Discussion Paper Series

No.235

Measuring 'The Wealth and Poverty of Nations': Methodological Problems and Possible Solutions

Jan-Pieter Smits

January 2008

Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences

A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

Measuring 'The Wealth and Poverty of Nations': Methodological Problems and Possible Solutions[1]

Jan-Pieter Smits/ Groningen Growth and Development Centre/ Statistics Netherlands

1. Introduction

One of the main debates in economics and economic history concerns the analysis of the great divergence between the West and the Rest. Historical National Accounts are of vital importance to address this issue. Ever since Simon Kuznets embarked upon his ambitious project to chart the long-run economic development of countries, historians and economists have tried to construct datasets on long-run economic growth and structural change. This research seems to be flourishing nowadays as many "regional initiatives" have been taken to construct the historical national accounts for East Asia, Latin America and Europe.

Now that more and more detailed historical national accounts are made available, it is time to focus more on the comparability of these estimates. Even a superficial glance at the data reveals that data are constructed on the basis of very different methods, which undermines the possibility of sound international comparisons. It is for this reason that we have recently started a new research project at the *Groningen Growth and Development Centre* (GGDC) into the construction of comparative historical national accounts. This project is supported by the *International Economic History Association* which stimulates the formation of data-hubs in various research domains in economic history.

In this paper I want to present an outline of this new project and discuss some of the methodological issues involved (see section 2). I will emphasise three of the most problematic areas which hamper a sound comparison of levels and rates of growth, i.e. the treatment of services (section 3), the use of deflators (section 4) and the construction of historical purchasing power parities (PPP's) (section 5).

2. Standardising Historical National Accounts

Ever since Kuznets initiated the research into long-term economic growth, researchers have been active in constructing datasets for their respective countries. All these research efforts culminate in the regular publications by Angus Maddison[2], who devotes an incredible amount of time and energy in tracking down these data, encouraging scholars to start working on the historical accounts for countries for which no data are available, and to make the various estimates more comparable. Thanks

[1] I thank Jop Woltjer for helping me out with the section on deflation. Comments and feedback welcome at: j.p.h.smits@rug.nl

[2] See for example: A. Maddison, *The world economy: a millenial perspective* (Paris-OECD, 2001).

to these huge efforts we now have a dataset at our disposal which covers large parts of the world economy for long periods of time.[3]

Of course the work of Maddison can be criticised on a number of points. For example, for a couple of countries the GDP series could and should be improved. There is a general consensus among German scholars that the original Hoffmann estimates are flawed and need to be revised. Besides, in Maddison's dataset only one recent benchmark is used to convert the current price estimates of the various countries into one common denominator. It would, of course, be better to take relative price changes over time into account by making Purchasing Power Parity (PPP) comparisons for earlier benchmark years. These points of criticism should not be aimed at Angus Maddison, 'le Grand Maitre', himself. In case individual series for particular countries are flawed, this is first of all the responsibility of the researchers of the countries in question. In presenting his grand overviews, Maddison can't offer more than researchers in the field of historical national accounting have been producing themselves.

The new research project that has been started at the Groningen Growth and Development Centre aims to build on the previous –ongoing- work of Maddison. We intend to present historical national accounts data for the period 1870-present, but in more detail than has been done so far. In order to enhance the analysis of structural change, we will disaggregate the GDP data at least by giving sector estimates (agriculture, industry and services) and preferably by covering 25-30 underlying industries. Besides, for the western countries attempts will be made to include estimates from the income- and expenditure side. Last but not least, due attention will be paid to the standardisation of these estimates.

The project 'An International Historical Database on Economic Growth and Development' (carried out by Jan-Pieter Smits and Herman de Jong) is supported by the International Economic History Association and financed by the Dutch science foundation (NWO). In this project we will closely co-operate with existing networks in the field of historical national accounting. It is our intention to build a data hub, a central place were the data will be stored, described and updated, but which is linked to existing datasets that are constructed by other researchers. For example, our database will be linked to the Eastern Asian research project carried out at Hitutsubashi University under the supervision of Fukao. For the Latin American countries, contacts have been established with Carreras, Hofman and Yanez who attempt to extend and standardise the time series of the various Latin American countries. Last but not least, the GGDC project will be embedded in ESF/SCSS Scientific Programme *Globalizing Europe Economic History Network* (GLOBALEURONET), co-ordinated by Batilossi (Carlos III, Madrid) in which 13 countries among which the Netherlands participate. Within this programme the Groningen Growth and Development Centre will organise workshops and seminars aimed at standardising economic growth estimates for the countries both in 'old' and 'new' Europe.

[3] Publications and data from Angus Maddison can be found at the website of the Groningen Growth and Development Centre: http://www.ggdc.net/Maddison/.

Furthermore, this historical research is linked to more contemporary work that is already carried out within the research group of the Groningen Growth and Development Centre.[4] Above all the EU KLEMS project (Productivity in the European Union: A Comparative Industry Approach) should be mentioned. This project aims to create a database on measures of economic growth, productivity, employment creation, capital formation and technological change at the industry level for all European Union member states from 1970 onwards. Productivity measures will be developed, in particular with growth accounting techniques. Several measures on knowledge creation will also be constructed. Substantial methodological and data research on these measures will be carried out to improve international comparability. There will be ample attention for the development of a flexible database structure, and for the progressive implementation of the database in official statistics over the course of the project.

This project is funded by the European Commission, Research Directorate General as part of the 6th Framework Programme, Priority 8, "Policy Support and Anticipating Scientific and Technological Needs". The research consortium is led by the Groningen Growth and Development Centre of the University of Groningen (The Netherlands) and the National Institute of Economic and Social Research in London.

The new *historical* database project is closely linked to these projects. However, except from sharing and linking data, this GGDC project also aims to provide some important, and perhaps crucial, new elements. First of all, we aim to put the disaggregated, standardised data on the GGDC-website (http://www.ggdc.net) and make the datasets downloadable. Apart from that, there are some specific areas in which additional research will be done to standardise the various time series:

1) *The treatment of the service sector*: during the last two decades new estimation procedures have been developed in order to measure output and productivity in services. However, in many of the first generations of historical national accounts (such as those of Great Britain, France and Germany) unsound estimation procedures have been followed. The measurement of service output needs to be standardised (section 3)
2) *The use of deflators*: researchers in various countries use different deflation techniques. Often, without a clear methodological justification, use is made of Paasche, Divisia, or other price indices. Especially the question whether weighting schemes should be changed annually, or that we should stick to weighting schemes for period of 15-20 years, must be discussed more thoroughly. In this section 4 of this paper I will show how different deflation techniques may impact on the GDP growth rates that are calculated.
3) *The construction of historical benchmark PPP's*: economists and historians often use data expressed in dollars of 1985 or 1990 in order to make international comparisons of income

[4] The information of EU Klems is derived from the website: http://www.euklems.net.

and productivity levels. However, the relative price structure of the late 20th century may not be representative for earlier periods in time. Recently interesting attempts have been made to develop expenditure PPP's for early benchmark years. In this last part of the paper I will discuss the advantages and disadvantages of the expenditure- vis-à-vis the industry-of-origin approach. Moreover, new industry-of-origin estimates will be presented for a couple of Western countries around 1910.

3. The treatment of services

The measurement of service production has always been problematic.[5] There are different measurement techniques of service output in historical national accounting as well as contemporary national accounting.[6] In addition, historical estimates of service output often faces data problems, because we do not have all of the required information at our disposal. Output measurement for agriculture and manufacturing is less difficult, as it basically consists of summing the quantities times the prices of goods. This procedure is impossible to follow for many service industries. Only for branches such as transport and communications, such a 'commodity' approach is applicable. Output in many other service industries is arrived at by means of input valuation, i.e. adding the value of inputs, mostly wages. Of course, this procedure has a major shortcoming. Output is not estimated independently of inputs, and, as a consequence, cannot be used for productivity purposes.

In this section an overview is presented of the estimation procedures that have been applied in the first generation of historical national accounting –for example the United Kingdom and Germany- and how these estimates can and should be improved..

Transport

Output volume in transports is ideally estimated by multiplying the number of passengers and the tonnage of goods transported with the distance of transportation. In this way the output volume of the transport sector is expressed as the number of passenger kilometres and the number of ton kilometres. Next these output figures are linked to freight rates, that is the weighted price of transportation per passenger kilometre or per ton kilometre. This results in the value of output. Finally, the value of inputs are deducted in order to arrive at the estimates of value added.

In most countries such estimates have not been made. Often value added has been calculated by multiplying employment figures with data on the average income. For instance Feinstein and Hoffmann use this method to estimate production in the transport sector for a benchmark year around

[5] The first part of this section, with the overview of the different ways in which service output can be calculated, is taken from: A. Burger and J.P. Smits, "A benchmark comparison of service productivity between Europe and the United States for 1910", *Economic and Social History in the Netherlands*, special issue: Historical National Accounts in the Netherlands 7 (1996) 133-159.

[6] Cf. Z. Griliches (ed.), *Output measurement in the service sectors* (Chicago, 1993); O. Krantz, "Service production in historical national accounts", *Review of Income and* W*ealth* (1994).

1910.[7] Direct estimates from the output side (i.e. the actual calculation of outputs and inputs) are scarce. Yet, in the Swedish and Dutch historical national accounts much work has been done in this respect.[8] For the Unites States the work of Gallmann and Weiss should be mentioned.[9] By using annual reports of important transport companies, gross earnings and the value of inputs can be calculated. However, in many other countries 'guestimates' from the input side are made.

The quality of benchmark estimates may be questionable, even more serious problems arise if we try to analyse growth patterns for the pre-1913 period. It was already mentioned that in most countries 1910/13 estimates were projected backwards using all sorts of production (volume) indices. Hoffmann was one of the few who used actual data on output volume such as the number of passenger kilometres and ton kilometres. In the British historical national accounts, however, rough indicators were used to make backward projections. For instance, Feinstein was able to estimate value added for transport for the benchmark year 1907. Next, series on production volume (derived from Lewis) were used in order to extrapolate the 1907 estimates backwards.[10] In this way a transport series in constant prices was constructed. For railways it was simply assumed that that production in the period 1855-1907 increased at a rate of 3% a year. As far as sea shipping is concerned Lewis used used the total tonnage of Britain's merchant fleet as an indicator for production. Because of the fact that steam ships were sailing at greater speed, the tonnage of steam ships was multiplied by a factor 3. Furthermore, on the basis of data from the foreign trade statistics, it was assumed that the rate of utilisation in sea transport increased with 0,5% a year.

Of course serious criticism can be levelled against such assumptions. First of all, no estimate of the actual volume of transportation was made. For example, no attention is paid to possible changes in the average distance. Furthermore, changes in input/output relations were disregarded. Thus, the effect of technological and organisational changes in transport is not measured. Moreover, it should be noted that by using a fixed set of prices, changes in relative prices are not reflected in the production estimates for earlier years.

Trade

The analysis of productivity in trade causes even greater problems. Mostly value added in trade was calculated for a benchmark year around 1910 on the basis of data on employment and average income (see the United Kingdom, Germany and the United States). As far as the construction of time-series in concerned the growth of distribution is often calculated on the basis of agricultural and industrial

[7] W.G. Hoffmann, *Das Wachstum der Deutschen Wirtschaft seit der Mitte des 19. Jahrhunderts* (Berlin, 1965); C.H. Feinstein, *National income, expenditure and output of the United Kingdom, 1855-1965* (Cambridge, 1972).
[8] O. Krantz, "Historical national accounts –some methodological notes', *Scandinavian Economic History Review* 2 (1983); J.P. Smits, E. Horlings and J.L. van Zanden, *Dutch GNP and its components* (Groningen Growth and Development Centre, Research monograph no. 5, 2000).
[9] Gallman and Weiss, "Gross national product in the United States 1834-1909', *Output, employment and productivity in the United States after 1800* (New York, 1966).
[10] W.A. Lewis, *Growth and fluctuations 1870-1913* (London, 1978).

output figures as well as the value of imports. In the British historical national accounts, for example, the growth is considered to be equal to the growth of the output volume in agriculture, manufacturing, mining and the volume of imports (including re-exports).

In Germany more or less the same method is followed. Hoffmann, however, stressed that his estimates are of a tentative nature.[11] In his calculations it was assumed that no elimination of intermediate trade took place. This means that trade margins (i.e. the difference between sales and purchases values) are supposed to be constant throughout the period 1850-1913. Such an assumption may not plausible. For example, in the period 1890-1913 trade margins in the Unites States were rising.

There are strong indications that trade margins did not remain constant during the nineteenth century. First of all, important structural changes took place in retail trade. The rise of multiple shop organisations and department stores had a decisive impact on productivity and on the development of trade margins. Furthermore, innovations in the field of transport and communication made an elimination of intermediate trade possible, as a result of which the costs of distribution could be lowered substantially.

To arrive at more reliable estimates for value added in trade it is therefore important to establish trade margins for wholesale and retail trade. On the basis of records of trading companies and by comparing levels of different types of prices (producer prices, wholesale prices and retail prices) trade margins can be established. Next we can check the validity of such estimates by calculations from the income side, i.e. by using employment data and information on average incomes.

Other services (mainly government and domestic servants)

Production in this segment of services is mostly estimated by combining employment data with figures on average income. For the period around World War I we are amply provided with such data. Analysing productivity changes over time, however, leads to serious problems. If the production series (calculated as employment multiplied by income) are deflated by wage indices, it is supposed that productivity did not rise. Of course this is not plausible. In the Swedish historical national accounts Krantz therefore tried to construct indicators for the development of productivity. For example, he used the number of patients in hospitals as an indicator for the production in medical services, and for education the number of pupils was used.

It should be noted that the income estimates are not always comparable because of different methods of deflation and the underlying assumptions concerning productivity changes.[12] Lewis, for instance, supposes that productivity in the 'income related' services (i.e. government, domestic servants, professions and other personal services) increased with an average of 0.5% a year. In most

[11] Hoffmann, *Das Wachstum*.
[12] O. Krantz, "Historical national accounts".

other countries such an adjustment was not made, as a result of which the British estimates on services for the pre-1907 period are not comparable with those of other countries.

It is clear that especially in the historical national accounts of the first generations, short-cuts were used to estimate service output. This might distort the international comparison of growth rates, especially as services comprise a substantial part of GDP in high-income countries such as the United Kingdom and Germany.

In an updated version of this paper, new service estimates will be presented for the UK and Germany, by applying modern estimation techniques. For now I just present a sensitivity analysis on the basis of the Dutch database[13]. The following tables compare estimated of service output as they were calculated in the Dutch historical national accounts, with counter-factual outcomes using the short-cut estimation procedures as they were used in the British accounts (mainly the work of W.A. Lewis).

Distribution:

Table 1 shows the outcomes of the original value added estimates for distribution made in the Dutch historical accounts, compared to the short-cut method that was used in the British accounts and which was applied to the Dutch basic data.

Table 1: Value added of distribution (at constant prices; Dutch Guilders of 1913)

	shortcut	original
1850	100,1	108,9
1890	186,1	246,2

The 'shortcut' method seems to underestimate the levels of output of trade (in 1850 8% lower, in 1890 even 24.1% lower!). Especially after 1890 the trade sector shows much more growth when we use the shortcut method. The main reason why the two methods yield such different results is that the short-cut method does not take changes in trade margins into account. However, these margins were far from constant during the nineteenth century. For example, large parts of intermediate trade were eliminated as a result of better transport and communication facilities. Besides, a distinction should be made between wholesale trade and retail trade. The larger growth of trade until 1890 according to the calculations carried out in the Dutch national accounts project, stems from the fact that retail trade – which is characterised by relatively high trade margins- witnessed strong growth in the period 1850-1890.

[13] Cf. J.P. Smits, E. Horlings and J.L. van Zanden, *Dutch GNP and its components, 1800-1913* (Groningen Growth and Development Centre, 2005). This monograph can be downloaded at: http://www.ggdc.net. Go to DHNA (Dutch Historical National Accounts). Here the monograph as well as some of the time series can be downloaded.

Transport:

Also in the transport sector substantial differences can be demonstrated.

Table 2: Value added of maritime shipping (at constant prices; Dutch Guilders of 1913)

	shortcut	original
1850	5,3	6,0
1890	29,2	5,5

In the case of maritime shipping, the shortcut method clearly produces implausible results. Apparently, capital-output ratios are not constant as was assumed in the Lewis series (which takes the capacity of the merchant fleet as the basis for the output estimates). Perhaps even more important, the shortcut method ignores changes in input/output relations. The Dutch data, however, indicate that the costs of input showed a relative increase. From the late 1860s the freight rates in merchant shipping plummeted, whereas input costs remained constant or even increased.

Table 3: Value added of railways (at constant prices; Dutch Guilders of 1913)

	shortcut	Original
1850	9,1	0,8
1890	29,6	23,1

Here, the differences for 1890 are quite limited. However, with the shortcut method -which simply assumes a fixed 3% growth rate in the pre-1913 period- we arrive at a much higher level. This should not come as a surprise, as the Dutch railway network was still in its infancy around 1850 and in these early years of course higher rates of growth were realised.

Other services:

The main difference in the estimation methods for the other services concerns the assumption that has been made in the shortcut method, i.e. that productivity in the other services increased with 0,5% a year. Of course, when the absolute figures for 1910 are being projected backwards on indices with and without such a productivity adjustment, we arrive at different estimates for the earlier benchmark years.

Table 4: Value added of other services (at constant prices; Dutch Guilders of 1913)

	shortcut	original
1850	133,0	151,2
1890	285,4	320,3

In the years 1890 and 1850 the output of the remaining services is 11 respectively 12% lower than the figure that was calculated in the Dutch historical national accounts. Especially over longer time periods, productivity assumptions have a huge influence on output levels for earlier years.

Total services:

In table 5 these results are aggregated to a sector level.

Table 5: Value added of total services (at constant prices; Dutch Guilders of 1913)

	Shortcut	original
1850	267,1	302,8
1890	610,6	655,6

Adding all the information together, we find that the shortcut estimates which were followed in the United Kingdom and partly also in Germany, yield lower levels of service output in earlier benchmark years. In 1890 value added is 6,7% lower than the actual calculations show, in 1850 a difference of even 12% is reached.

These differences are considerable. Of course these figures do not *automatically* imply that the British and German figures are actually 7-12% too low in the pre-1913 period. However, this sensitivity analysis shows that the methods which have been applied in the older generations of historical national accounts, may yield implausible outcomes.

These problems can be overcome. It is not too difficult to 'upgrade' the estimates of these first generation accounts. With some additional research into (relative) prices and taking input/output changes into account, better estimates can be produced. This can be done mainly on the basis of of Statistical Yearbooks and Foreign Trade statistics. In an updated version of this paper, such estimates will be presented.

4. Deflation procedures

The problems regarding deflation are often dismissed as technical number-crunching, an issue that only national account nerds are interested in. And indeed, testing the different methods to transform current price series into constant prices may not be one of the most exciting tasks for economic historians. Yet, deflation is a problem which should be taken seriously as it might have serious analytical implications.

Even a quick review of the ways in which the problem of deflation has been dealt with, reveals that there are almost as many deflation techniques as there are historical national accounts. Some of these differences stem from source-related reasons. In some countries, where estimates could only be made at a sector level (agriculture, industry and services), the construction of the deflator is based on a limited number of price indices. In countries with a large number of underlying series, the aggregate GDP deflator is often based on 20-30 sub series.

A second difference, concerns the weighting schemes that are used. Especially in the first generation of historical national accounts, mostly one fixed price set (often around 1910) is used. In the later generations, weighting schemes are periodically changed. In the case of Sweden and the

Netherlands the weights are changed every 15-20 years. Some researchers even argue that it is desirable to change weights on an annual basis. Last but not least, different types of deflators are used. Many economic historians follow the United Nations National Accounting (SNA) guidelines and use Paasche-type deflators, but often also Fischer, Thornqvist and Divisia price indices are used.

Of course, different types of deflation techniques yield different results. Therefore, databases on long-run growth unfortunately contain heterogeneous time series. This raises concern with respect to the comparability of growth estimates, as we might end up analysing different deflation techniques rather than actual differences in GDP growth rates.

It is not easy to standardise deflation procedures for the present set of historical national accounts that we have at our disposal. For example, in the first generation of historical national accounts (such as the United Kingdom and Germany), a fixed price set for 1910 is used and there are no underlying price series at industry level given to re-calculate the deflator.

I therefore use another kind of sensitivity analysis to try to figure out whether differences in deflation techniques might indeed disturb our view of the long-run growth patterns. As we have rather detailed information at our disposal in the Dutch Historical National Accounts, we have re-estimated Dutch real GDP growth (i.e. at constant prices), using different types of deflators. However, it should be noted that the Dutch economy witnessed only a modest degree of structural change (in terms of shifts from agriculture to other branches of the economy), which implies that the effects of choosing different deflators might be less important than for countries which witnessed more structural change.

In this sensitivity analysis 14 different types of deflators were applied. The first seven series are calculated on the basis of a sub division of GDP in only three sectors: agriculture, industry and services:

Deflation based at price indices for three sectors

A fixed price set for 1913:
-Series 1: fixed price set for 1913

Periodically changing weighting schemes. The period 1890-1913 is weighted using the relative prices for 1913, 1870-1890 with 1890 etc etc. The weighting schemes are changed every twenty years, and subsequently all series were linked and the final series are also expressed at constant prices of 1913

-Series 2: Paasche index
-Series 3: Laspeyres index
-Series 4: Fischer index

Annually changing weighting schemes, a chain index
-Series 5: Paasche index
-Series 6: Laspeyres index
-Series 7: Fischer index

The seven different types of deflation have been applied on the current price series for agriculture, industry and services. This results in seven new series on GDP at constant prices. These series have been compared to the 'ideal' outcome, and that is the series the way we calculated them in our Dutch historical accounts (Paasche deflator, on the basis of 20-yearly changing weighting schemes). I characterise this series as 'ideal', as we follow the United Nations SNA guideline to use a Paasche index and Olle Krantz' recommendation to use periodically changing weighting schemes. Table 6 reports to what extent GDP levels as calculated according to the 7 different ways of deflating, differ from the original Dutch estimates.

Table 6: Comparing the effects of different deflators on the levels of GDP (in %)

	1	2	3	4	5	6	7
1807	6,2	3,0	-3,4	-0,3			
1830	2,6	2,3	-1,6	0,3	1,5	0,9	1,2
1850	5,8	4,9	0,9	2,9	4,7	2,4	3,5
1870	3,9	3,6	2,2	2,9	3,2	3,1	3,1
1890	2,4	2,4	1,8	2,1	2,5	2,3	2,4

Note: The differences in GDP levels are calculated as follow. With each deflator (1-14) we estimated GDP levels. These levels were compared to the one we arrived at using method no 14 (periodically changing weighting schemes of a Paasche type). The difference we found was expressed as a percentage of the GDP level according to method 14.

Even a deflation on the basis of only three underlying price series gives substantial differences. Especially in case of using a fixed price set for 1910 –such as has been done for the United Kingdom and Germany- the results are quite different. This fixed price approach results in GDP levels that are 6% lower in 1807 and 1850 than when we use periodically changing weighting schemes of a Paasche type.

Exactly the same procedure has been repeated, but now with a GDP deflator which is not based on just the three sector deflators, but which takes into account the circa 25 underlying price indices for the main branches within agriculture, industry and services.

Deflation based on 25 price indices at industry level

A fixed price set for 1913:
-Series 8: fixed price set for 1913

Periodically changing weighting schemes
-Series 9: Paasche index
-Series 10: Laspeyres index
-Series 11: Fischer index

Annually changing weighting schemes, a chain index
-Series 12: Paasche index
-Series 13: Laspeyres index
-Series 14: Fischer index

Also here we compare the GDP outcomes at constant prices, using deflation techniques 8-14 with the original estimation procedures used in the Dutch Historical National Accounts

Table 7: Comparing the effects of different deflators on the levels of GDP (in %)

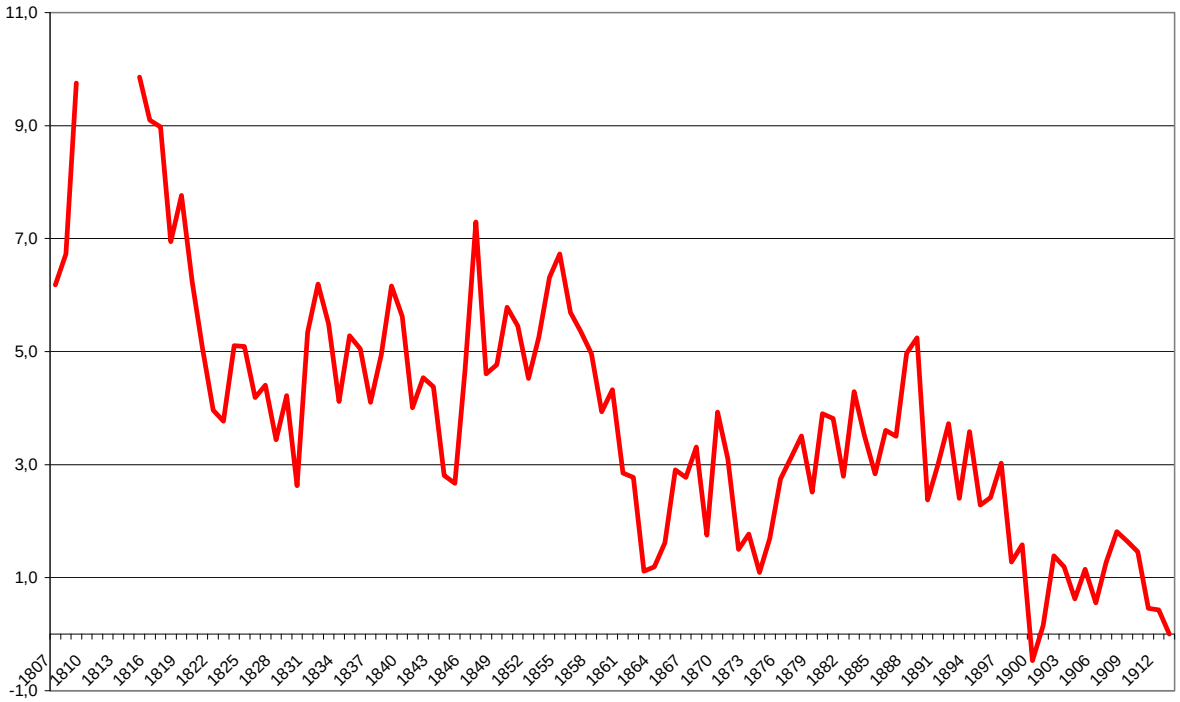
	8	9	10	11	12	13	14
1807	16,6	0,0	-33,5	-18,4	-11,0	-16,8	-13,9
1830	7,8	0,0	-25,6	-13,7	-6,0	-4,6	-5,3
1850	6,1	0,0	-19,0	-10,0	-4,6	-3,2	-3,9
1870	2,2	0,0	-8,3	-4,3	-2,6	-0,1	-1,3
1890	0,0	0,0	-4,0	-2,0	-2,1	-2,9	-2,5

Note: The differences in GDP levels are calculated as follow. With each deflator (1-14) we estimated GDP levels. These levels were compared to the one we arrived at using method no 14 (periodically changing weighting schemes of a Paasche type). The difference we found was expressed as a percentage of the GDP level according to method 14.

Table 7 shows that in case deflation procedures are applied to a larger number of underlying price indices, the margins of error are considerably larger. In case of a three-sector deflation procedure, a 6% difference between various types of deflation is the maximum. In the case of a more extensive deflation procedure –working with ca. 25 underlying price series- the differences are much larger. In case of GDP estimates for the year 1807, the use of fixed 1910 prices results in a GDP level 16,6% higher, whereas Fischer indices which are based on re-weighted and linked time series, result in GDP levels that are 10-15% lower than the original estimates. The margins of error are considerable.

Graph 1 reports the difference between two extreme types of deflation procedures. The ‘simple’ procedure applies to a three-sector model, with a fixed price set for 1910. This scenario is chosen as for many less developed economies or other regions with a lack of primary statistics, this might be the maximum level of sophistication that can be expected. The ‘sophisticated’ procedure relates to a Paasche price index, with a periodical change of weighting schemes. Graph 1 shows how much higher or lower GDP levels are according to the ‘simple’ procedure. This difference is expressed as a percentage of GDP calculated according to the ‘sophisticated’ procedure.

Graph 1: Differences in GDP levels according to two different deflation techniques, 1807-1913 (in %).



This graph shows that these two deflation techniques yield different results. In the early nineteenth century the 'simple' deflation procedure results in an over-statement of GDP of more than 9%. Besides, this graph indicates that the difference between the two deflation procedures does not change smoothly over time. Long-term trends as well as annual fluctuations are seriously influenced by the choice of deflation techniques.

5. The problem of PPP's

One of the key questions in economic history concerns the question which countries are at the technological frontier, and to what extent other countries are able to catch-up. In order to assess the relative economic strength of nations over long periods of time, we need to have reliable information about productivity differentials. Unfortunately, one of the main problems with the international comparison of GDP levels concerns prices. Well-known datasets such as the Penn World tables or Maddison's dataset, work with one set of relative prices (dollars of 1990). The use of such a recent set of prices might distort our view on the comparative productivity performance of countries in earlier time periods. It is necessary to make proper comparisons of purchasing power parities (PPPs) for earlier benchmark years, as relative prices have changed over time, .

Fortunately, in recent years a lot of interesting research is done in this field. First of all the work of Ward and Devereux should be mentioned.[14] Their new estimates of PPPs seem to have significant analytical implications concerning the relative strength of the British versus the United

[14] M. Ward and J. Devereux, 'Measuring British decline: Direct versus long-span income measures', *Journal of Economic History* 63-3 (2003) 826-851.

States economy in the nineteenth century. Also the re-estimation of East Asian PPPs by Fukao, Ma and Yuan opens new perspectives on the comparative economic performance of nations.[15] Both these studies construct historical PPPs from an expenditure perspective. Consumer prices are being selected and weighted on the share of the relating products in total consumption. This approach has the main attraction that with a relatively small sample of circa 20-30 products, reliable PPP estimates can be made. This method is aimed at measuring differences in *living standards*, as only final consumer goods are taken into account and consumer prices (including distribution costs and taxes) are being used.

The research group of the Groningen Growth and Development Centre follows a different route and estimates PPPs using the industry-of-origin method. This method is chosen as we are primarily interested in analysing *productivity* differentials, and therefore not only include final consumer goods, but also capital goods and intermediate products. Besides, we base our calculations not on expenditure- but on producer prices to get a clear view of the efficiency of production processes. The comparative productivity of economies can be analysed at quite a detailed level, as the PPPs are estimated at branch level.

Since 1983 a substantial research effort has been undertaken within the GGDC to carry out industry of origin comparisons of sectoral output and productivity across countries.[16] The International Comparisons of Output and Productivity (ICOP) project was initiated by Angus Maddison. The ICOP project now covers about 30 countries in Asia, East and West Europe, and North and South America. Comparisons of manufacturing output and productivity, which are based on unit value ratio comparisons and disaggregated into 16 manufacturing branches, are available for almost all of these countries. Comparisons for agriculture, transport and communication and wholesale and retail trade are available for a smaller group of countries. Comparisons for all sectors of the economy, adding up to total GDP, are available for Brazil, Japan, Korea, Mexico and the United States.

Over the years the ICOP has extended cooperation with other universities and research institutes which carry out ICOP-type productivity studies. These include the Eindhoven Centre for Innovation Studies (ECIS) at the Technical University Eindhoven, the National Institute for Economic and Social Research (London, the Centre d'Études Prospectives et d'Informations Internationales (Paris), and the Department of Business Studies at the Hong Kong Polytechnic University. Since 1997 the Economics Department of The Conference Board is sponsoring this programme, and makes use of its results for economic analysis.

A key characteristic of the ICOP is the use of industry-specific purchasing power parities (or unit value ratios) to convert output in national currencies to a common currency, e.g. US dollars. In contrast to other studies, these comparisons are therefore not based on exchange rates or PPPs for total

[15] K. Fukao, D. Ma and T. Yuan, 'Real GDP in Pre-War East Asia: A 1934-36 benchmark purchasing power parity comparison with the US' (research paper, january 2006)
[16] For more information about the ICOP project, see: http//www.ggdc.net

GDP, but take into account differences in relative price levels between industries. In most cases countries are compared on a binary basis with the USA as the numéraire country. In some cases (for example East European countries) comparisons have been made with West Germany; in some other cases with the United Kingdom (Spain and Portugal); in some other cases with France (Morocco, Tunisia). ICOP has mostly focussed on comparisons of value added per person employed or per hour worked. For some countries, the ICOP data base not only provides labour productivity, but also estimates of total factor productivity and unit labour cost. The ICOP project has put out over 80 research reports, in which methods, procedures and basic data are given, so that methods can be replicated and revised.

ICOP data for manufacturing, transport and communication and wholesale and retail trade are now also included in a regular publication of the International Labour Office (ILO), called Key Indicators of the Labour Market, 2001-2002. These estimates make use of ICOP's unit value ratios, but - in contrast to the estimates presented here - output and employment are directly taken from national accounts of individual countries instead of from industrial surveys and manufacturing censuses.

The Groningen Growth and Development Centre aims to extend the work on industry-of-origin PPPs and to construct benchmark estimates for earlier time periods. In this section new industry-of-origin PPPs are presented for the benchmark year 1910[17]. The industry-of-origin estimates should be seen as complements and not as substitutes of the expenditure-PPPs. At later stages, a detailed comparison of both types of PPPs should be made. This might reveal important information regarding international differences in price structure (such as the share of distribution costs and taxes).

On the basis of *industrial census* material, Burger estimated Unit Value Ratios (UVR's) for ca. 20 different industrial products. These product UVR's were aggregated to branch UVR's (weighted on the output value). Subsequently, these data were weighted to an aggregate sector estimate on the basis of the share of the various branches in total industrial value added. Branch PPPs could be calculated for mining, utilities, building materials, chemicals, textiles, food products, paper, leather and metal products.

The percentage of production matches ranges from 15-39% of the United K industrial production (and 15% for Germany, 23% for France, 31% for the Netherlands, and 39% for the United

[17] The work on industry is derived from: A. Burger, 'A five-country comparison of industrial labour productivity 1850-1990', paper presented on the seminar *Comparative historical national accounts for Europe in the 19th and 20th centuries* (Groningen, 1994); For the service estimates, see: A. Burger and J.P. Smits, "A benchmark comparison of service productivity between Europe and the United States for 1910", *Economic and Social History in the Netherlands*, special issue: Historical National Accounts in the Netherlands 7 (1996) 133-159.

States). This percentage is comparable with the standards achieved in modern-day research in productivity comparisons.[18]

Table 8 presents the outcomes of this new benchmark estimate of comparative labour productivity.

Table 8: Comparative labour productivity in industry in 1913: extrapolations and direct estimates (UK=100)

	Germany	France	Netherlands	United States
Burger	88	79	78	248
Broadberry	119	79	103	213

Source: Burger, 'a five-country comparison'; S.N. Broadberry, 'Historical patterns of comparative productivity in manufacturing' (mimeo, Warwick, 1993).

The new, direct estimates and Broadberry's extrapolated figures agree reasonably for the United States and closely for France. The differences for Germany and the Netherlands are quite large. Burger thinks that these differences are due to problems with the historical time series which were used for extrapolation. A sensitivity analysis shows that using alternative time series for the numéraire country, the United Kingdom, indeed leads to substantial differences. Therefore, the quality of various benchmark estimates can hardly be evaluated without involving the quality of the time series into the discussion.

Estimates on comparative labour productivity in services around 1910 are estimated by Burger and Smits. Their assessment of comparative productivity is based on a combination of UVR's and so-called Quantity Relatives (in the last case comparisons are made on the basis of output volume per worker). Even though the use of Quantity relatives is not ideal, at least it allows us to include this important sector the analysis.

Table 9: Comparative productivity per branch at the geometric average of national and United Kingdom weights (indices; United Kingdom=100)

	Transport	Communication	Trade	Government	Domestic serv.
France	102	73	90	71	66
Germany	101	79	55	131	78
The Netherlands	102	81	118	87	66

[18] B. van Ark, *International comparisons of output and productivity. Manufacturing productivity performance of ten countries from 1950 to 1990* (Groningen, 1993), pp. 18-23.

United States	155	142	96	59	84
United Kingdom	100	100	100	100	100

Source: Smits and Burger, 'A benchmark comparison', p. 151.

The United States have a clear lead in transport, especially due to the high relative productivity levels in railways. Probably the high efficiency in railways is caused to the long average distance of freight transport which undoubtedly resulted in scale economies. Also in case of communication, the United States clearly outperform the other countries, with a lead of 30-50% over the United Kingdom. The trade sector presents a different picture, with the Netherlands as the productivity leader. This phenomenon can be explained from the strong contribution of foreign trade tot total distribution in this small and open economy. Measured in terms of trade margins, Dutch productivity levels in trade were 18% higher than in the United Kingdom. The United States performed similar to the UK, whereas France and Germany were clearly lagging behind. For the other services (government as well as domestic servants), value added per worker was deflated with cost of living indices, assuming that differences in real wages actually reflect productivity differentials (a heroic assumption!).

Therefore, in table 10 two sets of aggregate service productivity estimates are presented. The basic estimates include government and domestic servants. In the alternative estimates the productivity of these last two categories is set at 100 in all countries.

Table 10: Comparative productivity in services in 1910 (indices; United Kingdom=100)

	Basic estimates		Alternative estimates	
	national weights	UK weights	national weights	UK weights
France	84	95	91	103
Germany	91	86	84	86
The Netherlands	106	112	110	118
United States	124	100	129	106
United Kingdom	100	100	100	100

Source: Smits and Burger, 'A benchmark comparison', pp. 152-153.

Table 10 clearly indicates that around 1910 the United States was the productivity leader in services, followed by the Netherlands which enjoyed a productivity level that was 6% higher than the figure for the United Kingdom. Labour productivity in services was lower in France and Germany, but the productivity gap between leader and followers was much smaller than in the case of industrial production.

Table 11 combines the estimates for industry and services, and also includes data for agriculture. These data do give an indication of differences of economy-wide comparative labour productivity on the eve of World War I. The new, direct benchmark estimates are compared to Maddison's estimates expressed at constant dollars of 1990.

Table 11: Comparative economy-wide productivity in 1910; GDP per worker compared to Maddison's GDP per capita. (indices, 1910=100)

	National	UK	Average	Maddison	Difference
France	79	77	78	64	22%
Germany	88	75	82	73	12%
The Netherlands	95	87	91	82	11%
United States	144	114	129	108	19%
United Kingdom	100	100	100	100	

Source: Burger and Smits, 'A benchmark comparison', p.154.

The direct estimates indicate that the productivity gap between continental Europe and the United States is smaller than the Maddison data show. Besides, the productivity lead of the Unites States over the United Kingdom is much larger.

Of course, this new view on productivity differentials around World War I may also shed a light on comparative productivity performance in earlier periods of time. Table 12 presents the income differentials –measured in terms of GDP per caput- in the period 1820-1910. These estimates were made by extrapolating the new 1910-estimates on the existing time series as presented by Maddison (=panel B). These data are compared with the original estimates made by Maddison himself (see panel A).

Table 12: Difference in income levels, 1820-1910 (income per head of population as a percentage of the UK level)

Panel A:

	1820	1850	1870	1890	1910
France	67	69	59	59	64
Germany	63	61	58	61	73
Netherlands	108	102	86	83	82
United States	74	77	77	85	108
United Kingdom	100	100	100	100	100

Panel B:

France	71	74	63	64	78
Germany	71	69	65	68	82
Netherlands	119	113	96	92	91
United States	88	93	92	101	129
United Kingdom	100	100	100	100	100

Especially a comparison between the three technology leaders of the period 1700-1913, i.e. the Netherlands, the United Kingdom and the United States is interesting. The new data indicate that even as late as 1850, in the Netherlands levels of GDP per head of population were more than 10% higher than in the United Kingdom. Besides, these data also shed a somewhat different light on the forging ahead of the United States. The new data show that already in the 1880s income per head of population in the US was higher than in the UK. Therefore, these new data come closer to the

'revisionist' view of Ward and Devereux, even though they argue that already at a much earlier stage the US surpassed the British income- and productivity levels[19]. The difference in the timing of the overtaking can be caused by numerous factors, of which the possibility of flawed time series (in the lights of the problems with services as well as with deflation techniques) should not be underestimated.

6. Concluding remarks

This paper gave an overview of the main challenges that we are facing when trying to make historical national accounts more comparable. Previous generations of economic historians have accumulated a rich amount of data. This material has been systematically processed, presented and analysed by Angus Maddison.

This paper points at three areas of concern when it comes to standardising historical national accounts, i.e. the treatment of services, the deflation of time series, and the use of recent sets of purchasing power parities. In all three cases, a sensitivity analysis was presented which showed the (potential) problems which may arise when working with the present, non-standardised, data. If we are not careful we might end up interpreting differences estimation techniques, rather than differences in growth and development.

However, this paper also indicates that many of the problems can be overcome. In the case of services additional research into (relative) prices and into changing input/output relations, might substantially improve the quality of the older sets of service estimates. Also the construction of historic benchmark PPPs may enhance a better comparison of income- and productivity levels over time. This research is time consuming, but recent work in this field shows that it is a fruitful area of research. Last but not least, the issue of deflation should be mentioned. Real GDP growth rates are strongly affected by the deflation techniques that have been used. Until now research has strongly focused on benchmark estimates of price levels, but the sensitivity analysis presented in this paper indicates that also additional research needs to be done in order to construct reliable time series.

[19] M. Ward and J. Devereux, 'Relative British and American income levels during the first industrial revolution' (Baltimore-New York City, 2004)