at least as great a problem prior to the appearance of a slowdown (Byrne et al. 2016; Syverson 2017).

Statistical agencies will no doubt implement improved methods as time goes on and research delivers new solutions.[5] But a point to bear in mind is that price indices are almost never revised. So the shortcomings of earlier methods will remain in the historical record, even if the most recent years are better measured.

Missing or inappropriate price indices

Real GDP can be measured either from the expenditure side, GDP(E), or from the output side, GDP(O). Consistency requires that the two measure should be equal. On the expenditure

---

[5]   See Bean (2016) for a comprehensive set of recommendations tailored to the British case for improving economic statistics.

side we have the familiar formula GDP = C + I + G + X − M. Private consumption (C) typically accounts for 60-65% of GDP and here we can rely on the prices gathered for the Consumer Price Index. The CPI program is the largest and best-funded of all price-gathering programs. Gross fixed investment (I) accounts for another 20% or so of GDP. Here we have to rely on the much less well-funded Producer Price Index programme. Exports (X) and imports (M) account for a large fraction of GDP (in some small countries a multiple of GDP) but what matters for GDP is the balance, typically a small proportion of GDP (plus or minus 1-3%). Since nowadays rich countries tend to trade mainly with each other and the goods imported and exported are similar, any errors in export and import price indices will tend to cancel out.

That leaves government consumption (G) — defence and public administration, education, and health — as the remaining major component of GDP(E) and here there is a serious deficiency. Until recently, most countries measured real government output by real government input (essentially hours worked adjusted for the composition of the labour force) which left very little room for productivity improvement and allowed for no improvements in quality. Nowadays some countries try to do better by using a collection of output measures weighted together by costs. For example, the output of the education sector can be measured by a weighted average of the numbers passing through each stage of the school system, weighted by the costs of providing each stage. This is better than measuring education output by hours worked in this sector but hardly addresses the quality issue. The quality issue is perhaps greatest in health where there have been large improvements in health outcomes, sometimes achieved at low cost; for example the incidence of heart attacks and strokes has been greatly reduced by statins and aspirin. It is clear that these improvements are not reflected in the price indices for health output and expenditure. Improving these indices is an active area of research (Groshen et al. 2017).

On the output side GDP is the sum of value added across industries. Here the appropriate price indices are Producer Price Indices and (where they exist) Service Producer Price |Industries. (Ideally, inputs need to be deflated separately from outputs but this is not always the case). In practice, statistical agencies tend to put much more weight on the expenditure side for estimates of real GDP. The reason is that the bulk of GDP(E) is private consumption where prices indices are comparatively well-measured, So (for example), the UK's Office for National Statistics adjusts the annual estimates of the growth of real GDP(O) so that they

conform to the growth of real GDP(E) to within 0.1% per annum (Lee 2011). They do this by adjusting the growth rates of private service industries. The reason, no doubt, why the adjustment falls on private services, is that this is where price indices are either inadequate or missing so that they have to be replaced by proxies like the CPI. A large fraction of the output of a modern economy (often larger than the proportion accounted for by manufacturing) is made up of industries supplying mainly intermediate services to business, such as finance and business services of all kinds (accountancy, advertising, contract cleaning, design, legal, management consultancy, computer and software services, etc). Here price indices are often of low quality or missing altogether (Timmer at al. 2010, pages 90-94). To the extent that we care just about GDP this does not matter, since these problems are largely absent on the expenditure side: business services are an intermediate product so drop out of GDP(E). But if we care also about what is happening in individual industries, say because we want to trace the origins and impact of the Great Recession, then we will also need better price indices for important industries like finance and business services.

Cross-country comparisons of price and income levels

Though it has attracted far less attention than possible deficiencies in consumer and producer price indices, the accuracy of PPPs is just as pressing an issue. There are conceptual problems which are yet to be fully resolved. To take one example, the relative income levels yielded by successive rounds of the ICP are not consistent with extrapolating from one round to the next using the national accounts of the countries studied. Whether this should be treated as a fact of life or adjusted for in some way is still a matter of debate. One extreme is to largely ignore national accounts and base international comparisons solely on successive PPPs. The other extreme is to pick the "best" single set of PPPs and ignore the others; this approach makes maximal use of national accounts. Debate continues on these and some compromise alternative (Oulton 2015).

4. Should we still care about GDP?

The commonest criticisms of GDP as a target of policy are the following:

1. GDP is hopelessly flawed as a measure of welfare. It ignores leisure and women's work in the home.

2. GDP ignores distribution. In the richest country in the world, the United States, the typical person or family has seen little or no benefit from economic growth since the 1970s. But over the same period inequality has risen sharply.

3. Happiness should be the grand aim of policy. But the evidence is that, above a certain level, a higher material standard of living does not make people any happier. So we should stop looking for policies to raise GDP and look instead for policies which promote happiness.

4. Even if higher GDP were a good idea on other grounds, it's not feasible because the environmental damage would be too great. The planet is finite, so if the truly poor in the third world are to be allowed to raise their standard of living by a modest amount, then consumers in the rich countries will have to accept a lower standard of living, i.e. lower not higher GDP per capita should be the aim for them.

I consider the first three criticisms in turn. Space precludes a discussion of the fourth though see Oulton (2012a) for some comments on this.

"GDP is hopelessly flawed as a measure of welfare"

GDP is and always was intended to be a measure of output, not of welfare. In current prices it measures the value of goods and services produced for final consumption, private and public, present and future; future consumption is covered since GDP includes output of investment goods. Converting to constant prices allows one to calculate growth of GDP over time (or differences between countries across space). The exclusion of home production and leisure is not due (I believe) to prejudice against women but to the desire on the part of national income accountants to avoid imputations wherever possible. However, it is not very difficult to include values for leisure and home production provided the necessary data on time use are available and provided one can decide on an appropriate wage rate to value time spent in non-market activities.

Though not a *measure* of welfare, GDP can be considered as a *component* of welfare. The volume of goods and services available to the average person clearly contributes to welfare in the wider sense, though of course it is far from being the only component. So one can

imagine a social welfare function which has GDP as one of its components along with health, inequality, human rights, etc: see the comments below on the Stiglitz-Sen-Fitoussi Report.

GDP is also an *indicator* of welfare. In practice, in cross-country data GDP per capita is highly correlated with other factors which are important for human welfare. In particular it is positively correlated with life expectancy, negatively correlated with infant mortality, and negatively correlated with inequality. Charts 1-3 illustrate these facts (actually these charts, from Oulton (2012a), plot household consumption per capita rather than GDP per capita against each welfare measure, but the picture for GDP would be very similar), for some 126-146 countries in 2005. In other words, richer countries tend to have greater life expectancy, lower infant mortality, and lower inequality (though this last relationship is not a linear one: some middle income countries have high inequality but nonetheless the richest countries are also the most equal ones). Correlation is not necessarily causation, though one might certainly make the case that higher GDP per capita causes improved health (Fogel, 2004; Deaton 2013).

Life expectancy rose steadily throughout the twentieth century and is still rising on average in the twenty-first. This means that people have more years in which to enjoy the higher consumption which they now receive, an aspect of welfare which is not captured just by the GDP statistics. But recently the United States has seen a rise in mortality amongst less-educated, middle-aged whites due it seems to self-harming behaviour — drug and alcohol dependency, accidents and suicide (Case and Deaton 2017). Whether this is a specifically American phenomenon, related perhaps to deficiencies in the US social safety net (Edin and Shaefer 2015), or whether the same phenomenon will appear in other developed countries remains to be seen.

According to the Commission on the Measurement of Economic Performance (the Stiglitz-Sen-Fitoussi Commission), policy should be concerned with well-being and well-being is multi-dimensional (Stiglitz et al., 2009, page 15):

"To define what well-being means a multidimensional definition has to be used. Based on academic research and a number of concrete initiatives developed around the world, the Commission has identified the following key dimension that should be taken into account. At least in principle, these dimensions should be considered simultaneously:
i. Material living standards (income, consumption and wealth);
ii. Health;
iii. Education;

iv. Personal activities including work
v. Political voice and governance;
vi. Social connections and relationships;
vii. Environment (present and future conditions);
viii. Insecurity, of an economic as well as a physical nature."

Few will disagree that these dimensions of life are important for human welfare and no-one can object to improved measurement. There is clearly a role for government in measuring and tracking these dimensions. To what extent though a dimension like "Social connections and relationships" should be objects of government policy is open to question. It is doubtful that effective policy levers exist. And even if they did the scope for a vast extension of the reach of government is worrying.

If one sticks to measurement and is a bit less ambitious than the Stiglitz-Sen-Fitoussi Report then further progress is possible. Jones and Klenow (2016) use an expected utility framework to combine measures of life expectancy, inequality and consumption to construct what they call a consumption-equivalent welfare measure for a large sample of countries. Their measure turns out to be highly correlated with GDP per capita.

"Growing GDP is pointless since most people don't benefit"

This claim is most often made in relation to the United States. Many people assert that real household income levels there have stagnated since the 1970s, despite labour productivity and GDP per capita growing quite rapidly.[6] It is non-controversial that income inequality has been rising for decades in the U.S. but does this mean that the typical household has received no benefit from growth? A comprehensive examination of these issues has recently appeared in an article by Wolff *et al*. (2012). Their results reveal quite a different picture.

They define a number of income concepts which are superior to GDP as a measure of household welfare: Comprehensive Disposable Income (CDI), Post Fiscal Income (PFI), and their preferred measure, the Levy Institute Measure of Economic Well-Being (LIMEW). CDI is household income, including property income (on an annuitized basis), less taxes plus cash and non-cash benefits. PFI adds to this individual public consumption (e.g. publicly-provided

---

[6] There is considerable evidence that mean real wages, analysed by age, sex and educational level, have stagnated since the 1970s. But this does not quite establish that living standards have also stagnated since the composition of the labour force might have shifted to better-paying jobs. And property income, taxes and benefits have to be taken into account too.

health and education but not things like defence). Finally LIMEW adds the value of household production. These measures are all per household. For LIMEW they also report equivalent median income; "equivalent" means that corrections are made for changing household size and composition. They estimated each of these income measures over the period 1959-2007 and for various sub-periods. Since measuring economic welfare over time is the objective, they convert each measure to real terms using the CPI and consider the median household values.

The growth rates of these four concepts of household income appear in lines 1-4 of Table 1. Let's concentrate on the last column initially, which shows growth over the whole period 1959-2007. The big point to take away is that median LIMEW grew at 0.67%, and equivalent median LIMEW at 1.01% p.a. Furthermore if we look at the sub-periods in the table we can see that there is no sign of a slowdown, except perhaps in 2004-2007. The period 1959-1972, supposedly the golden age of economic growth, was actually a comparatively poor one for households. Far and away the best period for households was 1982-1989 which coincides roughly with the Reagan presidency if we are allowed to ignore 1980-1981, the Volcker deflation and recession.[7]

The second big point to take away from Table 1 is that all these measures grew much less rapidly than GDP per capita, shown in Line 9, which grew at 2.18% p.a. over this period. None of the household measures grew at anything like this rate, e.g. their preferred measure, median LIMEW, grew at only 0.67% p.a. What accounts for this huge gap? Wolff *et al.* don't discuss this much but here is my explanation:

- Household size and composition have been changing: there are fewer children and more single households (Gordon, 2009). Hence equivalent median LIMEW grew faster than median LIMEW by some 0.34% p.a. (the same household income is spread over a smaller number of people).

- If the distribution of income had stayed the same, then mean LIMEW would have grown at the same rate as the median. In fact, the mean grew faster than the median by 0.30% p.a. Equivalent mean LIMEW (line 6) therefore grew, I estimate, at 1.31% p.a.

---

[7]   GDP per capita was 2.8% below its 1979 level in 1982 which helps to explain some of the rapid growth after 1982. GDP per capita grew at 2.43% p.a. over 1980-88, still faster than any sub-period except 1959-72.

- LIMEW is deflated by the CPI while GDP is deflated by the GDP deflator (more precisely, each component of GDP is deflated by its own price index). It so happens that the CPI grew more rapidly than the GDP deflator: the difference was 0.45% p.a. over 1959-2007 (line 10). Employing the GDP deflator rather than the CPI raises the growth of equivalent mean LIMEW to 1.76% p.a. (line 7). Arguably it would be better to use the price index for personal consumption expenditure (PCE) from the U.S. NIPA as a deflator. Methodologically, the PCE is superior to the CPI since it is an annually chained Fisher index while the CPI is a biannually chained Laspeyres.[8] Line 8 shows that the result would then have been much the same as deflating by the GDP deflator.

- Much of the remaining gap between median LIMEW and GDP per capita can probably be explained by two factors. First, investment has grown faster than consumption over this period, pulling up GDP in relation to consumption.[9] Second, household production is included in LIMEW but not in GDP: household production grows slowly because by assumption there is zero technical progress. These factors may account for the remaining 0.42% p.a. of the difference between the growth rates of median LIMEW and GDP per capita over 1959-2007.

These remarks are not meant to suggest that GDP per capita is a better measure of welfare than (equivalent) median LIMEW, but rather to explain how there can be such a large difference between the growth rates of the two.

The conclusion is that the median US household has gained significantly from economic growth since 1959. This remains the case even though the median household would have gained more (to the extent of 0.30% p.a.) if inequality had not widened. However most of the gap between the growth of GDP per capita and of median LIMEW is not due to rising inequality but to the other factors detailed above. Furthermore, and contrary to the common view, there were large gains in the 1980s which continued, albeit at a slower rate, in the 1990s and even into the 2000s.

---

[8] McCully et al. (2007) show that over 2002Q1-2007Q2 almost half of the 0.4 percentage point difference between the two deflators in annual growth rates was explained by the formula effect; most of the rest was explained by differences in relative weights due to the use of different surveys.

[9] This is probably because the prices of investment goods have been falling in relation to consumption goods, i.e. technical progress has been more rapid in investment goods. To keep the capital output-ratio constant in current price terms, investment has to grow faster than consumption in steady state (Oulton 2012a).

The little bit of analysis above is an attempt to show how while still making use of the SNA one can move "beyond GDP" to explain how household welfare relates to GDP. The main point to take away is that rising inequality has certainly reduced the gains form higher productivity which would otherwise have accrued to the typical US household but has not eliminated these gains completely.

This analysis stops in 2007, the last year of the boom. The median household has certainly done worse during the Great Recession and its aftermath, mainly because of lower productivity growth and declining labour force participation. Whether these adverse headwinds will continue to operate is an important question which will have to be left open for the moment.

Should GDP be adjusted for inequality?

There have been a number of suggestions for discarding GDP in favour of a measure which takes explicit account of inequality. One of the best-known is the measure based on the Atkinson index of inequality (Atkinson, 1970):

$$Z = \left(\frac{1}{N} \sum_{i=1}^{N} y_i^{1-\varepsilon} \right)^{1/(1-\varepsilon)} , \ 0 \leq \varepsilon < 1$$

where y_i is the income of the *i*-th person (or household), N is the number of people (or households) and ε is a parameter measuring "inequality aversion". If $\varepsilon = 0$ then society cares nothing for inequality in which case the Atkinson measure reduces to GDP per capita (or per household).

In the standard treatment of which the Atkinson index is an example, inequality is *per se* bad, though people may differ in the extent to which they are inequality averse. I would argue that our moral intuitions about inequality are too complex to be wholly captured by this formulation. In particular the crucial issue of desert is omitted. If the Atkinson/Sen approach were the whole story, then social welfare would be raised by abolishing two institutions (amongst others): the national lotteries run in many countries and the Nobel prizes. Both increase inequality unambiguously. Indeed Nobel prizes must be the most unequally distributed of all forms of income: only a dozen or so individuals receive one each year out of a world population of some 7.5 billion. Nobel prizes could be justified on Rawlsian grounds:

monetary incentives are needed to induce the effort required to make discoveries which benefit everyone, including the worst off. But suppose that it could be conclusively shown that the monetary rewards are not necessary, and that the prize winners (and their less successful colleagues) would have expended the same effort in exchange for just the honour and glory alone? I suspect that most people would still be quite happy to see the winners receive a monetary reward, even if it was not economically required. This is because they are perceived to deserve it. With national lotteries a different form of desert comes into play. In the UK version some winners receive £20 million pounds or more and in one sense no-one is worth this amount. But anyone can buy a lottery ticket and so long as the lottery process is perceived as fair most people are quite happy with the outcome.

Desert is a complex issue and it may be that people's views are not entirely consistent. Who gets the money and for what may well make a difference. The large rewards paid to professional footballers are seen by most people as justified (as long as they are playing well) but not the similar-sized rewards paid to bankers, especially after the global financial crisis.

Then there is the issue of redistribution, particularly welfare payments. Here it is obvious that notions of desert play a major role in most people's thinking. Paying welfare benefits to a former soldier with post-traumatic stress disorder may well be seen as one thing, paying the same amount to a drug addict with addiction-induced mental health problems may seem quite another. Whether what I take to be common moral perceptions can be justified philosophically is not the point here. The point is that they exist and in a democracy they should be taken into account.

In summary, it is not clear that the Atkinson index would meet with universal approval, even setting aside the issue of varying "taste" for inequality (the parameter ε). There is certainly a case for developing an index which takes explicit account of inequality as does the Atkinson index. But fortunately we do not need to choose between GDP and the Atkinson index (or any similar one). We are free to use and argue for both.

"Raising GDP per capita is pointless as it doesn't make people any happier"

Surveys of well-being or happiness repeatedly show that within any given country at any point in time richer people report themselves to be happier than poorer people. But when the

same survey is repeated in the same country over time there is no rise in the average level of happiness despite the fact that per capita income has gone up. Most of the time series evidence is for the U.S. and this result is known as the Easterlin paradox.[10]

The commonest explanation for the paradox and the one suggested by Easterlin himself, is that, at least above a certain level of income, people care more about their relative position in the income scale than they do about their absolute position. They are motivated by envy and ideas of "keeping up with the Joneses", and also by the satisfaction obtained by looking down on the less successful, more than by the pure desire for material stuff. This explanation reconciles the cross-section and time series evidence. But it leaves the implication that stopping growth would have no effect on happiness. Also, more redistribution from rich to poor would raise overall happiness (provided it did not reduce GDP too much through adverse incentive effects).

I must admit that I am puzzled by these survey results, mainly because they are inconsistent with other facts about people's behaviour. First, one might ask, if people care mainly about their relative position, why has there been so much fuss about the financial crisis? After all, for most people in most countries the drop in income has been (on this view) trivially small, no more than five per cent, and furthermore it fell disproportionately on the rich (at least initially). Second, if people care about their relative position, why does this have to be expressed in terms of annual income? After all, most workers today can work part time if they want to. So why can't A boast that his daily rate of pay is higher than B's even if B's *annual* earnings are higher because smart A works only 3 days a week while poor dumb B, a slave to the rat race, works five? Also surveys of part time workers regularly show that many would like to work longer hours if only they could. It is true that some leisure activities like skiing require a lot of complementary expenditure on expensive equipment but many other activities such as watching TV, surfing the Internet, or chatting with friends in pubs or cafés, do not.

In fact, people's leisure choices provide powerful evidence against the view that only relative position matters. The Classical economists argued that the amount of time people were prepared to work depended on the range of goods and services available for consumption.

---

[10]   Easterlin (1973). The time series evidence has been disputed by Stevenson and Wolfers (2008).

This was the basis for Adam Smith's "vent for surplus" theory of international trade, which was elaborated by John Stuart Mill (1871, Book III, chapter XVII):

"A people may be in a quiescent, indolent, uncultivated state, with all their tastes either fully satisfied or entirely undeveloped, and they may fail to put forth the whole of their productive energies for want of any sufficient object of desire. The opening of a foreign trade, by making them acquainted with new objects, or tempting them by the easier acquisition of things which they had not previously thought attainable, sometimes works a sort of industrial revolution in a country whose resources were previously undeveloped for want of energy and ambition in the people: inducing those who were satisfied with scanty comforts and little work, to work harder for the gratification of their new tastes, and even to save, and accumulate capital, for the still more complete satisfaction of those tastes at a future time."

Let's perform a simple thought experiment. Imagine that over the 220 or so years since the Industrial Revolution began in Britain *process* innovation has taken place at the historically observed rate but that there has been no *product* innovation in *consumer* goods (though I allow product innovation in *capital* goods). The UK's GDP per capita has risen by a factor of about 12 since 1800.[11] So people today would have potentially vastly higher incomes than they did then. But they can only spend these incomes on the consumer goods and services that were available in 1800. In those days most consumer expenditure was on food (at least 60% of the typical family budget), heat (wood or coal), lighting (candles) and clothing (mostly made from wool or leather). Luxuries like horse-drawn carriages were available to the rich and would now in this imaginary world be available to everyone. But there would be no cars, refrigerators, washing machines, dishwashers or smartphones, no radio, cinema, TV or Internet, no rail or air travel, and no modern health care (e.g. no antibiotics or antiseptics). How many hours a week, how many weeks a year and how many years out of the expected lifetime would the average person be willing to work? My guess is that in this imaginary world people would work a lot less and take a lot more leisure than do real people today. After all, most consumer expenditure nowadays goes on products which were not available in 1800 and a lot on products not invented even by 1950.[12]

---

[11]   Source: spreadsheet accompanying Maddison (2003).
[12]   Only about a tenth of the family budget goes on food nowadays and even within the food basket many items were not available in 1800.

Overall the proportion of time devoted to market work has not changed much in the last century or so, though this masks differences between women whose contribution has been rising while that of men has been falling. But the rough constancy of the labour/leisure choice may be a bit of an accident, produced by a kind of battle between product and process innovation. There is no guarantee that this constancy will persist. If consumer product innovation falters then I would expect leisure to rise. Of course other factors are at work here too: increased longevity, itself probably a product of economic growth, is generating pressure for increased work effort.

In summary, people's choice between labour and leisure demonstrates that they value higher consumption in an absolute and not just a relative sense. So rising GDP per capita would be in accordance with people's desires and preferences. Philosophers and social critics may object that the average person's desires and preferences are trivial, ill-informed and misguided (an attitude which can be traced back at least as far as Plato's *Republic*), but policy should take people as they are, not as others would have them.

Conclusion: not fade away?

The thought experiment just discussed suggests another one. Assume that technical progress continues to raise labour productivity over the next century at something like the rate experienced in the last 100 years. Will the typical consumer in Western societies take the benefits in the form of ever-increasing leisure? If so, consumers would be increasingly satiated with the goods and services which GDP measures. So in this era of material abundance GDP might come to seem not wrong but increasingly irrelevant. Such societies would probably not lack for problems due to the uses to which some people might put their ever more abundant leisure. But the analysis of such problems would not be much helped by the GDP statistics.

This second thought experiment envisages the same scenario as the first: no new consumer goods or services. We know that the two centuries since 1800 have seen an enormous variety of new consumer goods invented and made available on the market. It seems to me very unlikely that this inventiveness will simply come to a dead halt in the foreseeable future. So I expect new consumer goods to appear in a steady stream. On this count alone GDP and the

SNA will continue to be useful. Also much of the rest of the world outside the magic circle of Western societies remains poor. Today's poorer countries will likely retain an interest in GDP for many decades to come.

Throughout its more than 60 year official life the SNA has expanded to address new concerns. The "core" SNA is now buttressed by satellite accounts covering interactions between the economy and the environment and household activities. I expect this process to continue and deepen as international discussions proceed towards agreeing a successor to the 2008 SNA.

References

Advisory Commission to Study the Consumer Price Index [The Boskin Commission] (1996). *Towards a More Accurate Measure of the Cost of Living.* Final Report to the Committee on Finance of the U.S. Senate. Washington, DC. [http://www.ssa.gov/history/reports/boskinrpt.html]

Atkinson, Anthony B. (1970). "On the measurement of economic inequality". *Journal of Economic Theory*, 2 (3), pages 244–263.

Bean, Charles. (2016). *Independent Review of Economic Statistics*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/29049 36_Bean_Review_Web_Accessible.pdf.

Broadberry, Stephen, Bruce M.S. Campbell. Alexander Klein, Mark Overton, Bas van Leeuwen. (2016). *British Economic Growth, 1270-1870*.

Byrne, D.M., J.G. Fernald and M.B. Reinsdorf (2016). Does the United States have a productivity slowdown or a measurement problem?". *Brookings Papers on Economic Activity*, Spring, pp. 109-174.

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

Deaton, Angus (2013). *The Great Escape: Health, Wealth and the Origins of Inequality*. Princeton: Princeton University Press.

Deaton, Angus and Bettina Aten (2017). "Trying to understand PPPs in ICP 2011: why are the results so different?" *American Economic Journal: Macroeconomics*, vol. 9(1), pages 243-264.

Easterlin, Richard A. (1973). "Does money buy happiness?". *The Public* Interest, vol. 30, pages 3-10.

Edin, Kathryn J. and H. Luke Shaefer (2015). *$2.00 a Day: Living on Almost Nothing in America.* Houghton Mifflin Harcourt: New York.

European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, World Bank (2009). *System of National Accounts 2008.* New York. (the latest version of the SNA).

Feenstra, Robert C. (1994) "New product varieties and the measurement of international prices". *American Economic Review*, vol. 84(1), pages 157-177.

Fogel, Robert William (2004). *The Escape from Premature Hunger and Death, 1700-2100: Europe, America and the Third World*. Cambridge: Cambridge University Press.

Gordon, Robert J. (1990). *The Measurement of Durable Goods Prices.* Chicago: University of Chicago Press for NBER.

Gordon, Robert J. (2009). "Misperceptions about the magnitude and timing of changes in American income inequality". NBER Working Paper 15351. [http://www.nber.org/papers/w15351].

Gordon, Robert J. (2016). *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War.* Princeton: Princeton University Press.

Groshen, Erica L., Brian C. Moyer, Ana M. Aizcorbe, Ralph Bradley and David Friedman (2017). How government statistics adjust for potential biases from quality and new goods in an age of digital technologies: a view from the trenches". *Journal of Economic Perspectives*, Vol. 31, Number 2 (Spring), pages 187-210.

Hausman, Jerry (1997). "Valuation of new goods under perfect and imperfect competition". Chapter 4 in *The Economics of New Goods*, edited by Timothy F. Bresnahan and Robert J. Gordon.

Hausman, Jerry (2003). "Sources of bias and solutions to bias in the Consumer Price Index". *Journal of Economic Perspectives*, 17(1), Winter, pp. 23-44.

Hicks, John R. (1940). "The valuation of the social income". *Economica*, vol. 7(2), pp. 105-124.

Jerven, Morten (2013). *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It*. Ithaca and London: Cornell University Press.

Jones, Charles I. and Peter J. Klenow (2016). "Beyond GDP? Welfare across countries and time". *American Economic Review*, vol. 106(9), pages 2426-2457.

Jorgenson, D.W. (1989), "Capital as a factor of production", in D.W. Jorgenson, and R. Landau, eds., *Technology and Capital Formation*, The MIT Press, Cambridge, MA.

Jorgenson, D. W. and Z. Griliches (1967). 'The explanation of productivity change'. *Review of Economic Studies*, vol. 34, pages 249-83.

Jorgenson, D.W., F.M. Gollop, and B.M. Fraumeni. (1987). *Productivity and U.S. Economic Growth*. Cambridge, MA: Harvard University Press.

Kuznets, Simon (1959). *Six Lectures on Economic Growth*. The Free Press of Glencoe.

Kuznets, Simon (1966). *Modern Economic Growth: Rate, Structure, and Spread.* Yale University Press: New Haven and London.

Lee, Pete (2011). *UK National Accounts: A Short Guide.* Office for National Statistics. [http://www.ons.gov.uk/ons/rel/naa1-rd/national-accounts-concepts--sources-and-methods/august-2011/index.html].

McCully,Clinton P., Brian C. Moyer and Kenneth J. Stewart (2007). "A reconciliation between the Consumer Price Index and the Personal Consumption Expenditure Price Index". [http://www.bea.gov/papers/pdf/cpi_pce.pdf].

Maddison, Angus (2003). *The World Economy: Historical Statistics*. Paris: OECD.

Mill, John. Stuart. (1871) [1909]. *Principles of Political Economy*. 7th edition. London: Longmans, Green and Co.

OECD (2001b). *OECD Productivity Manual: A Guide to the Measurement of Industry-Level and Aggregate Productivity Growth.* Paris: OECD.

OECD (2009), *Measuring Capital: OECD Manual 2009: 2nd Edition*, OECD, Paris. Available at http://www.oecd.org/dataoecd/16/16/43734711.pdf.

Oulton, Nicholas (2012a). "Hooray for GDP". Centre for Economic Performance, Occasional Paper no. 30.   [http://cep.lse.ac.uk/pubs/download/occasional/op030.pdf].

Oulton, Nicholas (2012b). "Long term implications of the ICT revolution: applying the lessons of growth accounting and growth theory". *Economic Modelling*, vol. 29, 1722-1736.

Oulton, Nicholas (2015). "Understanding the space-time (in)consistency of the national accounts", *Economics Letters*, vol. 132.

Redding, Stephen J. and David E. Weinstein (2016). "A unified approach to estimating demand and welfare". NBER Working Paper No. 22479.

Stevenson, B. and J. Wolfers (2008). "Economic growth and subjective well-being: reassessing the Easterlin paradox." *Brookings Papers on Economic Activity,* Spring. Retrieved from *www.gallup.com/poll/File/116605/EasterlinParadox.pdf* .

Stiglitz, Joseph E., Amartya Sen and Jean-Paul Fitoussi (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris. [www.stiglitz-sen-fitoussi.fr/].

Syverson, Chad (2017). "Challenges to mismeasurement explanations for the US productivity slowdown". *Journal of Economic Perspectives*, vol. 31(2), Spring, pages 165-186.

Timmer, M.P., R. Inklaar, M. O'Mahony and B van Ark. (2010). *Economic Growth in Europe: A Comparative Industry Perspective.* Cambridge: Cambridge University Press.

Wasshausen, Dave and Brent R. Moulton (2006). "The role of hedonic methods in measuring real GDP in the United States". 31st CEIES Seminar. https://www.bea.gov/papers/pdf/hedonicGDP.pdf.

Wolff, Edward N., Ajit Zacharias and Thomas Masterson (2012). "Trends in American living standards and inequality". *Review of Income and Wealth*, series 58, number 2, June, pages 197-232.

World Bank Group (2008). *Global Purchasing Power Parities and Real Expenditures: 2005 International Comparison Program*. Available at http://siteresources.worldbank.org/ICPINT/Resources/icp-final.pdf.

World Bank Group (2015). *Purchasing Power Parities and the Real Size of World Economies*. [http://siteresources.worldbank.org/ICPEXT/Resources/ICP_2011.html]

Table 1
Real income measures, per capita and per household, in the United States:
annual percentage rates of growth, 1959-2007

	1959-1972	*1972-1982*	*1982-1989*	*1989-2000*	*2000-2004*	*2004-2007*	*1959-2007*
Deflated by CPI-U							
1. Median CDI	1.22	-0.29	2.16	0.88	0.62	0.16	0.85
2. Median PFI	1.55	-0.38	2.16	1.00	0.69	0.65	0.98
3. Median LIMEW	0.36	-0.68	2.82	0.93	0.96	0.22	0.67
4. Equivalent median LIMEW	0.94	-0.13	3.22	0.97	0.84	0.42	1.01
5. Mean LIMEW	0.53	-0.41	2.87	1.90	0.22	0.73	0.97
6. Equivalent mean LIMEW	1.11	0.14	3.27	1.94	0.10	0.93	1.31
Deflated by GDP or PCE deflator							
7. Equivalent mean LIMEW (deflated by GDP deflator)	1.02	1.26	3.64	2.74	0.25	0.94	1.76
8. Equivalent mean LIMEW (deflated by PCE deflator)	1.35	1.16	3.25	2.55	0.47	1.29	1.77
9. GDP per capita	2.73	1.34	3.37	2.03	1.26	1.58	2.18
Memo items							
10. CPI-U deflator less GDP deflator	-0.09	1.12	0.37	0.80	0.14	0.00	0.45
11. PCE deflator less GDP deflator	-0.32	0.10	0.39	0.19	-0.22	-0.36	-0.01
12. CPI-U deflator less PCE deflator	0.23	1.02	-0.02	0.61	0.37	0.36	0.46

Table 1, continued

Notes
CDI: Comprehensive Disposable Income. CDI equals LIMEW less the value of household production and public individual consumption, per household.
PFI: Post Fiscal Income. PFI equals LIMEW less the value of household production, per household.
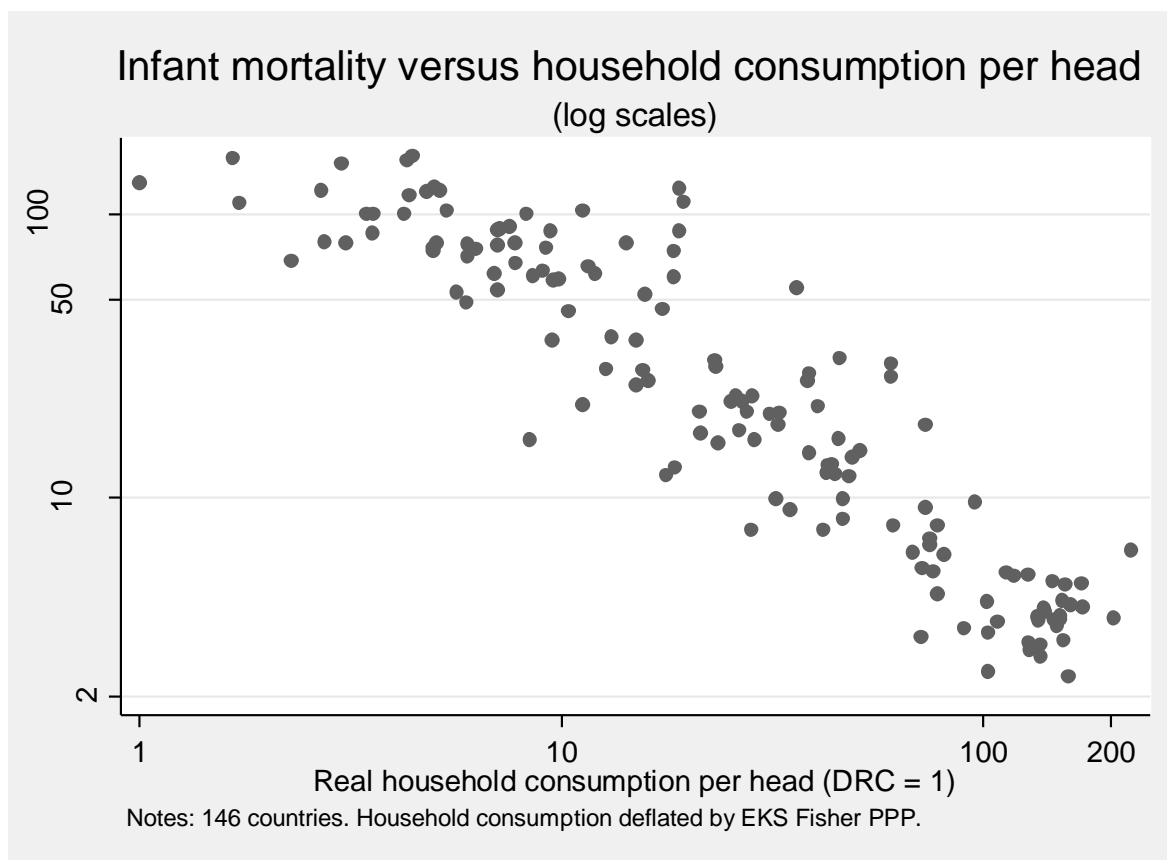LIMEW: Levy Institute Measure of Economic Well-Being, which is income less taxes plus cash and non-cash benefits plus individual public consumption plus household production, with property income valued on an annuity basis, per household.
Equivalent median LIMEW: median LIMEW per equivalent household, i.e. after adjusting for household size and composition.
Equivalent mean LIMEW: calculated as growth of *equivalent median* LIMEW plus growth of *mean* LIMEW minus growth of *median* LIMEW.
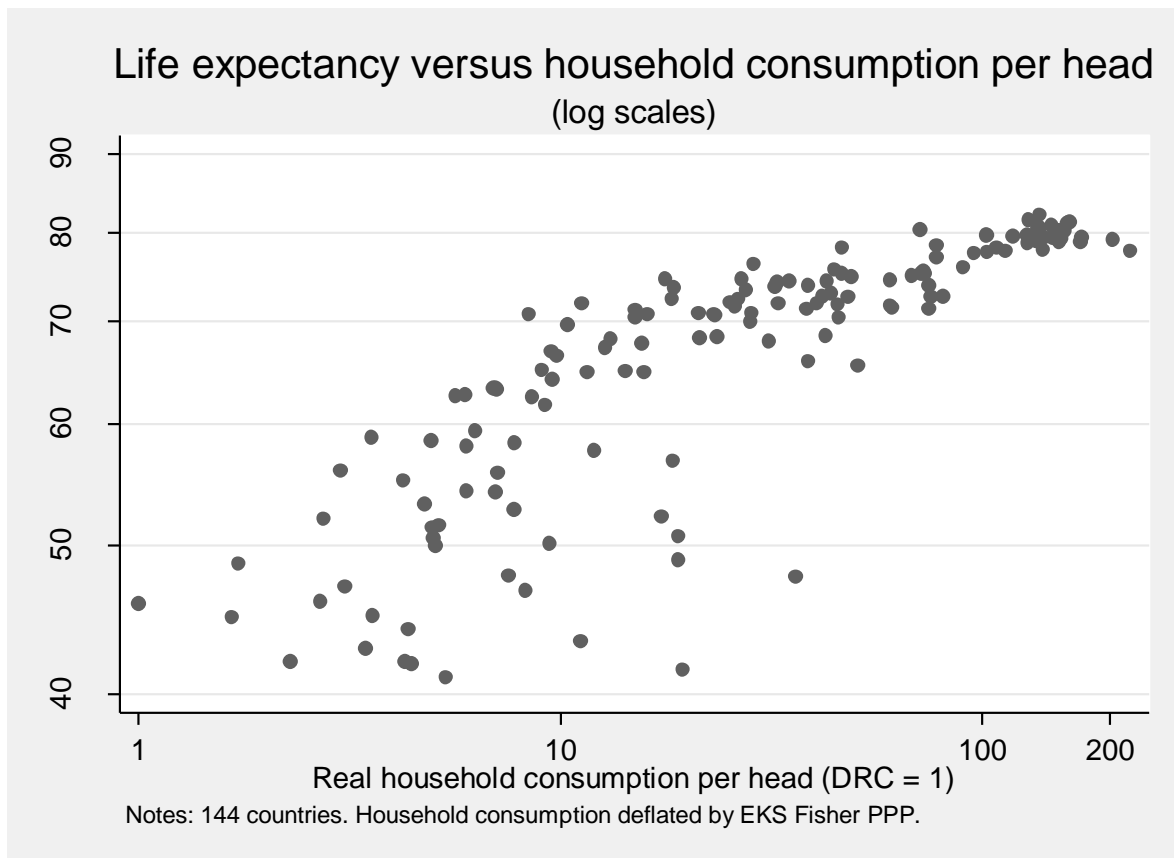In lines 1-6 the deflator is the CPI-U. GDP per capita (line 9) is deflated by the GDP deflator.
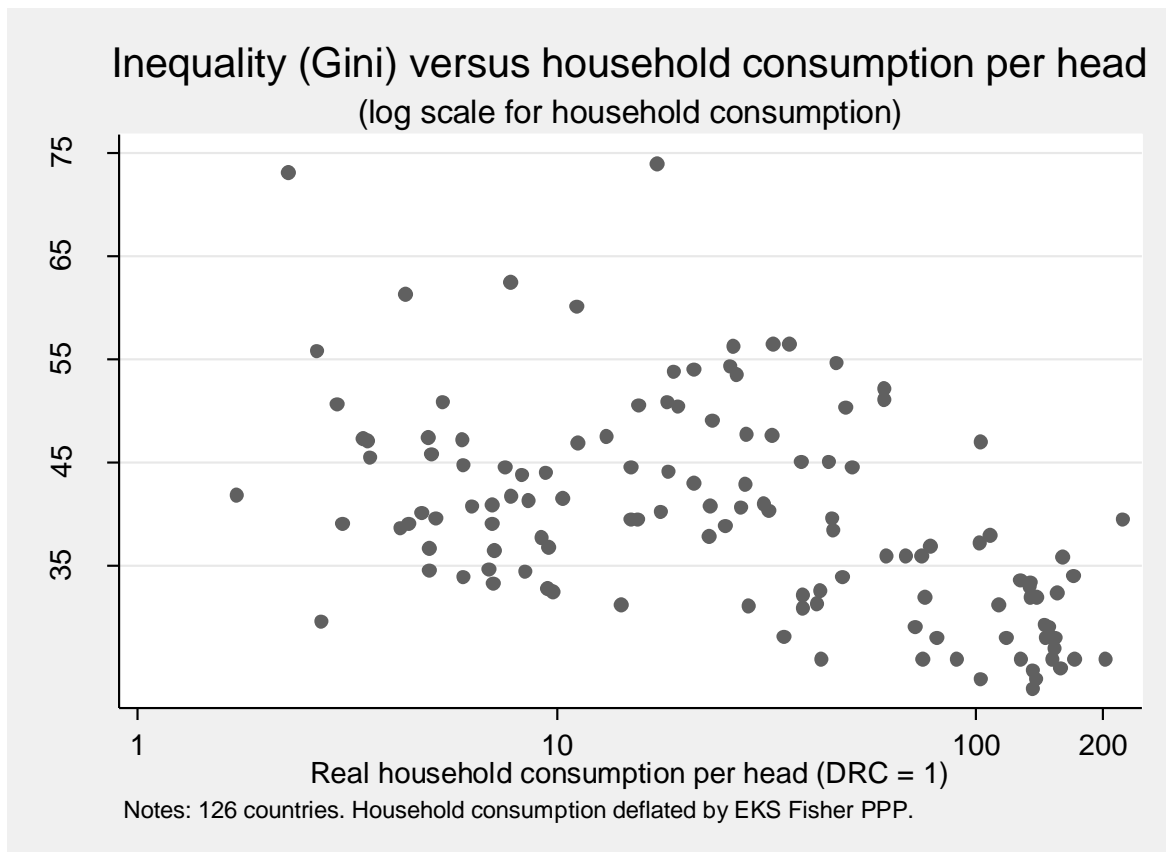
Chart 1

Infant mortality versus household consumption per head
(log scales)

Real household consumption per head (DRC = 1)

Notes: 146 countries. Household consumption deflated by EKS Fisher PPP.

Source: Oulton (2012a).

Chart 2

Life expectancy versus household consumption per head
(log scales)

Notes: 144 countries. Household consumption deflated by EKS Fisher PPP.

Source: Oulton (2012a).

Chart 3

Inequality (Gini) versus household consumption per head
(log scale for household consumption)

Real household consumption per head (DRC = 1)

Notes: 126 countries. Household consumption deflated by EKS Fisher PPP.

Source: Oulton (2012a).