CHAPTER 7

National accounts in world history

Methodological problems and possible solutions

Rodney Edvinsson

Did the European economy overtake that of China as early as the fifteenth century, or was China more advanced economically than most of Europe up until the early nineteenth century? To answer these questions, historical national accounts of all the countries and regions in the world must be constructed leading back to the Middle Ages. In the last ten years, major progress in this direction has been made (Broadberry & Gupta 2009; Broadberry, Campbell, Klein, Overton & van Leeuwen 2010; Bassino, Broadberry, Fukao, Gupta & Takashima 2011; Broadberry, Guan & Li 2012). The main pioneering work in this field was carried out by Angus Maddison (2010). In his database he has extended the GDP series for all countries back to the year 1 CE.

Despite the progress made, there are a number of problems with various estimates concerning several countries. Co-ordination among researchers and comparisons between countries are still not fully developed, and there is a lack of international standards for historical national accounts. Gregory Clark (2009) remarks that:

> All the numbers Maddison estimates for the years before 1820 are fictions, as real as the relics peddled around Europe in the Middle Ages. Many of the numbers for the years 1820, 1870, and 1913 are equally fictive.

Despite this, Maddison's database has gained wide currency among economists analyzing economic growth. His books are among the most quoted among social scientists, historians and economists in Google Scholar. Clark continues:

> Just as in the Middle Ages, there was a ready market for holy relics to lend prestige to the cathedrals and shrines of Europe – Charlemagne secured for the cathedral in Aachen, his capital, the cloak of the Blessed Virgin, and the swaddling cloths of the infant Jesus – so among modern economists there is a hunger by the credulous for numbers, any numbers however dubious their provenance, to lend support to the model of the moment. Maddison supplies that market.

Since Maddison's death in 2010, a project has come into existence to update his database, but it also reproduces the problematic methods applied by Maddison, for example, his method to use the purchasing power parities of 1990 as a benchmark for all other periods.

The main concern of national accounts is how to measure different aspects of the production and distribution process. In the general debate, the impression is often given that the value of aggregate production, often taken as synonymous with GDP, is unequivocal once you have reliable sources. Sometimes the data is taken for granted even if it is not based on reliable sources at all. Aggregate production can, however, be calculated using different methods and definitions, which can lead to quite divergent interpretations of economic development. GDP is a controversial measure from many points of view. There are also different methods to calculate GDP that can give quite different results (Inter-Secretariat Working Group on National Accounts 1993: 14).

The constant price fallacy

Volume or real value is a kind of constant price estimate. The effect of different price levels must be eliminated when volume growth or volume values are measured. In national accounts, this is achieved by making a comparison in constant prices of the production at two different points in time or in two countries. The nominal series is deflated by a price index to arrive at volume values. The difficulty does not end there, since the question still remaining is which constant prices and index formulas to use. Various techniques to eliminate the inflation component result in different conclusions concerning economic growth and the relative standing of various

countries. This drawback is often glossed over by economists. Part of the problem is theoretical in origin. Many theories of economic growth, mostly of neoclassical origin, use a one-commodity model. In international comparisons the focus is on aggregate GDP, often disregarding comparison of the various components of GDP. Real economies, however, consist of many goods and services.

A simple example can illustrate the deflation procedure. Take an economy that in year 1 produces one billion tons of apples and in year 2 one billion barrels of oil. Has the economy experienced a positive, negative or zero growth rate in constant prices? This depends on how a barrel of oil is valued in comparison to a ton of apples, i.e. on the relative prices of oil versus apples. If a barrel of oil is valued at more than a ton of apples, the economy has experienced positive growth. If a barrel of oil is valued at less than a ton of apples, the economy has experienced negative growth. Finally, if a barrel of oil and a ton of apples are valued equally, the economy has experienced zero growth.

A further issue is which year's relative prices should be used. In the above example, assume that in year 1 a barrel of oil is worth more than a ton of apples, and in year 2 a barrel of oil is worth less than a ton of apples. In the prices of year 1 there has then been positive economic growth, while in the prices of year 2 there has been negative economic growth. The first is called a Laspeyres volume index, while the second is called a Paasche volume index (for both indices, the first year is here considered the base year, and the second the compared year). The difference between the two indices is often small, but can accumulate to large differences over time. Over longer periods, a Laspeyres volume index tends to display a higher growth rate than a Paasche volume index if the base year is earlier in time than the compared year, the so-called Gerschenkron effect (Jonas & Sardy 1970: 83). The Gerschenkron effect arises when activities whose relative prices are falling tend to increase their volume shares of total production and vice versa. This is what happened during the industrial revolution. Manufacturing expanded its volume share of GDP at the same time as the relative prices of manufactured goods decreased. This was due to faster increases in productivity than for other sectors.

If the Laspeyres volume index displays just a 0.3 per cent higher growth per year than the Paasche volume index, over 200 years this amounts to 82 per cent. Comparing the relative development of two countries that use different index formulas over a longer period of time can have huge consequences. Maddison's database consists of such series, and the difficulty has been ignored or glossed over by many economists and social scientists who use his data. Some researchers have argued for the use of the geometric average of the two indices – a so-called Fisher price index – or equivalent formulas, but this has seldom been adopted in historical national accounts.

In poor countries, the prices of goods and services for domestic consumption are, in general, lower than in rich countries. In addition, exchange rates tend to fluctuate considerably. Therefore, when the GDP or GDP per capita of various countries are compared, Purchasing Power Parities are constructed to eliminate differences in price levels, which is equivalent to the comparison of a country over two time periods. These parities are expressed in national currency units per United States dollar (similar to the exchange rate). The relative price level in one country is the ratio of the purchasing power parity to the exchange rate. Even if the prices of many goods did not converge internationally until the nineteenth century (O'Rourke & Williamson 2002), the price differences between countries today for many services may be as large or even larger. The problem of different prices is as much a problem today as it is for the reconstruction of historical national accounts of the preindustrial period.

A difficulty with purchasing power-adjusted GDP is that it can be computed in different ways. PPP-adjustment removes the difference in the absolute price level, but not the relative price differences. In addition, there are two ways to compare countries' GDP per capita over time: the use of current and constant PPPs.

Current purchasing power parity entails two countries' GDP for one year being compared in that year's PPPs. This means that the new PPPs have to be calculated each year, and this method is very time-consuming.

The other method entails PPPs only being calculated for one year, the benchmark year. When comparisons are made for earlier or later years, they are based on the estimated economic growth

rates of the various countries. For example, assume that the GDP per capita of country *A* grows by 100 per cent between 1950 and 1990 and of country *B* by 200 per cent during the same period. If the two countries' GDP per capita is valued at 20,000 dollars in 1990's prices and PPPs (the benchmark year), country *A* had a GDP per capita of 10,000 dollars in 1950 and country *B* a GDP per capita of 5,000 dollars in 1990's constant PPPs. No current PPPs are needed for 1950 to make this calculation. This is basically the method applied by Angus Maddison in his comprehensive database. All the data back to the year 1 CE is expressed in 1990 so-called Geary-Khamis dollars.

There are drawbacks with both methods. Let us continue our example with oil and apples. Suppose that Norway's GDP consists of only one billion barrels of oil, and Sweden's of only one billion tons of apples. Which country has the highest purchasing power-adjusted GDP? It depends on the relative price of oil and apples. If a barrel of oil is valued at half as much as one ton of apples, Norway's real GDP is half that of Sweden. However, if world oil prices increase significantly in relation to apples, so that a barrel of oil is valued at twice as much as one ton of apples, Norway has twice the real GDP of Sweden (here we assume one world price, but if the relative prices are different in Sweden and Norway, using Sweden's or Norway's prices results in different PPPs as well, as in the example of growth from one year to the next). This would illustrate why, for example, Norway's purchasing power-adjusted GDP per capita fluctuates sharply from one year to another, despite the fact that Norway's real GDP per capita does not display any equivalent fluctuation. The explanation is the high share of oil production in Norway's GDP and the high volatility of the relative price of oil on the world markets.

Using constant PPPs has the advantage that annual fluctuations are reduced. However, we then have another dilemma pertaining to long-term growth and terms of trade. Assume that between 2010 and 2020 Norway increases its oil production by 100 per cent, from one billion to two billion barrels of oil, while Sweden's production remains stagnant at one billion tons of apples. In constant PPPs, Norway's GDP has doubled compared to Sweden's. However, assume that the price of apples remains the same, while the price

of oil declines by 75 per cent. The growth in Norway would then be counteracted by a decline in its terms of trade. In fact, while in constant PPPs Norway's GDP would double in comparison to Sweden, in current PPPs it would be halved compared to Sweden's, between 2010 and 2020. In the example, Norway became relatively poorer since it could buy fewer apples and less oil on the world market than Sweden, despite experiencing faster economic growth. Again, the differences are usually quite small for shorter periods. For longer periods, which are what matters for economic historians, the accumulated effects could be huge.

The method of constant PPPs is primarily used in international economic-historical research, such as in Maddison's database. Most historical reconstructions of GDP transform their figures into 1990 Geary-Khamis dollars, to make the new series comparable with Maddison's database. This entails comparisons of countries' volume GDP for earlier centuries being distorted and not being the same as if current PPPs had been used. If a country produces a large share of goods and services whose relative prices fall faster than for other goods and services, this country shows a higher per capita growth than other countries if the position compared to other countries does not change in current purchasing power parities. In this case, current PPPs provide a more accurate picture of a country's long-term economic development in relation to other countries, although the constant PPPs provide a more accurate picture of annual changes. The only way that Maddison's method could work is if the terms of trade of all countries were unchanged for longer periods of time, an assumption that cannot be supported theoretically or empirically.

Against this background, Prados de la Escosura (2000) argues that current purchasing power parities are preferable when examining how countries' relative positions change over time. Ward and Devereux (2003) show that estimates of current purchasing power parity entails the United States having had a per capita GDP that was well above that of the UK as early as the 1870s, while Maddison's method of constant purchasing power parity implies that the United States did not overtake Britain until around 1900. Broadberry (2003) argues that there are problems with Ward's and Devereux's calculations, and maintains that both methods should

be used. Recently, Broadberry and Klein (2012) presented the PPPs for the years 1905 and 1927. Their result is that Sweden's purchasing power-adjusted GDP per capita for the year 1905 is 10 per cent higher if 1905 year's purchasing power parity is used instead of 1990 international Geary-Khamis dollars.

The definitional dilemma

A great difficulty for international comparisons is that the historical national accounts of different countries are not based on the same methods and definitions. Alternative measures of production and volume growth are mostly ignored. Generally, historical national accounts attempt to follow modern national accounts. However, due to the constant revisions implemented by modern statistical offices, historical national accounts tend to use older methods and definitions. Some definitions of the production boundary are also specific to individual studies.

Certain assumptions and definitions are necessary for the reconstruction of historical national accounts and trends for economic growth. One problem is taking for granted that assumptions and definitions are based on a fully developed market economy. For example, assumptions that various assets, such as land and equipment, were used competitively even during the Middle Ages, can lead us to the wrong conclusion concerning the level of income these assets generate. Even today, much of production is not profit-maximizing.

One issue concerns which activities to include in aggregate production – i.e. where to put the "production boundaries". According to modern international guidelines for national accounts, a distinction is made between activities that are "productive in an economic sense" and those that are not. There are many inconsistencies in this respect when it comes to official national accounts. Many economic activities, mostly outside the market economy, are not included in the GDP – most importantly, unpaid domestic or voluntary work – while all goods produced for own use are to be included since a revision of international guidelines implemented in 1993, which is of particular importance for developing countries (United Nations et al. 2009: 99).

One issue in historical national accounts is whether manufacturing in homes is to be included in the measured GDP or not. For some countries this has significant repercussions on the estimate of industrial production. Modern guidelines today entail the inclusion of manufacturing in homes. Internationally, historical national accounts often calculate the approximate scope of manufacturing in the 19th century by following the intermediate consumption of raw materials. Indirectly, that entails the inclusion of manufacturing other than factory production and handicrafts (Bourguignon & Lévy-Leboyer 1990: 266; Grytten 2004: 249; Horlings et al. 2000: 37–45). However, the production boundary of SNA 2008 is not used consistently for all countries. For example, the historical national accounts of Finland include manufacturing in homes for sale on the market, but not for own use (Hjerppe 1996: 33). In Swedish national accounts two different definitions have been used, one including manufacturing in homes (Edvinsson 2013) and one excluding it (Edvinsson 2005; Schön & Krantz 2012). A problem with Maddison's database is that he mixes GDP data in which different definitions of the production boundary are applied.

One of the largest drawbacks in constructing national accounts is the reliance on price. This is especially problematic when valuing non-market production. In the agrarian societies production was mostly for self-use. At least during the early modern period a large part of it was traded. Prices exist for most goods and services. One important aspect with indices is that even if they have low validity, if they are measured consistently they may still be quite good indicators of economic growth and fluctuations. That is the main, although quite shaky, argument for using modern definitions of GDP for the preindustrial period as well. It is also reasonable to argue that it should be possible to compare production in one period to the level in a later period.

A major weakness is how to deal with products and services that are not priced, or where the pricing mechanism is distorted. These products and services can either be excluded or assigned a fictitious price tag. Historical national accounts deal with economies where market relations only affected a small part of production; but modern economies also consist of large sectors that are not priced. Although

the international guidelines recognize that unpaid household services are "productive in an economic sense", it is argued that the "inclusion of large non-monetary flows of this kind in the accounts together with monetary flows can obscure what is happening on markets and reduce the analytic usefulness of the data" (Inter-Secretariat Working Group on National Accounts 1993: 5), and that "there are typically no suitable market prices that can be used to value such services" (Inter-Secretariat Working Group on National Accounts 1993: 124). For historical national accounts this statement is particularly awkward, since the purpose of reconstructing a GDP series is to measure production rather than to provide data suitable for economic policy. Some researchers therefore argue that there are good reasons for historical national accounts to include unpaid household services (Jonsson 1997: 49). Especially for earlier times it is difficult to apply the distinction between "paid" and "unpaid" labor, since most of the production was for final self-use.

Different methodologies have been developed to measure the value of unpaid household work by putting a price tag on it (Edvinsson 2009). One method is to equal the value of these services to the labor input, utilizing the wage of paid domestic labor as an indicator. This method has been put into operation in *The National Income of Sweden* (Lindahl, Dahlgren & Kock 1937: 213–215) and later in Swedish historical national accounts (Krantz 1987: 17). However, doing this runs the risk of putting the wrong value on the actual work performed. A more appropriate method is to estimate the market output of these services (Nyberg 1995: 22–28). This also gives different results, depending on the indicators used. One solution may be to equal the value added per unpaid household working hour to the average value added per working hour within the market sector (Folbre & Wagman 1993: 285). Nevertheless, such a measure has little to do with how such services would actually be valued on the market, and it does not add any new information apart from that already provided by the estimates of unpaid household work in terms of actual working hours. The question of the labor productivity of unpaid household work in relation to market activities needs to be empirically investigated and not taken as given.

According to Anita Nyberg (1995: 25–27), the monetary estimates

of the value of unpaid household work in different industrialized countries vary between 30 and 60 per cent of GDP, which is quite sizable. It is likely that the proportion is even greater for earlier periods.

How do we measure changes in the level of production if we are dealing with an economy that does not know of any prices, such as a self-subsistence economy? Surely, it should be possible to construct volume indices for economic growth for such an economy as well. One solution is to use the relative prices of a modern economy, but such relative prices might be completely different from the relative valuations of the economy under study. Not all economies or activities are priced, but labor is the foundation of all human production.

An alternative to relative price may be to compare how much labor time it takes to produce two goods, i.e. to use relative labor times instead of relative prices to construct a volume index, accounting for the labor time embodied in intermediate consumption as well. Rickard Warlenius discusses in this book how embodied labor can be used to analyze flows at the international level. In studies of hunter-gatherer societies, where there are neither money nor prices, anthropologists often use the number of hours spent on different activities of the total worked per week to describe the economic structure of these societies (Cashdan 1989: 23). The labor time is a cost, in terms of foregone free time. Such a volume index is, in contrast to the usual one applied in national accounts, completely independent of price relations. If prices are proportional to labor values, then this type of volume index gives exactly the same result as the volume index based on relative prices. But when prices and labor values diverge, the labor value volume index favors activities that have a low value, reckoned per working hour, if it is assumed that labor productivity is the same as in other types of activities such as, for example, government and household services. It is of course desirable that the productivity differences between laborers should be considered. However, without information on price, such comparisons can only be made if similar types of product are considered. Table 1 presents an example of using labor productivities as an alternative to prices when weighting quantities produced. Imagine an economy *A*, consisting of one household (one woman and one

man), with a strict, gendered division of labor. The woman and the man each work 8 hours a day. In total there would, therefore, be 16 working hours per day. The woman produces 8 kilograms of grain, while the man produces 8 kilograms of meat. This means that the productivity of economy *A* is 1 kilogram of grain per working hour and 1 kilogram of meat per working hour. However, because of gender discrimination and due to the strict, gendered division of labor, women's work is valued less than men's. In our example, the price of grain is set to 1 dollar per kilogram and for meat to 2 dollars per kilogram. This means that the total daily work of the woman is valued at 8 dollars, while that of the man is valued at as much as 16 dollars, in total 24 dollars. It is also possible to make this calculation if no prices are known for economy *A* (i.e. if there is complete self-sufficiency), if we apply the prices of a market economy at about the same level of development.

Imagine now that we observe another economy, *B*, which produces 16 kilograms of grain and 4 kilograms of meat. We may not know the productivity or the prices of this economy. Economy *B* may, for example, be an adjacent household or the same household at another period in time. To estimate the volume of production of economy *B* compared to economy *A* we can use the usual method based on constant prices. In the prices of economy *A*, economy *B* is valued at 20 dollars, i.e. one sixth less than the value of economy *A*. However, calculating how many hours it would take for the economy *A* to achieve the production of economy *B*, we arrive at the figure 20 hours, which is 25 per cent longer labor time than in economy *A*. The difference arises because in the prices of economy *A* men's work is valued more than women's. Both indices of production are Laspeyres volume indices, if the economy *A* is considered the base period. The Paasche volume indices, which use the prices and labor productivities of economy *B*, produce other results that are, however, not presented in the table.

Under perfect market conditions, if men and women were to produce the same amount of grain and meat per hour, then there would be no price difference between grain and meat. In reality, gender discrimination would probably come into play, preventing women from doing men's work and vice versa. Furthermore, we may

not know the prices of economy *A* and *B*, if both were self-sufficient. In that case only a volume index based on labor productivity should be calculated, since using the prices of a third economy could be misleading (the possible discrimination of the third economy may not apply to economies *A* and *B*).

Table 1. Illustration of how to compute a volume index to compare two economies using prices and labor productivity respectively.

	Production	Productivity	Price	Volume index, prices of economy A	Volume index, productivity of economy A
Economy *A*	8 kg grain (women) 8 kg meat (men)	1 kg grain per hour 1 kg meat per hour	1 dollar per kg grain 2 dollars per kg meat	24 dollars	16 hours
Economy *B*	16 kg grain 4 kg meat			20 dollars	20 hours

Errors and lack of transparence

The empirical material of historical national accounts consists of both primary and secondary sources. These sources do not use the same classifications and definitions. Breaks often occur when different time series for the same variable but different periods are compared with each other. Modern standards for national accounts change constantly, which requires retrospective revisions ("Utredningen om översyn av den ekonomiska statistiken" 2002: 18–23). Such revisions are made regularly by the statistical offices, but unfortunately, often only for an insufficient number of years, which creates new breaks in the series. When time series are linked with each other, the figures of the original time series are changed, and there is a risk that the linked time series will provide an inadequate picture of the actual values or levels.

Some of the difficulties when constructing historical national accounts stem from the status of doing this type of research. Often it

does not fit the format of journal articles, which means that writing extensive documentation of how data has been constructed is not very rewarding for individual researchers. This dilemma is shared with other fields of historical statistical reconstruction. Even so, historical national accounts often have better documentation than that of official statistics. One solution would be to redefine historical statistics as a separate academic field, and to create new academic journals specializing in this field. These could then also publish detailed documentations.

Since historical national accounts deal with quite extensive quantities of material, it is almost inevitable (according to the laws of probability) that they should contain errors, even if much time has been spent on double-checking and calculating series in different ways. Such errors can be found in most studies dealing with a large amount of quantitative material, and this is further aggravated by the lack of documentation and transparency. For countries where different researchers have constructed different historical GDP series, these often deviate substantially from each other. In spite of this, the impression is often given that the estimates are very accurate.

Feinstein and Thomas (2002) argue that the publication of new historical data should present the margin of errors in a transparent way. This is seldom done however, and the reader cannot gain any idea of how reliable various estimates are.

The users of historical national accounts must be much more conscious of the weaknesses and assumptions underpinning various series. What historical national accounts usually provide is in statistical terms the expected value of various aggregates. For example, if there is a 40 per cent probability that the GDP per capita of Sweden in the year 1000 was 600 dollars and 60 per cent probability that it was 800 dollars, the expected value of the two numbers is 720 dollars ($0.4\times600+0.6\times800=720$). Even if we knew that only a GDP per capita of 600 or 800 dollars is possible, and never a GDP per capita of 720 dollars, presenting the number 720 dollars minimizes the squared error in the estimate. The most important thing is, however, that the expected value presented is not biased. For example, if we know that the calculated estimate probably underestimates the

actual value, it is better to increase the estimate, even if it is based on very rough judgments.

Deciding the actual margin of error is very difficult. Instead, Feinstein and Thomas (2002) propose that researchers should present subjective margins of error for various series based on their calculation methods and assessments of the material. If the margin of error of an aggregate series is to be calculated we also need to estimate a correlation matrix of the different errors, which might also be based on the subjective judgments of the researcher. For example, the root mean square error (RMSE) of the sum of the estimates of *A* and *B* is calculated as:

$$RMSE_{A+B} = \sqrt{(RMSE_A)^2 + (RMSE_B)^2 + 2Corr(E_A, E_B)RMSE_A\,RMSE_B}$$

Corr (E_A, E_B) is the correlation between the two errors E_A and E_B which can range from -1 (perfect negative correlation) to +1 (perfect positive correlation). If there is no correlation this term is set to zero. The above formula shows that a correlation closer to +1 increases the margin of error of the sum of the two estimates. If there is no correlation between the errors of the individual series, this in turn means that the margin of error of the sum is reduced. We can also calculate a coefficient of variation of the error, as the RMSE divided by the estimate, which can be presented as a percentage.

Assume, for example, that the estimate of the GDP of Norway is 100 billion dollars and of Sweden 100 billion dollars. Assume, furthermore, that the RMSE in both instances is 10 billion dollars, which means that the coefficient of variation of the error for each country is 10 per cent. The estimated GDP for Norway and Sweden together is then 200 billion dollars. The estimated margin of error of this sum depends on the correlation between the two errors. Using the formula above, if the correlation is zero, then the estimated RMSE of the sum is 14 billion dollars and the coefficient of variation 7 per cent. If the correlation is +0.5 then the RMSE is 17 billion dollars and the coefficient of variation 8 per cent. If the correlation is +1, then the RMSE is 20 billion dollars and the coefficient of variation 10 per cent. Under the very unlikely cir-

cumstances that there is a perfect negative correlation, the RMSE of the sum would be zero.

This reasoning can be extended to the sum of many estimates. For example, assume that we estimate the GDP of 100 countries, encompassing the whole world, at 100 billion dollars each. The estimate of global GDP will then be 10,000 billion dollars. Assume that the RMSE of each individual country's estimate is 10 billion dollars and the coefficient of variation 10 per cent. If the errors of the individual countries are perfectly correlated then the coefficient of variation of the error for global GDP is also 10 per cent. However, if the errors of the individual countries are not correlated with each other, then the coefficient of variation for the error of global GDP is just 1 per cent.

The point is that given that there is no perfect correlation, the sum is relatively more accurate than its individual components. The various errors partly even each other out. That is one of statistical theory's important insights. Despite the problems of reconstructing historical national accounts for individual countries, estimates of global GDP might actually be more accurate than most or even all of the estimates individual countries. A major problem, however, is whether or not there is a systematic bias across countries, which renders the correlation of their errors close to unity. One such bias complicating our analysis of global GDP in the preindustrial period is Maddison's assumption of a 400-dollar subsistence level.

The 400-dollar subsistence level versus a real wage

There are divergent interpretations of per capita economic growth in the Middle Ages and the early modern period in Western Europe, one, Malthusian, assuming stagnation or even decline and one, Smithian, assuming steady growth. While Maddison advocates the Smithian view, Clark (2009) criticizes Maddison's assumptions of continual economic growth in the Middle Ages and the early modern period in Western Europe. Different indicators point in different directions. The real wage series supports the Malthusian view, while this indicator is heavily criticized by Maddison (2007). Historical national accounts have been used to support either one

or the other of the views. The quandary is that these series display quite different developments, despite the fact that they claim to measure the same variable (GDP per capita).

To reconstruct global GDP back to the Middle Ages or even earlier, we need data on population, the agrarian sector and the non-agrarian sector. All three are problematic to estimate. Maddison and others use outdated population data, which is further discussed by Janken Myrdal in this book. An important check for the agrarian sector is whether or not it yields reasonable estimates of the implied calorie consumption, but that is sometimes forgotten, and many estimates are too low. The non-agrarian sector is probably the most difficult component to calculate. An indicator that has been used, the rate of urbanization (Persson 2008: 170), might be inappropriate given that in some countries most of the non-agrarian sector was located in the countryside.

The Smithian viewpoint mainly rests on Maddison's assumption of a 400-dollar subsistence level in 1990 Geary-Khamis dollars. This is based on the data showing that the poorest countries only had a GDP per capita at this level, or slightly above, in that year. In fact, Maddison uses that to estimate the GDP per capita for most countries in the year 1000 CE. As Gregory Clark (2009) puts it:

> One crucial element is his assumption that the basic subsistence GDP per capita of all societies is $400 (1990 international prices). This is the fundamental constant in Maddison's world, the basic unit of human existence. Any society without a sophisticated production technology, without significant urbanization, and without a substantial rich class, or just where nothing is known, is assigned this minimum. Thus around 1000 AD the various parts of the world are mostly assumed to have incomes either of $400 (uncivilized) or $450 (civilized)… What is that subsistence income in real terms? In 1990 US $ prices, a pound of white bread cost $0.70. So Maddison's $400 is the equivalent of 1.6 lbs of wheaten bread per person per day, or 1,500 kcal. That is an extraordinarily low income, rarely observed in practice. Since most societies have inequality, the poorest in such a subsistence economy would have lived on the equivalent of much less than that daily 1.6 lbs of bread. So if the

> poorest people spent nothing on clothing, heat, shelter, light, and consumed only the cheapest form of calories such as bread, they would still be engaging in hard physical labor on a diet well below 1,500 kcal in the Maddison vision of subsistence.

Since many Western European countries had a GDP per capita of over 1,000 dollars in 1820, Maddison draws the conclusion that there was substantial growth in Western Europe during the Middle Ages and the early modern period. The assumption that the subsistence level was 400 dollars in 1990 Geary-Khamis dollars distorts the comparison of Western European countries with the rest of the world in the early modern period and before (Jerven, 2012 and 2013). It seems unlikely that a country like Sweden on the periphery of Western Europe had a GDP per capita around 1800 that was twice the level of poor African countries in 1990. It is likely that the estimates for 1990 undervalue the actual GDP level for the poorest countries. For example, it should be considered that the new revisions of international guidelines in 1993 recommended the inclusion of all goods production in GDP, also the proportion that is only consumed by the producers themselves. Poor countries implementing this have revised their data upwards.

An alternative to calculating GDP is to use the income approach. This is applied by Gregory Clark (2010) for the English economy for example, by calculating the incomes derived from wages, capital and land retrospectively as far back as the Middle Ages. His estimate of English GDP per capita in the Middle Ages is set much higher than Maddison's, but is also higher than the corresponding estimate for England based on the production approach presented by Broadberry et al. (2010). Clark's series displays stagnating GDP per capita in England between the Middle Ages and the early nineteenth century, which runs counter to the views of most other economic historians.

The main predicament when applying the income approach to the preindustrial period is that most of the incomes have to be estimated theoretically, since only a small part of the economy was monetized. Clark makes the assumption that the ratio of capital income essentially followed wages and rent values, with a correction for long-term changes in the profit rate; and because of the risks

he adds 3 per cent. It is also possible that his rent values are overestimated. According to him, his series "measures rental values when land was rented in a competitive market, not the average rents paid by land occupiers which would often be lower because of customary leases" and that the estimated rent values "are much higher before 1820 than in the recent series of Michael Turner, John Beckett and Bethany Afton" (Clark 2002: 201). Altogether, this implies that his estimated property incomes may be overestimated. Clark (2010) estimates the wage share in 1200 at 47.8 per cent, which was even lower than its share in 1860 at 65.1 per cent.

There is a difference in estimating theoretical production values and incomes. For example, while estimates concerning theoretical values for building houses for own-use may be questionable, they are still based on actual production activity. However, when the contribution of capital or labor is included in the estimation, the researcher may be led astray since their contribution must first of all be measured in terms of how much they contribute to physical output. Estimating physical output is necessary before anything can be said about incomes. Even in developed countries today some of the incomes must be imputed as well. The category of mixed income concerns production units where the laborer also owns the capital. It is a sum of labor and property income. In a previous study, I have tried to estimate property income in agriculture in late twentieth-century Sweden (Edvinsson 2005). Since the self-employed constituted a large proportion of the labor force, most of the labor income must be calculated by using the wage rate of agricultural wage earners. Property income cannot be estimated directly, but only indirectly, as the value added less the estimated labor income. The net property income thus estimated was negative for the late twentieth century. The only conclusion that could be drawn was that both capital and labor in modern Swedish agriculture are paid significantly below their market values. In other words, not knowing physical output and using the income approach for Swedish agriculture in the late twentieth century would significantly overestimate value added if it was assumed that capital and labor were paid their full market values. Assuming market rates for the Middle Ages must be even more questionable.

A variant of the income approach is the estimation of agricultural production according to the so-called demand approach. For example, Malanima (2010) presents annual estimates for Italy back to 1300 and Álvarez-Nogal and Prados de la Escosura (2011) for Spain back to 1270, using the approach for agricultural output, while other activities are approximated from the rate of urbanization. The demand approach was advocated by Allen, based on positing a demand curve for agricultural products (Allen 2001: 13). Consumer theory requires that own price, income, and cross-price elasticities of demand add up to zero. Agricultural production is accordingly calculated from the development of real wages and the real prices of agricultural and non-agricultural products. The change in income is set equal to the change in the wage rate, and rests on the assumption of no change in the wage share or labor input per worker. The problem is, of course, that the assumption of a constant wage share has no empirical backing, even for the modern period. Theoretically it is motivated by the Cobb-Douglas function in neoclassical growth theories, but there are also other functional specifications that do not rest on constant shares in the incomes of labor and capital. It is also quite likely that the share of labor income was much higher in the Middle Ages than it is in a fully developed capitalist economy.

Both the 400-dollar approach and the identification of income with wages rest on shaky assumptions. To overcome the gap between Smithians and Malthiusians, new measures involving more direct indicators of production are necessary. One path forward would be to apply a kind of expenditure approach, where reasonable estimates are made of consumption and investment patterns to assess the level of GDP per capita. The expenditure approach can also be used to cross-check whether various estimates are reasonable, such as in the quote by Clark that even the assumption that people only live on bread must entail a GDP per capita above 400 dollars.

One example can illustrate the expenditure approach. Most studies show that food consumption in Sweden, as well as in other European countries, was better in the late Middle Ages than in the early nineteenth century. An indicator of the food nutrition standard is average heights. For Sweden, archeological findings record that the average height among men was 170–174 centimeters in the Middle

Ages, and 170–176 centimeters in the Viking Age (Gustafsson et al., 2007). An average height of 172 centimeters is recorded for conscripted men born in the 1890s, and a height of 173 centimeters for conscripted men born around 1910 (Öberg, 2014: 17). In 1900, food consumption made up around one third of GDP. In that year GDP per capita was 2,202 (1990 Geary-Khamis) dollars (Edvinsson 2013). Food consumption alone therefore contributed around 700 dollars of the GDP per capita. A similar calculation for the early nineteenth century shows that food consumption contributed around 600 dollars, consumption of clothes 100–150 dollars, housing around 50 dollars, wood products (mostly firewood) around 50 dollars, and building and construction around 50 dollars of the GDP per capita. Given the climatic conditions in Sweden it is unlikely that people could survive having much worse clothing, housing and warming in the Viking Age or the Middle Ages than in the early nineteenth century. Assuming that other parts of GDP contributed to less than 50 dollars per capita, GDP per capita in Viking age and medieval Sweden should have been around 1,000 dollars, which is at the same level as in the early nineteenth century.

This very simple application of the expenditure approach shows that the GDP per capita in Sweden probably did not change much before the nineteenth century. Any growth in some parts of GDP, such as trade and the public sector, was mostly offset by a decline in food standards. More importantly, the estimate proposed here is much above Maddison's assumption for Sweden of 400 dollars for the Viking Age and 650 dollars for the year 1500. This gives further support to Clark's view that no society could have survived on a 400-dollar subsistence level in 1990 Geary-Khamis dollars, even if we allow that different climatic conditions can be accompanied by different subsistence levels (Jerven, 2012: 119).

Summary and conclusions

This paper focuses on the methodological questions and dilemmas of reconstructing historical series of global GDP. Various problems stem in part from the endeavor to make international comparisons. If we want to compare countries and world regions we need common

definitions and methods. Sometimes differences must be assumed to be of minor importance, otherwise comparisons cannot be made. Data collection is extremely time-consuming even for a single country or a single branch for a short period. Estimating world GDP in the last centuries or millennia necessitates certain shortcuts.

It must be emphasized that constructing historical national accounts is not an exact science. It is inevitable that certain assumptions must be made and accepted. Even modern national accounts are not very exact, and are based on much guesswork, such as, for example, estimating the extent of illegal activities to be included in GDP, or other activities prone to a high degree of tax evasion. The solution may lie in publishing all the documentation and calculations that underpin the historical national accounts.

Interpreting history is never a neutral act and neither is the critique of such interpretation. Which methods and definitions to use is not just a purely objective question, but is also dependent on the purpose to which each series is to be used. Official national accounts are not socially neutral, as they may appear, and are adapted to the needs of the social community of the present day, not least its economic policy. Using the same definitions and methods to construct macroeconomic series back to the Middle Ages, or even the dawn of mankind, unavoidably introduces anachronistic elements. Writing history on the basis of the definitions of official national accounting is in a sense partly writing history from the perspective of the social system that has conquered the whole world in the last two centuries, namely the capitalist system. Putting a price tag on all the goods and services produced in societies mainly based on self-sufficiency could still be done, however, if there were at least some market activities. However, economies with no market activity, such as hunters-gatherers, cannot be valued in current prices. Similarly, some productive activities, such as unpaid domestic services, currently excluded from official estimates of GDP, are not priced. There is, therefore, a need to develop new alternative measures that are not necessarily based on valuations using market prices.

Acknowledgments

I am grateful for financial support from the Swedish Collegium of Advanced Study, Riksbankens Jubileumsfond and Vetenskapsrådet. I am also grateful to Stefan Öberg, Gothenburg University, who has helped me to find data on human heights.

References

Allen, Robert. 2001. "Economic structure and agricultural productivity in Europe, 1300–1800", *European Review of Economic History*, 3(1), 1–25.

Álvarez-Nogal, Carlos & Prados de la Escosura, Leandro. 2011. *The Rise and Fall of Spain (1270–1850),* http://e-archivo.uc3m.es/bitstream/10016/10877/1/wp_11-02.pdf.

Bassino, Jean-Pascal, Broadberry, Stephen, Fukao, Kyoji, Gupta, Bishnupriya & Takashima, Masanori. 2011. *Japan and the Great Divergence, 730–1870*, http://www2.warwick.ac.uk/fac/soc/economics/…/japangreatdivergence3a.pdf.

Bourguignon, François & Lévy-Leboyer, Maurice. 1990. *The French economy in the nineteenth century: An essay in econometric analysis*. Cambridge: Cambridge University Press.

Broadberry, Stephen. 2003. "Relative Per Capita Income Levels in the United Kingdom and the United States since 1870: Reconciling Time-Series Projections and Direct-Benchmark Estimates", *The Journal of Economic History*, 63(3), 852–863.

—— & Gupta, Bishnupriya. 2009. *Indian GDP Before 1870: Some Preliminary Estimates and a Comparison with Britain*, http://www2.warwick.ac.uk/fac/soc/economics/staff/faculty/broadberry/wp.

—— & Klein, Alexander. 2012. "Aggregate and per capita GDP in Europe, 1870–2000: Continental, regional and national data with changing boundaries", *Scandinavian Economic History Review*, 60(1), 79–107.

——, Campbell, Bruce, Klein, Alexander, Overton, Mark & van Leeuwen, Bas. 2010. *British economic growth, 1270–1870*. Unpublished paper.

——, Guan, Hanhui & Daokui Li, David. 2012. *China, Europe and the Great Divergence: A Study in Historical National Accounting*. Unpublished paper.

Cashdan, Elizabeth. 1989. "Hunters and Gatherers: Economic Behaviour in Bands", in Plattner, Stuart (ed.), *Economic Anthropology*. Stanford: Stanford University Press.

Clark, Gregory. 2002. "Land rental values and the agrarian economy: England and Wales, 1500–1914", *European Review of Economic History*, 6(3), 281–308.

—— 2009. "Contours of the World Economy, 1–2030 AD: Essays in Macro-Economic History by Angus Maddison", *The Journal of Economic History*, 69(4), 1156–1161.

—— 2010. "The macroeconomic aggregates for England, 1209–2008", *Research in Economic History*, 27, 51–140.

Edvinsson, Rodney. 2005. *Growth, Accumulation, Crisis: With New Macroeconomic Data for Sweden 1800–2000*. Stockholm: Almqvist & Wiksell.

—— 2009. "Ekonomisk tillväxt, genusarbetsdelning och obetalt hemarbete i historisk belysning", *Historisk Tidskrift*, 129(1), 5–26.

—— 2013. "New Annual Estimates of Swedish GDP in 1800–2010", *The Economic History Review*, 66(4), 1101–1126.

Eitrheim, Øyvind, Klovland, Jan T., & Qvigstad, Jan F. (eds.). 2004. *Historical Monetary Statistics for Norway 1819–2003*. Oslo: Norges bank.

Eloranta, Jari (ed.). 1997. *Nordiska historiska nationalräkenskaper: Workshop 2 i Järvenpää 20–22 september 1996*. Jyväskylä: University of Jyväskylä.

Feinstein, Charles & Thomas, Mark. 2002. "A Plea for Error", *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 35(4), 155–165.

Folbre, Nancy & Wagman, Barnet. 1993. "Counting Housework: New Estimates of Real Product in the United States, 1800–1860", *The Journal of Economic History*, 53(2), 275–288.

Grytten, Ola. 2004. "The gross domestic product for Norway 1830–2003", in Eitrheim, Øyvind, Klovland, Jan T. & Qvigstad, Jan F. (eds.), *Historical Monetary Statistics for Norway 1819–2003*. Oslo: Norges bank.

Gustafsson, Anders, Werdelin, Lars, Tullberg, Birgitta S. & Lindenfors, Patrik. 2007. "Stature and Sexual Stature Dimorphism in Sweden, from the 10th to the End of the 20th Century", *American Journal of Human Biology*, 19(6), 861–871.

Hansen, Svend Aage. 1974. Økonomisk *vækst i Danmark*, Band II: 1914–1970. Copenhagen: Akademisk Forlag.

Hjerppe, Riitta. 1996. *Finland's Historical National Accounts 1860–1994: Calculation Methods and Statistical Tables*. Jyväskylä: University of Jyväskylä.

Horlings, Edwin, Smits, Jan-Pieter & van Zanden, Jan Luiten. 2000. *Dutch GNP and its components, 1800–1913*. Groningen: Groningen Growth and Development Centre.

Inter-Secretariat Working Group on National Accounts. 1993. *System of National Accounts 1993*, Brussels, Luxembourg, New York, Paris & Washington: World Bank.

Jerven, Morten. 2012. "An unlevel playing field: National income estimates and reciprocal comparison in global economic history", *Journal of Global History*, 7(1), 107–128.

—— 2013. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. Cornell: Cornell University Press.

Jonas, Paul & Sardy, Hyman. 1970. "The Gerschenkron Effect: A Re-Examination", *The Review of Economics and Statistics*, 52(1), 82–86.

Jonsson, Gudmundur. 1997. "Non-market production: What to count and how?", in Eloranta, Jari (ed.), *Nordiska historiska nationalräkenskaper: Workshop 2 i Järvenpää 20–22 september 1996*. Jyväskylä: University of Jyväskylä.

Krantz, Olle. 1987. *Husligt arbete 1800–1980* (*Historiska Nationalräkenskaper för Sverige. 6*). Lund: Ekonomisk-historiska föreningen.

Lindahl, Erik, Dahlgren, Einar & Kock, Karin. 1937. *National Income of Sweden 1861–1930*, part one. London & Stockholm: P.S. King.

Maddison, Angus. 2007. *Contours of the world economy, 1–2030 AD: Essays in macro-economic history*. Oxford: Oxford University Press.

—— 2010. "Historical Statistics of the World Economy: 1–2008 AD", http://www.ggdc.net/maddison/Historical_Statistics/horizontal-file_02-2010.xls.

Malanima, Paolo. 2010. "The long decline of a leading economy: GDP in central and northern Italy, 1300–1913", *European Review of Economic History*, 15(2), 169–219.

Nyberg, Anita. 1995. *Hemarbetets volym och värde*. Linköping: Linköping University.
Öberg, Stefan. 2014. *Social bodies: family and community level influences on height and weight, southern Sweden 1818–1968*. Gothenburg: University of Gothenburg.
O'Rourke, Kevin & Williamson, Jeffrey. 2002. "When did globalization begin?", *European Review of Economic History*, 6(1), 23–50.
Persson, Karl-Gunnar. 2008. "The Malthus delusion", *European Review of Economic History*, 12(2), 165–173.
Plattner, Stuart (ed.). 1989. *Economic Anthropology*. Stanford: Stanford University Press.
Prados de la Escosura, Leandro. 2000. "International Comparisons of Real Product, 1820–1990: An Alternative Data Set", *Explorations in Economic History*, 37(1), 1–41.
Schön, Lennart & Krantz, Olle. 2012. "Swedish historical national accounts 1560–2010", *Lund Papers in Economic History*, no 123.
United Nations, European Commission, OECD, IMF & World Bank. 2009. *System of National Accounts 2008*. New York: United Nations.
Utredningen om översyn av den ekonomiska statistiken. 2002. SOU 2002:118, Bilaga 3: Beräkningsrutiner för nationalräkenskaperna. Stockholm.
Ward, Marianne & Devereaux, John. 2003. "Measuring British Decline: Direct Versus Long-Span Income Measures", *The Journal of Economic History*, 63(3), 826–845.