Maddison style estimates of the evolution of the world economy. A new 2020 update[1]

Jutta Bolt and Jan Luiten van Zanden

October 2020

[1] This update is the result of a long search for the best way to continue the pioneering work by Maddison, which has produced another update, Bolt et al. 2018, which sparked a debate which eventually led to this new version; many colleagues and members of the Maddison project were involved, but special thanks go to Bart van Ark, Steve Broadberry, John Devereux, Giovanni Federico, Alan Heston, Peter Lindert, Branko Milanovic, Nuno Palma, and in particular Robert Inklaar and Herman de Jong, the two (other) co-authors of the 2018 working paper; all errors are ours, of course.

1. Introduction

In this paper we explain the changes in the 2018 revision of the Maddison project, and briefly discuss their implications, in particular for the process of economic growth in the (distant) past. The Maddison project is an ongoing research project aimed at standardizing and updating the academic work in the field of historical national accounting in the tradition of the syntheses of long term economic growth produced by Angus Maddison in the 1990s and early 2000s. It is collective work, building on the research of a large group of economists and economic historians who focus on the growth trajectory of individual countries or regions, or on the further refinement of methods for estimating this in a systematic way. In 2014 we published the first update, which has been widely used, and this paper is part of the 2018 revision that has been completed now (together with Bolt et al., 2018).

The primary aim of the revision of 2018 was to incorporate the results of the 2011 ICP round. It resulted in the Bolt et al. 2018 revision that systematically followed the approach set out in Feenstra, Inklaar and Timmer (2015) and as implemented in the Penn World Tables. For economists, who prefer a theoretically grounded way of approaching the measurement issues involved, this was a satisfactory solution to the 'perennial' PPP problem. As economic historians we started to test the validity and the plausibility of the outcomes acquired in this way. The result was this paper, which develops standard for the testing of the outcomes of the various possible approaches in this field. We take a step back and start by exploring the effects of incorporating different sets of PPP's for estimating relative levels of GDP per capita in the distant past. The first PPP we consider is the 'traditional' Maddison approach uses the 1990 benchmark as introduced in Maddison (1995), and which currently is the benchmark used in first update of the Maddison project (Bolt and Van Zanden, 2014). The second PPP benchmark we incorporate is the most recent PPP estimate published by the World Bank, based on the 2011 ICP round. The last set of PPP's we use is based on the estimates by Penn World tables (PWT) for the period from 1950 to the present, linking the historic (pre-1950) time series to the PWT estimates in that year.

The aim of this paper is twofold. First, we present (as we do in each new update) all new historical income series that have become available since the previous update of the Maddison project. Second, and this is the main contribution, we aim to find out how large the biases are from the various PPPs for plotting long run income series in global set of countries. In this paper we therefore develop a way to determine the magnitude of the biases using the various PPPs. We start by collecting all available direct and indirect benchmark estimates of relative

income levels for groups of countries in the 19th and early 20th century from the literature. We then compare the relative income levels using the different PPPs, with the relative income levels as indicated by the independent (direct) and indirect benchmarks. Based on this comparison we test whether a revision based on the 2011 PPPs or on the PWT approach would produce better estimates of relative levels of GDP per capita in the period before 1940 than the 'traditional' Maddison approach that uses his 1990 benchmark.

The focus of the paper is on the consequences of these changes for the historical period – for the years before 1950. It is meant to explain to an audience of economic historians why we made such radical changes in the Maddison dataset, and in this way supplements and builds on the Bolt et al. (2018) paper. After explaining the choices made, we also discuss some of its implications – in terms of the relative and absolute growth of countries – and briefly deal with its weaknesses and strengths, including a few suggestions how this might even be improved in the future. In the remainder of this paper we will first, in Section 2, we review the aims of the Maddison project, and in section 3 we will introduce the newly included historical income series used to update the Maddison project database. This is followed by a discussion of the three different PPP benchmarks, the 1990 benchmark, the 2011 benchmark and the multiple benchmark that are scrutinized in this paper. In section 5 we compare the implications of the three alternative benchmarks for selected patterns in long-run economic development. We analyse the effects of using alternative benchmark approaches for historical income series, with respect to the level of subsistence incomes, and with respect to the comparison with independent and indirect benchmarks for earlier time periods. The paper ends with concluding remarks in section 6.

2. The Maddison project

It is important at this stage to explicitly state the goals of the Maddison project and its product, the Maddison dataset. It aims at charting the long term trends in the world economy over the past millennium, in order to be able to analyse the determinants of growth and stagnation in the world economy. Angus Maddison was interested in the big picture and the long view, but also took great care to link his results to the most recent advances in the measurement of the wealth and poverty of nations. The Maddison project dataset therefore has to fulfil two conditions: for the recent period it has to accurately reflect international differences in GDP per capita

(reflecting the state of the art estimates), and it has to summarize the available information about historical patterns of growth and decline in the best way possible.

The Maddison database on Historical Statistics of the World Economy has probably the widest coverage of data on GDP per capita across countries and over time currently available, including information for over 160 countries and covering the period from Roman times to the present. To compare income levels and developments for this period and set of countries, national income estimates are converted from a national currency basis to a common currency using purchasing power parities (PPPs)[2]. PPPs measure price differences between countries and therefore represent what a dollar of income buys in one country relative to another. As relative price levels are lower in less advanced economies, exchange rates – the alternative for exchanging national currencies into a common currency – typically understate the real (or cross-country comparable) GDP level of such economies relative to the richer ones.[3] But while the principle of using PPP-converted income levels for cross-country comparisons is clear, the choice for a particular method of estimating PPPs is more difficult to make, yet of first-order importance in many debates on historical living standards.

For scholars interested in income levels over a period of time, the challenge is to estimate PPPs for years other than the ICP benchmark year. There are three general methods that can be used for this purpose:

1. The *extrapolation* (or *projection*) method.
2. The *multiple/historical benchmark* method, and
3. The *econometric* method.

The current version of the Maddison database[4] relies on the extrapolation method, under which GDP or GDP per capita levels from the benchmark year 1990 are extrapolated to earlier and later years using growth rates from country National Accounts (or other growth estimates).[5] This method implicitly assumes that changes in PPPs over time are well-approximated by relative inflation rates. Since PPPs aim to measure the relative price level between countries,

[2] These purchasing power parities are not a regular part of official national accounts, but their compilation has evolved over time from academic research into a global statistical program, called the International Comparisons Project (ICP). The latest ICP round, with 2011 as the benchmark year and comprehensive coverage for 180 countries, was coordinated by the World Bank with regional support from agencies such as the OECD, Eurostat and national statistical agencies, see World Bank (2014).
[3] See Summers and Heston (1991) and Feenstra, Inklaar and Timmer (2015).
[4] As well as earlier version of the Penn World Table, see Summers and Heston (1991).
[5] See Bolt and van Zanden (2014).

there is a conceptual logic to this implicit assumption, but over time it may lead to unreliable results due to shifting economic structures and relative prices.[6] In addition there are a host of practical challenges in cross-country price measurement that makes reliance on a single benchmark estimate hazardous.[7]

The second and third methods have been developed as alternatives to the extrapolation method. The econometric method estimates relative price levels for historical periods using the contemporaneous relationship between price levels and variables that are observable over long time periods.[8] [9] The final method relies on either multiple PPP benchmarks or on historical PPP benchmarks. The most recent version of the Penn World Table (Feenstra, Inklaar and Timmer, 2015), which covers the period since 1950, takes as the 'best' estimate of the PPP in, for example, 1980 the value based on the ICP round for 1980, while for 2005, they rely on PPPs from ICP 2005.[10] The historical benchmark method takes this logic to earlier periods, estimating PPPs based on available price and quantity information in an earlier year between a pair of countries.

The need to regularly revise the Maddison dataset is caused by the ongoing research in this field, which results in new series of estimates of historical national accounts, in new estimates of benchmarks comparing levels of GDP of different countries in the past, and in new ICP rounds that result in updated estimates of PPPs of countries. One of the weaknesses of the Maddison dataset was that he made use of a set of PPPs from 1990 which was based on often limited information. The 2005 ICP round did however not produce the new set of PPPS which could be used for this purpose, explaining the delay in replacing the 1990 benchmark.

The 1990 benchmark was state-of-the-art at the time of creation, and its results have served as a point of reference in many studies since. Yet, the quality of PPP measurement has improved considerably, as especially the number of countries included increased during consecutive rounds and the methodology employed in their construction became much more sophisticated over the years. The 2011 ICP results are generally considered the best available (Deaton and

[6] See e.g. Inklaar and Rao (2017) for a more precise formulation.
[7] See e.g. Deaton and Heston (2010).
[8] See Prados de la Escosura (2000) or Klasing and Milionis (2014) for a recent analysis.
[9] Due to a fundamentally different approach to estimating long run estimates we do not use the econometric or indirect method in this paper as one of the alternatives but we do exploit its information to cross-check the plausibility of the alternative extrapolation methods.
[10] PPP estimates between benchmarks are interpolated; when no earlier or later ICP benchmark is available, the extrapolation method is used.

Aten, 2017). However, by using the extrapolation method for creating comparable historical income series, one assumes that the underlying price structure of each economy is fixed over time. As a result, a comparison of GDP levels, further away from the benchmark year becomes affected especially if the comparative price structures of the countries included in the comparison change in very different ways. Hence, the further one moves away from the benchmark year, and/or the more different the price structures between countries evolve, the less reliable the comparisons generally become. An alternative therefore would be to extrapolate income series backwards from the earliest available benchmark for each country, using the PWT approach.

Our project to (possibly) rebase the Maddison project dataset has to strike a balance between these two opposite forces. The necessity of rebasing the estimates is first of all the result of the development of a new set of PPPs for 2011, which results in new – and, we think, probably more accurate – estimates of relative and absolute income levels for the most recent period. The next issue is then how to incorporate this new information into the carefully crafted historical framework as developed by Maddison (and by and large copied by the first update of Maddison project in 2014). We already discussed that there are in principle three alternatives for combining the recent PPPs with the time series of GDP growth: making use of the 2011 PPPs, stick to the 1990 PPPs, or use all available ICP benchmarks. What we need is a yardstick to measure the accuracy of the outcomes of the three scenario's. To compare the implications of each alternative for reconstructing global inequality over time we have defined two criteria to test the accuracy (or the impact of the cumulative biases) of each alternative set of relative prices.

First, we need to establish the extent to which the combinations of benchmark and time series results in plausible historical outcomes, such that for example countries do not end up with incomes (far) below subsistence for prolonged periods of time. Second, we compare the relative performance of countries during the 19th and early 20th century as given by the alternative datasets discussed above, with the relative performance of countries as indicated by the newly created dataset of independent and indirect (econometrically estimated) benchmarks. For this we have collected a new dataset which includes all independent historical benchmarks of relative GDP levels of countries available in the literature. This comparison can tell us which dataset 'predicts' these independent relative output levels more accurately, and to what extent these alternatives result in really different stories of global inequality? Additionally, we examine how the three scenarios compare with the results of the *econometric* (indirect) method,

as suggested by Leandro Prados de la Escosura (2000), which can be seen as an alternative source of estimates of relative income levels in the 19th and early 20th century.

3. Updating historical series[11]

This new version of the MPD extends GDP pc series to 2016 and includes all new historical estimates of GDP per capita over time that have become available since the previous update (Bolt and Van Zanden, 2014). As new work on historical national accounts appears regularly, a frequent update to include new work is important as it provides us new insights in long term global development. Further, we have incorporated all available annual estimates for the pre 1820 period, instead of estimates per (half) century as was usual in the previous datasets.

For the recent period the most important new work is Harry Wu's reconstruction of Chinese economic growth since 1950, a project inspired by Maddison which produces state of the art estimates of GDP and its components for this important economy (Wu, 2014). Given the large role China plays in any reconstruction of global inequality, this is a major addition to the dataset. Moreover, as we will see below, the new results show that the revised estimates of annual growth are in general lower than the official estimates. Lower growth between 1952 and the present however substantially increases the estimates of the absolute level of Chinese GDP in the 1950s (given the fact that the absolute level if determined by a benchmark in 1990 or 2011). This helps to solve a problem that was encountered when switching from the 1990 to the 2011 benchmark, namely that when using the official growth estimates the estimated levels of GDP per capita in the early 1950s are substantially below subsistence back until 1890, and therefore too low. Including the new series as constructed by Wu (2014) gives us much more plausible long run series for China.

Most of the other additions to the Maddison project dataset relate to the period before 1914; see Table 2.1. Again, important new work has been done for China, in particular the papers by Xu et al. (2016) and Broadberry et al. (2018). It is reassuring that these two independent teams of scholars who set out to quantify Chinese economic growth before 1900 produced very similar estimates, showing a strong decline (by about one third) of GDP per capita in the 18th century and quasi stability in the 19th century.

[11] This section is based on Bolt et al. (2018).

Often, studies producing early per capita GDP estimates make use of indirect methods, particularly work on the early modern period (1500-1800). The 'model' or framework for making such estimates is based on the relationship between real wages, the demand for foodstuffs and agricultural output (Malanima (2010), Alvarez-Nogal and Prados de la Escosura (2013) among others). This model has now also been applied to Poland (Malinowski and Van Zanden 2016), Spanish America (Arroyo-Abad and Van Zanden 2016), and France (Ridolfi 2016). In this update we have now included annual estimates of GDP per capita in the period before 1800 for these countries.

Table 1. New Additions to the Maddison Project Database

Country	Period	Source
Latin America		
Bolivia	1846–1950	Herranz -Loncán and Peres-Cajías (2016).
Brazil	1850–1899	Barro and Ursúa (2008).
Chile	1810–2004	Díaz Lüders and Wagner (2007)
Cuba	1690–1895	Santamaria Garcia (2005).
Cuba	1902–1958	Ward and Devereux (2012).
Mexico	1550–1812	Arroyo Abad and Van Zanden (2016).
Mexico	1812–1870	Prados de la Escosura (2009).
Mexico	1870–1895	Bertola and Ocampo (2012).
Mexico	1895–2003	Barro and Ursúa (2008).
Panama	1906–1945	De Corso and Kalmanovitz (2016).
Peru	1600–1812	Arroyo Abad, and van Zanden (2016).
Peru	1812–1870	Seminario (2015).
Uruguay	1870–2014	Bertola (2016).
Venezuela	1830–2012	De Corso (2013).
Europe		
England	1252–1700	Broadberry, Campbell, Klein, Overton and van Leeuwen (2015)
Finland	1600–1860	Eloranta, Voutilainen and Nummela (2016).
France	1250–1800	Ridolfi (2016)
Holland	1348–1807	Van Zanden and van Leeuwen (2012)
Italy (north)	1310-1871	Malanima (2010)
Norway	1820–1930	Grytten (2015).
Poland	1409–1913	Malinowski and Van Zanden (2017)
Portugal	1530–1850	Palma and Reis (2019).
Romania	1862–1995	Axenciuc (2012).

Spain	1850–2016	Prados de la Escosura (2017).
Sweden	1300–1560	Krantz (2017).
Sweden	1560–1950	Schön and Krantz (2015).
Switzerland	1850-2011	Stohr (2016)
UK	1700–1870	Broadberry, Campbell, Klein, Overton and van Leeuwen (2015)
Asia		
China	1661–1933	Xu, Shi, van Leeuwen, Ni, Zhang, and Ma (2016) and Broadberry, Guan and Li (2018)
China	1952–2008	Wu (2014).
India	1600–1870	Broadberry, Custodis and Gupta (2015).
Japan	724-1874	Bassino et al. (2018)
Japan	1874-1940	Fukao et al. (2015)
Korea, Republic of	1911-1990	Cha, Kim, Park and Park (eds.) (2020)
Korea, DRP of	1911 – 1940; 1990-2015	Cha, Kim, Park and Park (eds.) (2020)
Malaysia	1900-1939	Nazrin (2016)
Turkey	1500-1820	Pamuk (2009).
Singapore	1900–1959	Sugimoto (2011)
Middle East		
Syria	1820, 1870, 1913, 1950	Pamuk (2006).
Lebanon		
Jordan		
Egypt		
Saudi Arabia		
Iraq		
Iran		
Africa		
Cape Colony/ South Africa	1700–1900	Fourie and Van Zanden (2013).

For some countries for the period before 1870 or 1800 we only have series of a certain province or similar entity. The British series switches to England in 1700, the Dutch to Holland in 1807, the Italian to Northern Italy in 1871. The switch from the national to the 'partial' series is clearly indicated in the dataset, and the 'correction' in terms of GDP per capita is indicated. For Italy in 1871 we made use of the estimate by Felice (2018).

Finally, we have extended the national income estimates up until 2016 for all countries in the database. For this we use various sources. The most important is The Total Economy Database

(TED) published by the Conference Board, which includes GDP pc estimates for a large majority of the countries included in the Maddison Project Database. The same approach was followed in the 2014 MPD update (Bolt and Van Zanden, 2014). For countries not present in TED, we relied on UN national accounts estimates extend the GDP per capita series. To extend the population estimates up to 2016, we used the TED and the US Census Bureau's International Data Base 2016.[12] The TED revised their China estimates from 1950 onwards based on Wu (2014). As discussed above, we also included Wu's (2014) new estimates in this update. Last but not least, we have extended the series for the former Czechoslovakia, the former Soviet Union and former Yugoslavia, based on GDP and population data for their successor states.

3. The benchmark options

3.1 The 'original' Maddison 1990 benchmark

For his latest dataset Maddison focused his comparisons mainly on the reference year 1990, which became the 'interspatial-intertemporal anchor' for his world estimates and this is still used in the updates provided by the Maddison Project dataset.[13] In practice Maddison could not always use PPPs based on 1990 relative prices, because many countries were not covered by a PPP exercise in 1990, so he had to make links with countries that were covered in earlier ICP rounds (see Maddison, 2006, p. 610). In his 1995 study Maddison made use of the 1990 ICP 6 round for the 22 member countries of the OECD and presented the results together with earlier ICP rounds. Many non-OECD countries were covered in one of the other ICP rounds, for example in 1975, 1985 or 1993. In the final comparison, data for 43 countries (representing almost 80 per cent of world GDP at the time) was based on ICP or ICP-equivalent estimates. The other 113 countries ('non-sample') were covered by PPPs from PWT and by proxy

[12] As Palestine is not included in these sources, we used data from the World Development Indicators.

[13] In *Phases of Capitalist Development* Maddison used PPPs for the year 1970. Later, in *Dynamic Forces in Capitalist Development* he used PPP's for 1985. Both PPP's were Paasche PPPs, measuring output at US relative prices, as in his view this best resembled the price structure to which the other countries were converging. Secondly, the price structure would correspond to "an identifiable reality" (Maddison, 1991, p. 196). In his later publications, Maddison switched to comprehensive long-term comparisons of countries' GDP on the basis of multilateral purchasing power parities for the year 1990. These comparisons were published in three successive volumes that (more or less) build upon each other, starting with the publication of *Monitoring the World Economy* (1995), followed by *The World Economy: A Millennial Perspective* (2001) and, finally, *The World Economy: Historical Statistics* (2003). The 2001 and 2003 volumes contain updates and revisions of the series and comparisons of the 1995 publication, and were reprinted and further updated in *The World Economy* (2006) and *Contours of the World Economy, 1-2030 AD* (2007).

estimates. The PWT PPPs were, in turn, estimated by Summers and Heston (1991) based on cost-of-living estimates relevant for expatriates and foreign diplomats.

Maddison showed that alternative estimators from different ICP rounds for non-OECD economies lead to larger differences than for OECD countries. The range between different converters is generally larger in the non-OECD countries, because price structures differ more from those in the United States. China deserves special mention here, due to its size and because China did not participate in any of the ICP rounds before 2005. Maddison collected the data for his 1990 reference estimate from four different estimates of 1990 GDP per capita in GK dollars, ranging from $1,135 to $4,264. Later on Maddison adjusted the Chinese per capita GDP level to $1,871 in 1990, based on an estimate for China by Ren (1997) to which he made some significant adjustments (Maddison, 2007a, 154-155).

Even though there had been new ICP rounds (ICP 7 in 1993 and ICP 8 in 1996), Maddison continued to use 1990 as the reference year for his comparisons as he saw considerable problems in putting together all the regional ICP estimates on a comparable basis. In his view it would have been a complex exercise to switch to later reference years on a consistent basis (Maddison, 2006, p. 172).

To summarize, the set of 1990 "Geary-Khamis PPPs" that serve as the 'anchor' for the Maddison data have shortcomings because of data limitations. Most notably, data from actual ICP price surveys was available for less than one-third of the countries in his dataset, with PPP estimates for the remainder based on surveys designed for other purposes than comparing GDP levels across countries. For the one-third of countries with ICP data, the PPPs are drawn from several different benchmark estimates, often years apart, which all used different methods in collecting prices. So while Maddison's set of 1990 PPPs may have represented the best estimate back then, time and measurement has moved on.

3.2 From 1990 to the ICP 2011 global price comparison

In retrospect, the period when Maddison was developing his dataset were a low point in the history of the International Comparison Program: lacking sufficient funding and suffering from poor management, the 1993 round of ICP was only partially completed and widely seen as imperfect (Ryten, 1998). This provided the impetus for a better-funded, more widely-supported continuation of ICP, which led to the 2005 round of ICP. Compared to earlier global comparisons, this round was a major improvement, covering more countries than ever before

(146), including China and India.[14] Furthermore, the round featured more precise product specifications and a regional setup. Under the regional setup, countries would first compare prices within a region based on a regional product list, before the different regional comparisons would be linked based on prices from a global price list to allow for a global price comparison.[15] Both improvements should help mitigate one of the key challenges in international price comparisons, namely to find and price products that are comparable across countries yet representative of what consumers buy in each country. A more extensive discussion, as well as a critical reflection on measurement and methods in ICP 2005, can be found in World Bank (2008).

But while ICP 2005 represented an important step forward, concerns quickly arose about the results. Most notably, the relative income levels of lower-income countries, like China and India, were notably lower than had previously been thought. While that could have been the result of improvement measurement in ICP 2005, Deaton (2010) and Deaton and Heston (2010) raised the possibility of biases in measurement. They raised concerns about price surveys in China, which were only carried out in urban areas. If rural prices are lower, that would be a source of bias relative to the 'national average' prices used in other countries; see also Feenstra, Ma, Neary and Rao (2013). Deaton (2010) and Deaton and Heston (2010) were also concerned about the second stage of ICP 2005, in which regional comparisons are linked together, since any misalignment in this linking procedure would shift prices and income levels of whole regions relative to each other; Deaton (2010) refers to these as "tectonic regional PPPs". More recently, Inklaar and Rao (2017) have shown that the regional linking in ICP 2005 was indeed biased and that, as a result, the relative prices of countries in Africa and Asia were too high and thus their relative income levels too low.

Mindful of the weaker points of ICP 2005, measurement methods were modified and refined for ICP 2011 (see World Bank, 2014). Most notably, a more representative global product list was drawn up to be used in estimating the "tectonic regional PPPs". As shown in Inklaar and Rao (2017), the resulting regional linking in ICP 2011 does not suffer from the biases that were present in ICP 2005. Surveys in China were also more extensive, covering urban and rural areas in all provinces (ADB, 2014). Finally, country coverage was even more comprehensive with

[14] China had not participated in any of the earlier ICP rounds, while India had most recently participated in the 1985 round.
[15] The regional setup has a political background: Eurostat, as representative of the European Union countries, will only participate in a global price comparison if the relative prices between EU countries would not be affected.

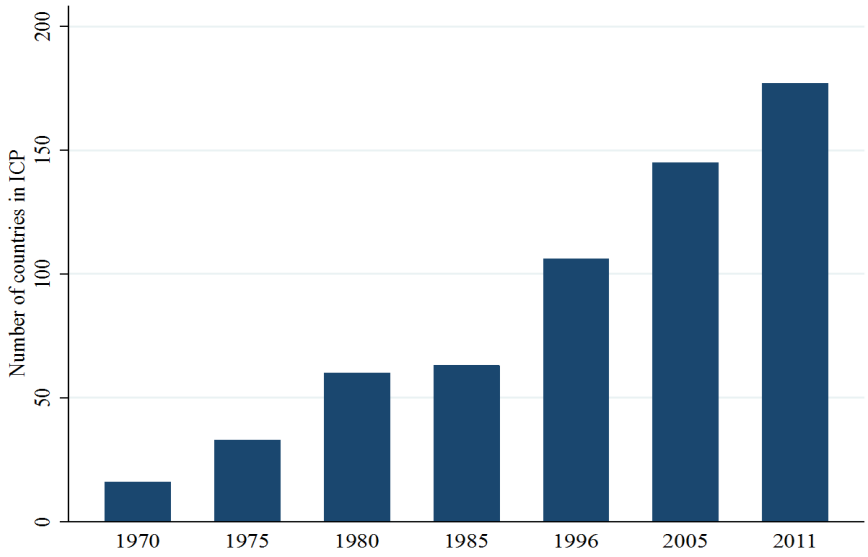
complete data for 177 countries. There are certainly areas of imperfect measurement with the prices for construction investment and for housing, government, health and education services viewed as 'comparison-resistant'. The 'output' of these activities is typically heterogeneous, hard to define and/or hard to price, so proxy methods are chosen (see World Bank, 2014). Still, these areas have been challenging since the start of ICP (see e.g. Kravis, Heston and Summers, 1982) and there is no indication that the measurement problems in this area are worse than used to be. In summary, the ICP 2011 PPPs represent the best-measured set of relative prices for the largest set of countries so far. In terms of measurement methods, ICP 2011 is better or at least no worse in all areas.

3.3 The PWT multiple benchmark approach

The third alternative that we explore in this paper is the methodology developed for version 8 of the Penn World Tables (Feenstra et al. 2015).[16] This approach makes use of all available post 1950 benchmarks and adapts the growth rates of GDP from SNA to fit the changing position of countries in the various benchmarks. There have been ICP rounds collecting prices in countries in the benchmark years 1970, 1975, 1980, 1985, 1996, 2005 and 2011. Over the years an increasing number of countries has participated in these global collection rounds, with 177 participating countries in 2011 being the most comprehensive round (see figure 1 below).

Figure 1: Historical global ICP participation[17]

[16] As indicated before, this paper only uses the indirect estimates to cross-check the extrapolation alternatives against.
[17] 1990 is not included as it was not a global round.



For each individual country, the PWT determines which benchmarks are available. These benchmarks are used directly as PPP converters. For the years between benchmarks, prices for final goods are interpolated between the benchmark years taking the benchmark price indexes as given and using the pattern of inflation from the national accounts in the intervening years to determine how the overall change in price indexes between these benchmark years should be distributed over the years in between (Feenstra et al. 2015, web appendix B: 5). For the years between 1950 and the first available benchmark for each country, prices are extrapolated using price deflators from the National Accounts. The same is done for years after the last benchmark (Feenstra et. al, 2015: 3167). For the purpose of this paper, the 1950 estimates obtained for each country using this methodology serve as the anchor from which the income estimates are extrapolated backwards using growth rates from existing series or from the newly included series as discussed above (Bolt and Van Zanden, 2014).

The PWT procedure gives a complete set of price indexes for all years and countries included in the dataset, providing the 'best' estimate of the PPP for each year. The consequence of fitting the national income series to the various benchmark estimates is that the resulting series are no longer equal to the original official National Account series as a result of the inherent inconsistency between benchmarks and time series[18]. The resulting pattern of income growth

[18] With a constant PPP approach (one benchmark as a base for extrapolating income estimates back in time) the underlying price structure of each economy remains fixed. A comparison of GDP levels further away from the benchmark year becomes affected especially if the comparative price structures of countries include of countries included in the comparison change in different ways. Further, inconsistencies can also result from measurement

as published by the PWT sometimes differs from growth as given by the National Accounts (for example the decline in GDP pc in Nigeria between 1977 and 1995 is much larger using the PWT approach than can be found in the official sources; other examples of substantial differences in growth rates between the PWT and National Accounts are Togo and Madagascar between 2005-2011).

5. Comparing the three approaches

In this section we will both use the latest PPP estimates produced by the World Bank for 2011 (World Bank, 2014) and the PWT benchmark approach to rebase the original Maddison database to both the 2011 relative prices and to (implicit) 1950 benchmark created by PWT benchmark for each country. Our starting point is the updated Maddison project dataset as presented above. Applying alternative reference years to the original GDP pc series involved two very simple steps. We first calculated yearly growth rates of the original series. Second, we use these growth rates to extrapolate the income series backwards using either the new 2011 PPP converted income measures taken from the PWT, or extrapolate income series backwards using the multiple benchmark approach as employed by the PWT (2015), in a similar fashion to Maddison (2003) and the Maddison Project in their 2013 update (Bolt and Van Zanden, 2014). This provides us with PPP converted GDP pc estimates expressed either in dollars and using 2011 constant prices, or constant prices from an earlier benchmark.

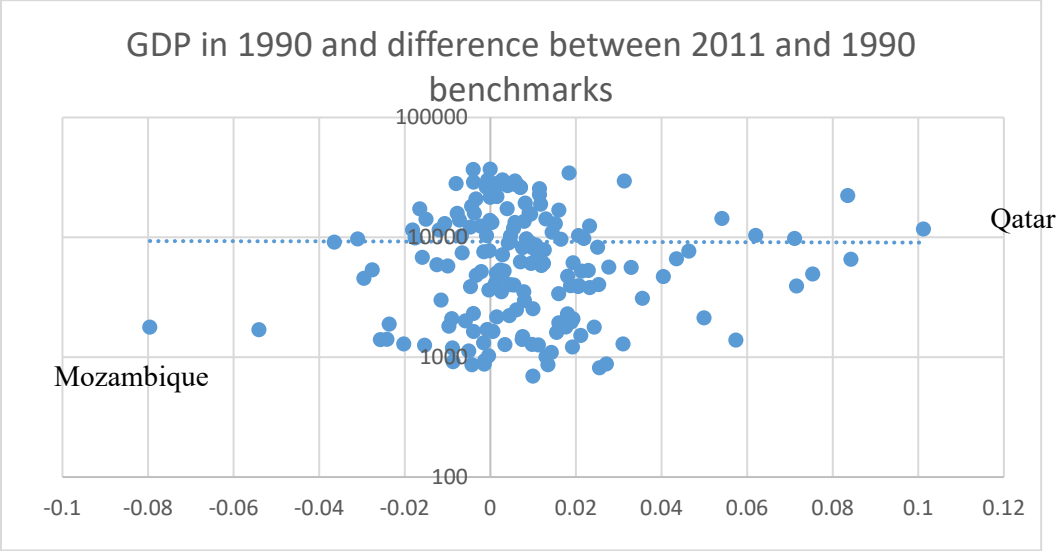
The outcome of our rebasing exercises essentially entails a level shift in income estimates. For example, those countries that experienced an increase in prices between 1990 and 2011 vis-à-vis the United States, original income series have been divided by higher relative prices in the 2011 PPP series, leading to a lower level of real income relative to the US. And for those countries that have experienced a decrease in prices during that period relative to the US, income estimates shift upwards. In a similar fashion, those countries that have experienced an increase in prices between the first benchmark available for that country and 1990 vis-à-vis the US, the level of real income relative to the US for the first benchmark was higher than their relative incomes in 1990. This leads to a reshuffling of the level of historical series for various

issues such as differences in methodologies between different PPP rounds and revision in time series which both influence the international comparison (Johnson, et al., 2009: 2013: Deaton, 2011: Kravis and Lipsey, 1989).

countries. For example it has an effect on the timing of the overtaking of the UK by the US, which we will discuss later on.

Figure 2 illustrates these shifts, by comparing the level of GDP per capita in 1990 (in 2011 prices) with the difference between the growth rate according to the national accounts between 1990 and 2011 and the actual increase in GDP per capita according to the benchmark estimates of 2011 and 1990. Qatar, one of the richest countries in 1990 and by far the most wealthy in 2011, has a strong positive difference, reflecting a rise in GDP per capita which is not explained by growth according to the national accounts. At the other extreme, Mozambique and Liberia, the countries with extremely low levels of real income, show very negative growth differences, pointing to the fact that despite some economic growth they were unable to improve their relative and absolute income level.

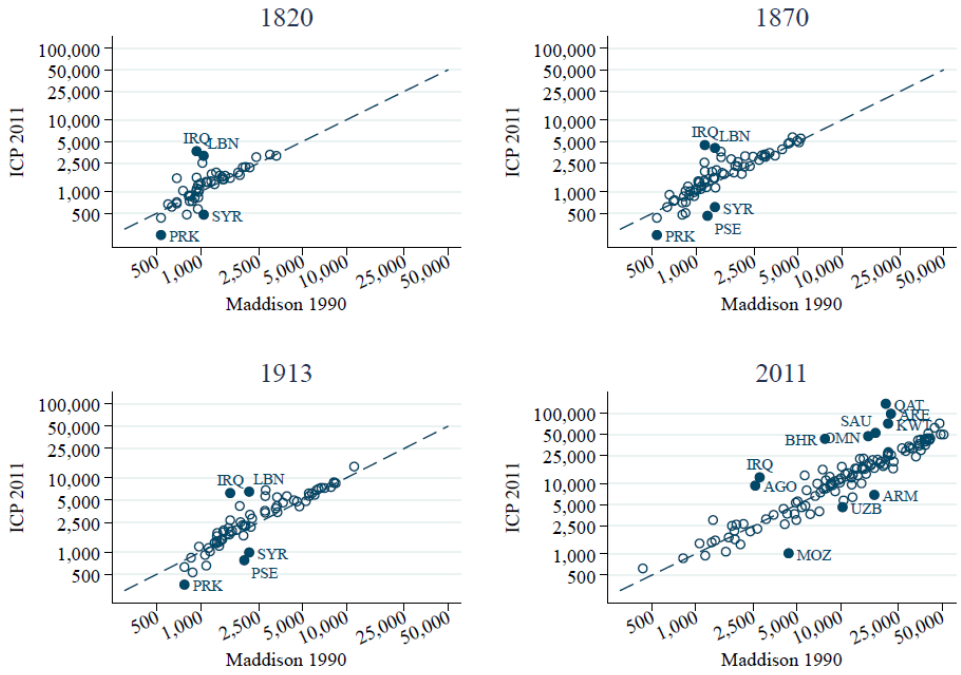
Figure 2: 1990 Income levels versus differences in growth rates 1990-2011[a]



[a] Comparison of GDP per capita in 1990 (2011 prices; vertical axis) and the difference between growth rate according to national accounts and the levels of GDP according to 1990 and 2011 benchmarks

Another way of presenting this comparison is by looking at the consequences of making use of the two benchmarks (2011 and 1990) for the historical estimates.

Figure 3: GDP per capita in various years based on Maddison 1990 versus ICP 2011 PPPs (in 2011 US dollars)[b]



[b] The original GDP per capita values based on Maddison's 1990 PPPs have been converted to 2011 US dollar values (to make the numeraire comparable) by multiplying all observations by 1.59, which corresponds to US inflation between 1990 and 2011

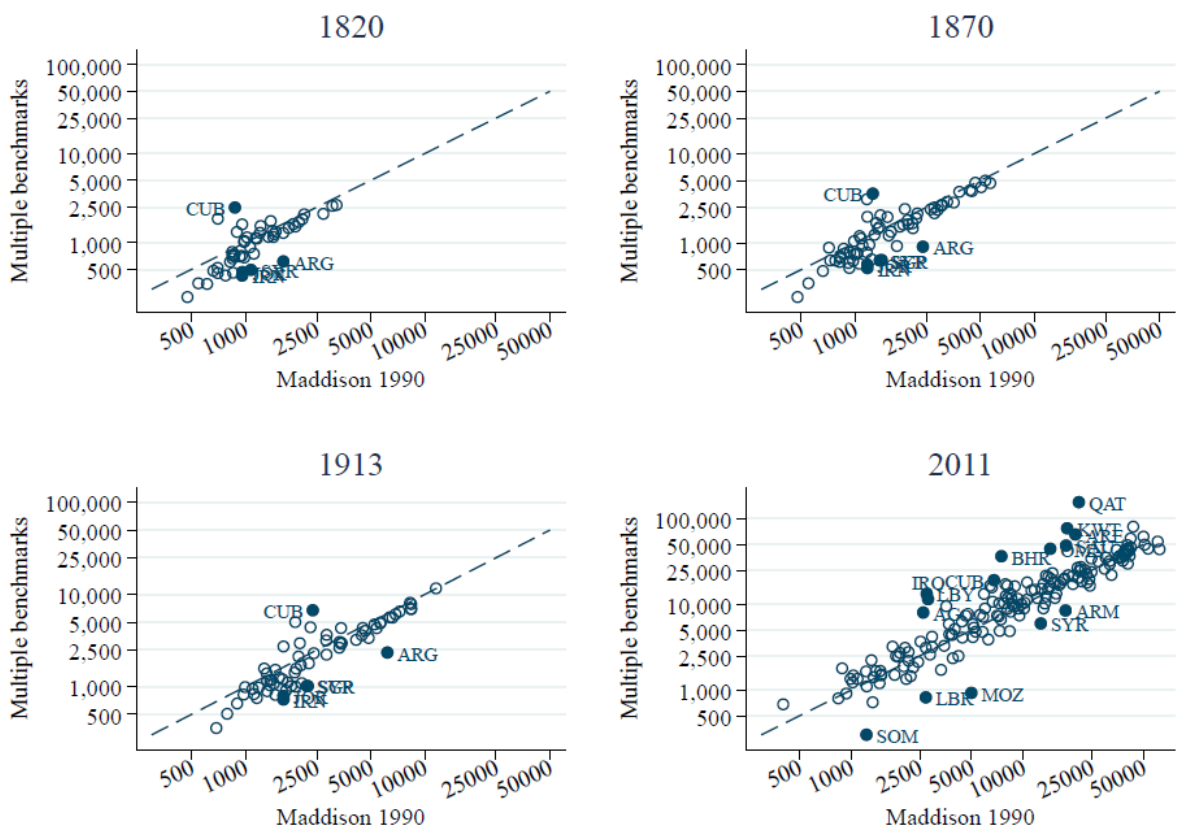
Again, the most notable exceptions for all years relate to the oil-rich countries in the Middle East: Qatar, Kuwait, United Arab Emirates, Bahrein, Oman and Saudi Arabia, Iraq and Syria. Their relative GDP per capita increases a lot when the switch is made to the 2011 benchmark. This may be explained by deficiencies in the Maddison 1990 estimated. These countries were in the original Maddison database part of the 'non sample' group, that is they were not included in the ICP rounds underlying Maddison's 1990 PPP calculations. Prevailing price estimates for 1990 were based on expat information and probably overestimated the actual prices in the economies, leading to an underestimation of actual income levels. Using more precise price estimates of the latest ICP rounds indeed shows a substantial upward shift in incomes for these countries.

But as argued already, the shift of the GDP levels of the Gulf states may also be caused by the strong increase in the relative price of oil between 1990 (when oil prices were rather low) and 2011 (when they peaked), which drove up the relative income levels of these oil producing countries. Extrapolating the 2011 levels back into the past on the basis of the estimated growth

rates of these countries may therefore result in implausible high incomes in earlier periods, as Figure 3 demonstrates. According to this combination of data (the 2011 PPPs and the estimated growth rates of GDP per capita) Iraq would be the most wealthy country in the world in 1820 (and close to that position in 1870), which is an implausible result. Other notable outliers are Public Republic of Korea and the Palestinian territory and Lebanon for 1820, 1870, and 1913 and Angola, Mozambique, Uzbekistan and Armenia for 2011.

Comparing the 1990 benchmark estimates to multiple benchmark estimates using the PWT approach, we find that most countries are located very close to the 45 degree line for all years analysed, see figure 4 below. Notable outliers for 1820, 1870 and 1913 are Argentina and Cuba. For Cuba the average per capita income is around one third of the level when based on the 1990 estimates compared to the PWT approach. In contrast, for Argentina income is around 3 times as large using the 1990 benchmark compared to the PWT approach. Likewise, incomes in Syria and Iran are also much higher using the 1990 levels compared to the PWT approach. For 2011, the graph is similar to the 2011 graph in figure 3.

Figure 4: GDP per capita in various years based on Maddison 1990 versus multiple benchmark (PWT) approach (in 2011 US dollars)



5.1 The first test: subsistence income[19]

One of the implications of the rebasing exercise is that the poverty level has changed. With the 1990 price levels, subsistence level income was between 350 and 400 international dollars per year (Maddison, 2003). The poverty line therefore was equal to around 1 dollar a day, and was based on the first international poverty line which was set at $1.01 per day using 1985 PPP's, which was later updated to $ 1.08 per day using 1993 PPP's (Ravallion, Datt and van de Walle, RDW, 1991; Chen and Ravallion, 2001). This made the interpretation of historical income series very intuitive. By using other relative prices, this subsistence level of income changes. The World Bank defined the absolute poverty line at 1,90 US dollars a day or 694 dollars per year, expressed in 2011 prices.

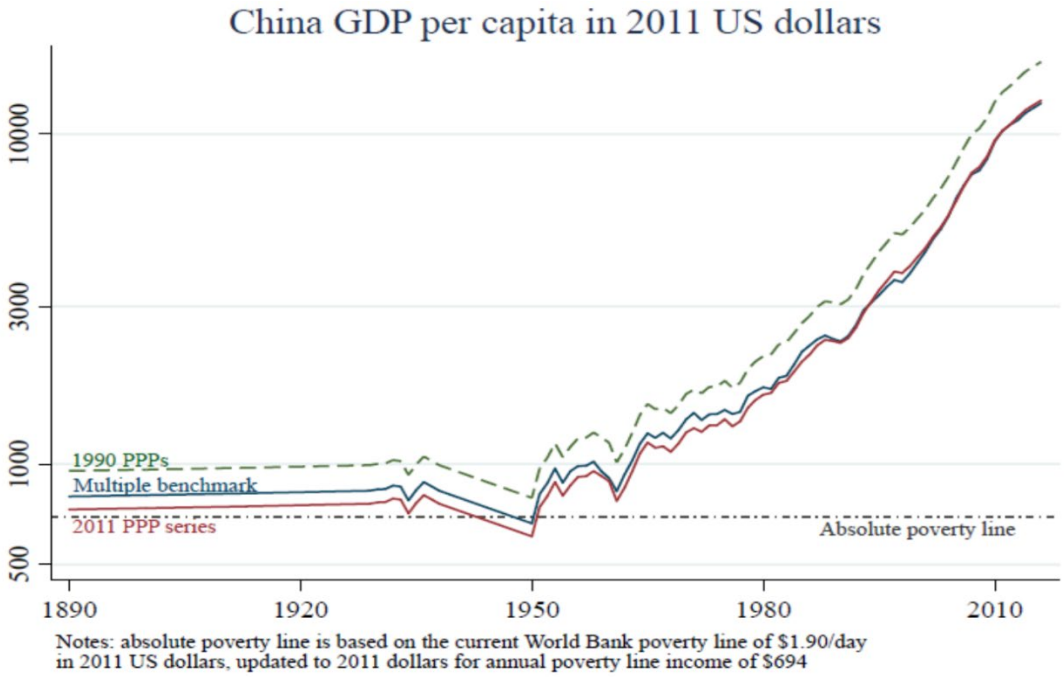
The effects of rebasing the original Maddison estimates has the most notable effects for countries who experienced substantial price changes relative to the US between the benchmarks years. China is an interesting case in this perspective. When the 2005 PPP's were released, the prices for China had increased so much relative to the US, that GDP pc levels ended up around 40% lower than China's relative income based on earlier price estimates (Deaton and Heston, 2010: 3; Feenstra et. al, 2013). This led to very implausible low historical income estimates for China, given that the original estimates were already very close to subsistence around 1950 (Maddison, 2007).

In the years after the release of the 2005 PPP's consensus arose about the 2005 shortcomings, most of which were corrected for in the 2011 ICP round. Still, relative prices for China relative to the US were substantially higher in 2011 compared to 1990 which lowers China's PPP adjusted income per capita in 2011 by 23%. Yet, in this paper we have updated the Maddison project incomes estimates for China based on Wu (2014) which show on average lower growth between 1952 and 2011 than the previous (official) estimates. Extrapolating backwards from the lower 2011 per capita income level using these lower growth rates leads to plausible historical income estimates for each alternative benchmark PPP (1990, 2011 and the multiple benchmark approach), see figure 5. The 2011 benchmark clearly gives a much lower level of average income due to the rapid increase in prices in China over the recent year, but income in China never falls substantially below subsistence, also not for earlier periods. The PWT

[19] This section is based on Bolt et al. (2018).

approach leads to similar results as the 2011 benchmark for the most recent years, but deviates away and closer to the 1990 benchmark results for earlier years.

Figure 5: Historical income series for China using alternative benchmark PPPs.



Looking more broadly into subsistence incomes, the original Maddison project dataset includes 184 observations below the subsistence income level of 400 dollars per year (on a total of 17,872 that is 1%), of which most are countries in times of civil war such as Afghanistan, Liberia, Burundi, and the Democratic Republic of Congo. Moving towards the 2011 PPPs, the poverty line becomes 694 dollars (using 2011 prices), and the number of below subsistence observations increase to 312 (1.8%). Using the Multiple Benchmark Method, the number of observations below subsistence increase even more to 386 (2.6%), and this includes various surprising countries: Peru during substantial parts of the 19th century, Egypt and Chile in 1820, or Korea during most of the period. Given the quality of the income series for some of those countries (the Peruvian historical series, for example, is one of the best available), concerns arise about the quality of their earlier benchmarks, which are explaining – in combination of course with the available time series of GDP per capita - these very low real income levels. The problem of the MBM is that the historical series are linked to the earliest available

benchmark of the ICP (of 1960 or 1970 for example), which may, in view of the further development of the project, have been rather crude and not the best available benchmark.

5.2 Testing the accuracy of backward projection: versus benchmarks

Even though the level effect on incomes of the new relative prices is for some regions substantial, this does not necessarily mean the view on past living standards completely changes as that view depends largely on relative income levels. In other words, how well off were certain countries compared to other countries during the same period. So another way to see how plausible the three scenarios of extrapolating income series backwards are, is to calculate relative income levels for countries for the years for which there are also independent benchmark estimates available. This comparison also confronts an often heard criticism of the work by Maddison as his dataset is based on the assumption that relative and absolute levels of real income could be measured in the relative prices of the base year 1990. Changes in relative prices were bound to lead to biases in the measured levels of GDP per capita; the longer the time period and the more dramatic relative prices changed over time, the more problematic the approach would be. This criticism has resulted in alterative attempts to reconstruct the (implicit) PPPs for historical periods – such as the econometric method introduced already – or to the use of alternative proxies for the wealth and poverty of nations, such as real wages.

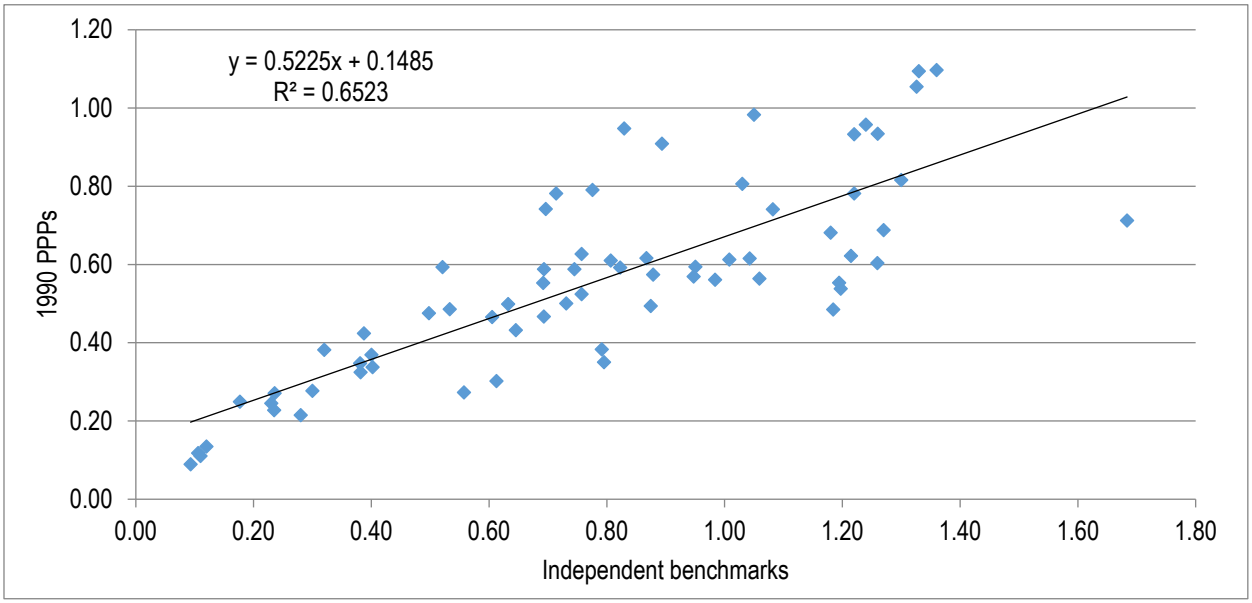
In this section we test the backward projections of the three scenarios (making use of the 2011, the 1990 and the multiple benchmark respectively), first of all to see which method results in the best 'predictions' of relative income levels in the 19th and early 20th century, but also to find out how large the biases are that result from the Maddison approach. In other words, do the critics have a point, or do these changes in relative prices cancel out in the medium long run?

We developed two ways of testing the accuracy of the backward projections. First and foremost we collected all available independent benchmarks available in the literature. For all of the independent benchmarks, we have calculated relative incomes for the same countries and years from the time series using the three different scenarios.

For this analysis, we collected all available independent historical comparisons of GDP per capita levels between sets of countries that are available in the literature.[20] For all of these independent benchmarks, we calculated relative GDP per capita for the same countries and years from the time series using the three different methods. Figure 6 shows that the R2 in panel A (1990 PPP benchmark) is 0.65, where the fit for both the 2011 PPP benchmarks (panel B) and multiple benchmarks (panel C) is much lower with an R2 of approximately 0.57. It appears that the 1990 benchmark produces a better for historical GDP estimates than the other two approaches.

Figure 6: comparison three scenarios with independent benchmarks

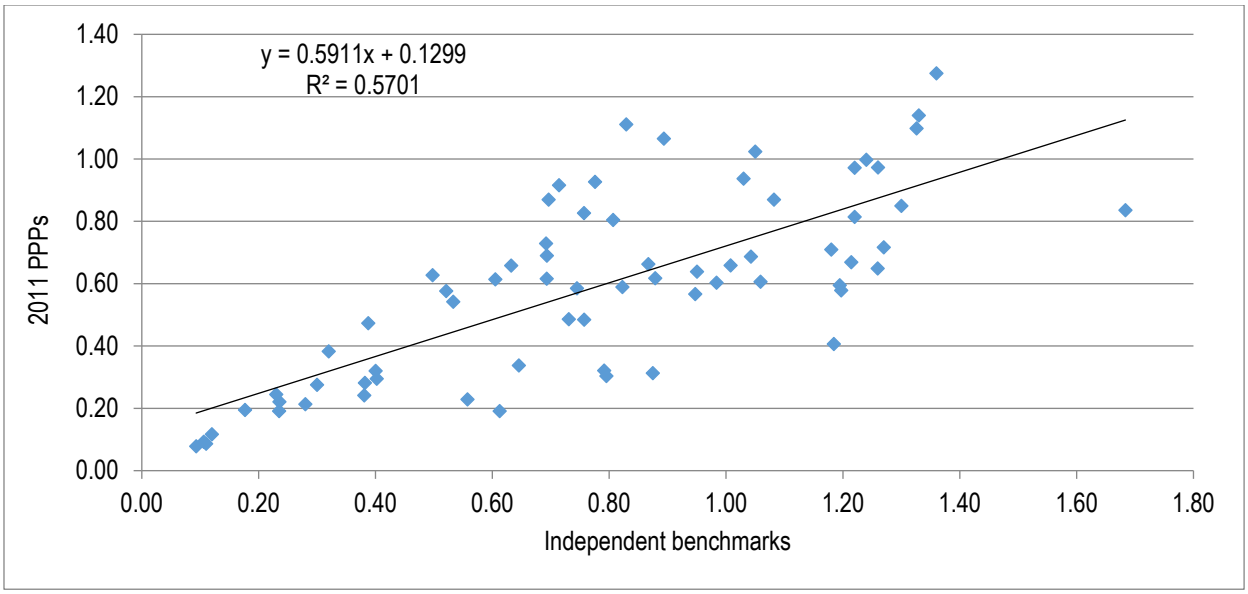
Panel A. 1990 PPPs versus independent benchmarks



Source: authors' calculation.

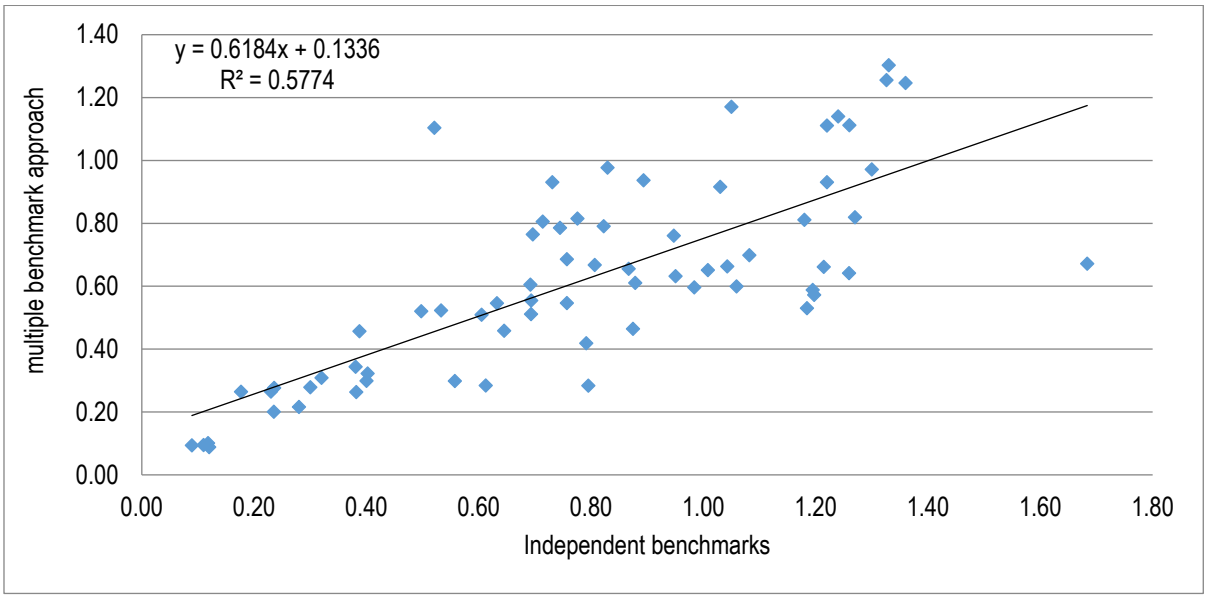
Panel B. 2011 PPPs versus independent benchmarks

[20] With one exception: we did not include the new Malaysia benchmarks as estimated by Bassino and Van der Eng (2020) as they diverge so much from earlier estimates that a more detailed analysis of these differences is required; we suspect that the high quality of the Nazrin (2016) estimates will help to explain the conundrum.



Source: authors' calculation.

Panel C. Multiple PPP benchmarks versus independent benchmarks



Source: authors' calculation.

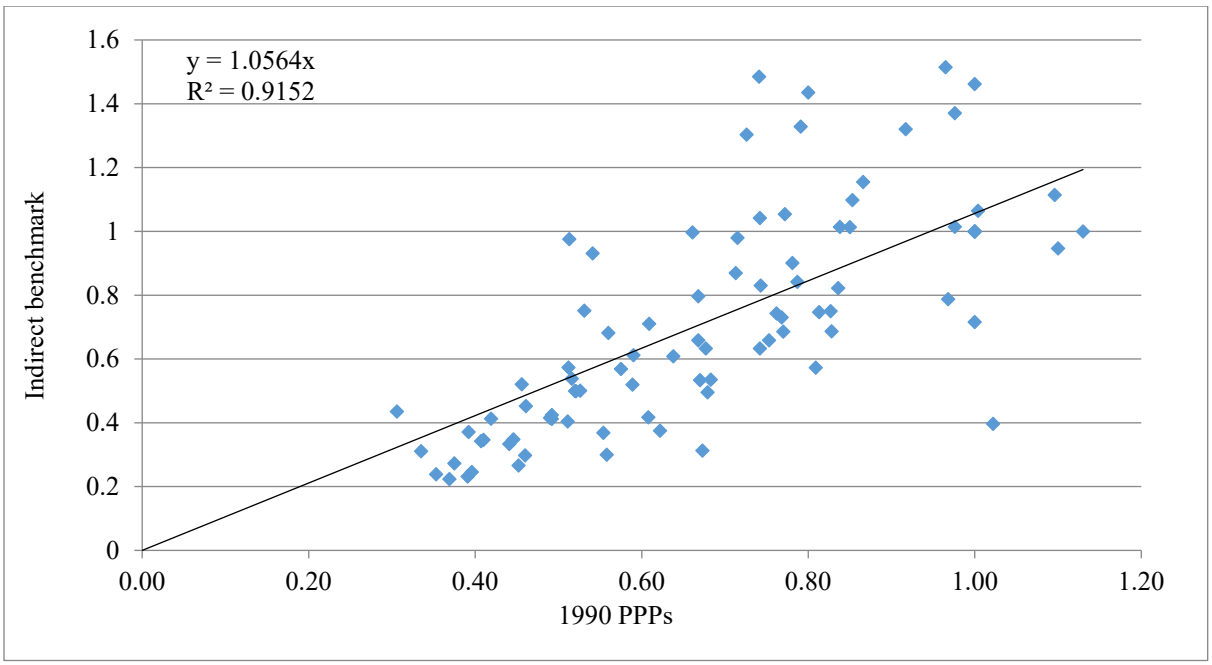
Additionally, we have also gathered *indirect* (econometrically derived) estimates of relative income levels in the 19th and early 20th century from Prados de la Escosura (2000).

Comparing the three scenarios of extrapolation with the indirect benchmarks from Prados de la Escosura (2000) shows systematic differences between the three scenarios on the one hand and the indirect benchmarks on the other hand, see figure 7 below. For all scenarios, the coefficient

of the trend line is close to 1, which suggests comparable results. But the scatterplots also shows a wide distribution of points around the trend line (R2 always smaller than 0.54) indicating that for many countries the relative income levels based on the time series are very different from the ones resulting from the indirect method. The results for the 1990 PPPs and the multiple benchmarks are again similar to each other and compare favourably to the results for the 2011 PPPs. The same exercise for the relative incomes derived from real wages produces results very alike the indirect benchmarks. The coefficient of the line is close to 1, but the variance around the line is even larger than for the indirect benchmarks. The R2 now becomes even lower, around 0.40 indicating an even wider distribution around the trend.

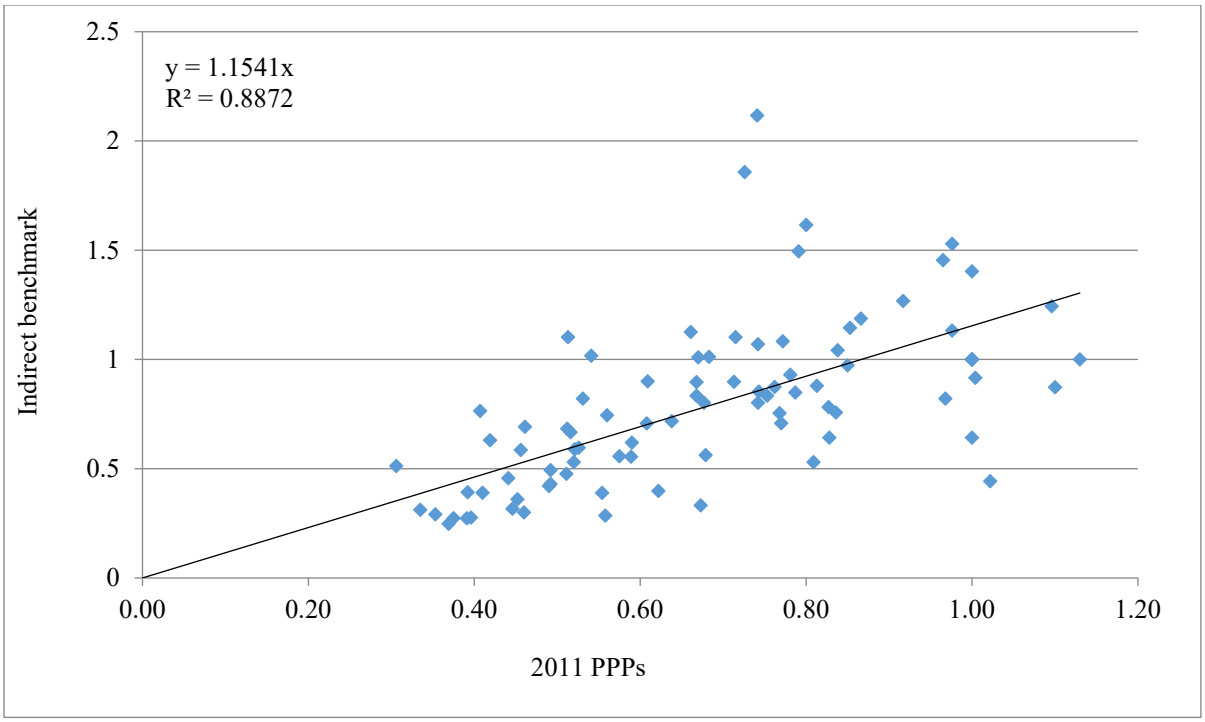
Figure 7: comparison three scenarios with indirect benchmarks

Panel A. 1990 PPPs versus indirect benchmarks



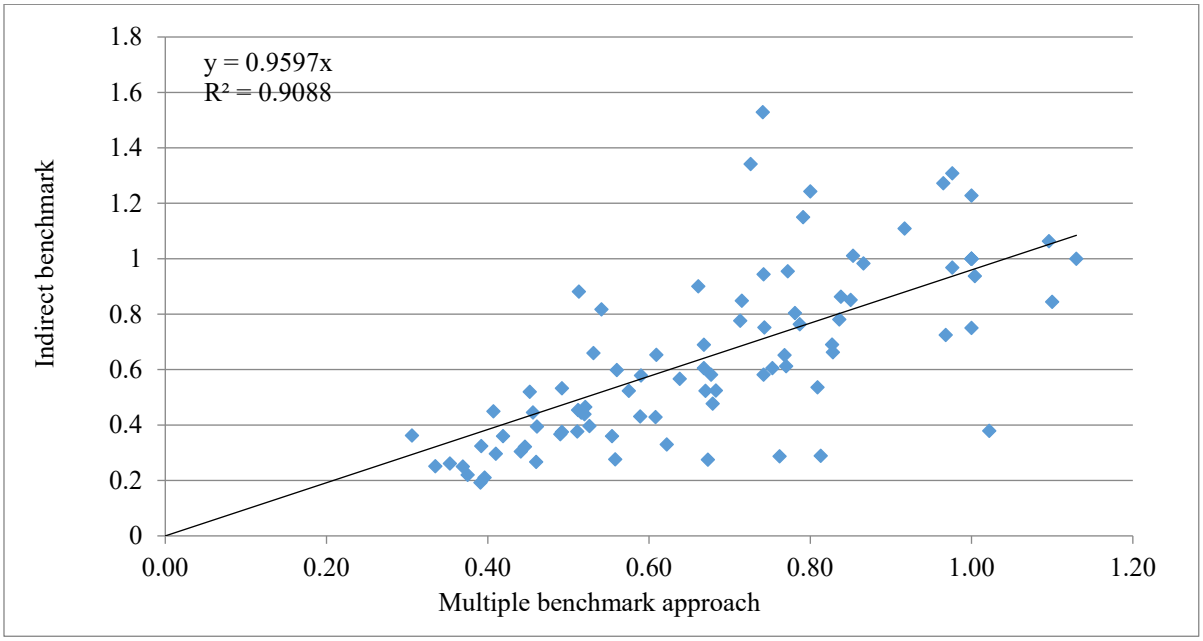
Source: authors' calculation.

Panel B. 2011 PPPs versus indirect benchmarks



Source: authors' calculation.

Panel C. Multiple benchmarks versus indirect benchmarks

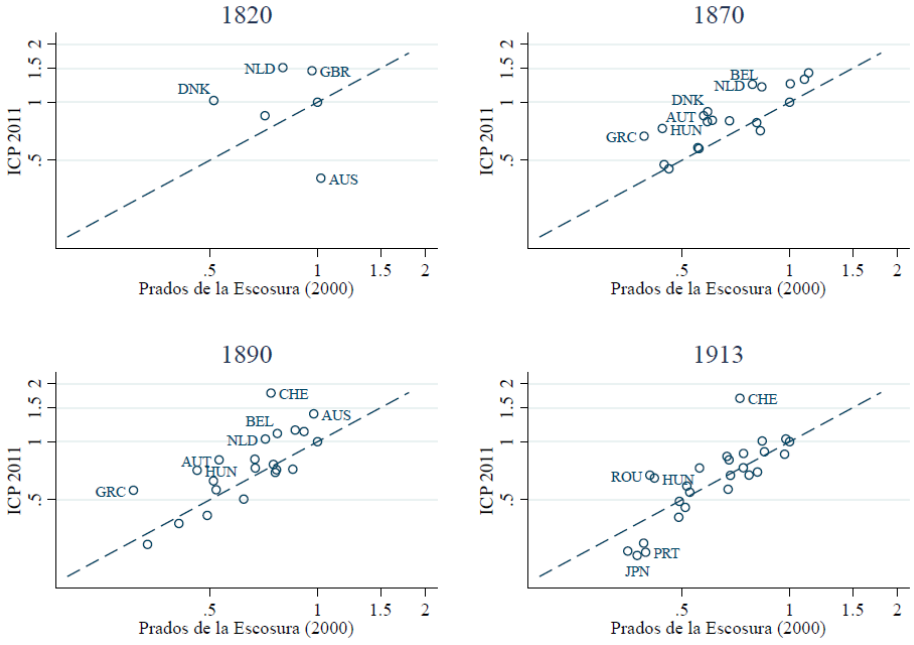


Source: authors' calculation.

As discussed extensively above, one disadvantage of the extrapolation method as applied in this paper is that by accepting one point of reference, the distortions and ambiguities in

intertemporal comparisons increase the further one moves away from the reference year. Prados de la Escosura (2000) provided an alternative to remedy this index number problem by constructing PPP converters based on an estimated relationship between price levels (PPP/Exchange rate ratios) and nominal product per head (exchange rate converted GDP pc) and a set of additional explanatory variables (Prados de la Escosura, 2000: 4). This method provides alternative real historical income levels. When we compare income levels of countries relative to the US as produced by Prados de la Escosura (2000) to the relative income levels as generated by the extrapolation method using the 2011 benchmark see that with the exception of 1820, both methods provide very similar relative income levels, see figure 8. For the years 1870 and 1890 the relative incomes produced by the extrapolation method tend to be a little higher on average (see figure below). For 1820, the differences are much more substantial.

Figure 8: Relative income levels from ICP 2011 versus indirect estimates from Prados de la Escosura (2000).



The conclusion of this long section is that simply moving from the 1990 to the 2011 benchmark is not improving estimates for the historical period, which is probably due to the greater distance in years between 2011 and the historical time series under review. However, the PWT approach, which uses all post 1950 benchmark estimates does not suffer from this bias, but does result in a rather high share of countries with below subsistence levels of GDP per capita, which is also considered problematic. An explanation of these results is that the quality of the benchmarks is probably subject to two changes: the quality of the PPPs and the coverage of the various ICP-rounds is increasing over time, whereas at the same time the distance to the 'historical' period is increasing as well, resulting in more biases due to changes in relative prices. Ergo, both a 'perfect' but 'distant' benchmark such as 2011, and the 'less than perfect' but 'early' benchmarks used by MBM, have sizeable biases. It follows that the 1990 benchmark created by Maddison may well be the best compromise solution. It certainly produces the best results in terms of the two tests that we carried out: it predicts historical benchmarks rather well (almost as good as the PWT approach) and it does not result in a high share of countries with a rather implausible low income level (below subsistence). This assessment therefore results in the conclusion that the best way forward is to stick to the 1990 benchmark for the overall architecture of the dataset.

At some points, however, this exercise leads to a number of departures from the original Maddison approach. There is, to begin with, no doubt that the 2011 PPPs and the related estimates of GDP per capita reflect the relative levels of GDP per capita in the world economy better than the combination of the 1990 benchmark and growth rates of GDP according to the national accounts. In the following way did we adapt the growth rates of GDP per capita in the period 1990-2011 to get a close fit between the two (1990 and 2011) benchmarks. We estimated (as shown in figure 2) the difference between the combination of the 1990 benchmark and the growth rates of GDP (per capita) between 1990 and 2011 according to the national accounts of the countries concerned (the Maddison-style estimates for 2011), and the results of the 2011 benchmark, taking the growth rate of the USA as the standard. This difference, as shown in Figure 2, is for example quite high for a few oil producing countries (which became rich without growing much), and negative for a few poor countries (Mozambique is the most extreme example), which did grow but apparently – according to the 2011 PPPs – did not improve their relative position. This difference is then evenly distributed to the growth rate of GDP per capita between 1990 and 2011; in other words, to the rate of growth of each year according to the national accounts is added a correction (which is constant for all years between 1990 and 2011)

Ito make it consistent with the two benchmarks 1990 and 2011. Growth after 2011 is in the current update exclusively based on the growth rates of GDP according to national accounts (following TED) (but in the future this might be changed when new PPPs become available). A concern here is that the two benchmarks 1990 and 2011 make use of different sets of international prices: Geary-Khamis prices for 1990 and GEKS prices for 2011. GEKS prices are now the standard and to be preferred over Geary-Khamis prices, but we cannot transform the 1990 PPPS into PPPS based on the new methodology as the underlying data are not available. Moreover, the concern that Geary-Khamis prices may result in an overestimate of GDP per capita of poor countries does not seem to be justified; one would expect a link between the level of GDP per capita in 1990 and the correction necessary to link it to the 2011 benchmark, but as Figure 2 shows, such a correlation does not exist (the estimated trend is flat). Moreover, we also developed a rather crude way to incorporate the available historical benchmarks.

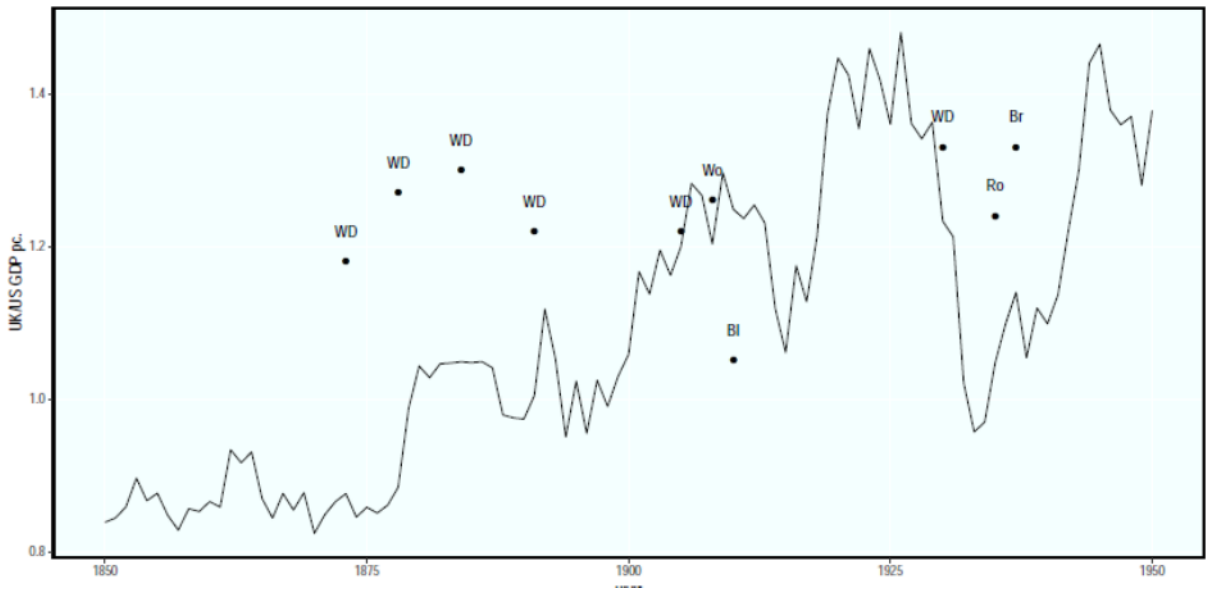
5.3 Incorporating important historical benchmarks

Having collected all available benchmarks, it makes sense to use the information contained in it. We have therefore used the benchmark estimates to fine tune the dataset for the pre-1940 period. When the quality of the benchmark was considered good and/or when multiple benchmarks pointed into the same direction, and the difference between the backward projected estimates based on the 1990 benchmarks was larger than 10 percent, we considered correcting the time series to fit the benchmark. It turns out that, when the Maddison 1990 benchmark is used, most historical benchmarks collected for this paper do not show a difference of more than 10 percent between the backward projections of the historical series, so the number of modifications is limited. Two exceptions are important, however.

The most important correction concerns the US/UK comparison. The conventional picture, based on the original 1990 Maddison estimates, indicated that the US overtook the UK as the world leader in the early years of the 20th century. This was first criticized by Ward and Devereux, 2003), who argued based on alternative measures of PPP-adjusted benchmarks between 1870 and 1930, that the United States led the United Kingdom in term of GDP per capita already in the 1870s, see figure 9. This conclusion was criticized by Broadberry, 2003). New evidence however also suggests that the overtaking by the United States indeed already

happened in the 19th century (Lindert and Williamson, 2015). This is also confirmed by direct benchmark comparison of the income of both countries for the period 1907-1909 (Woltjer, 2015). This shows that GDP per capita for the United States in those years was 26% higher than in the United Kingdom (Woltjer, 2015). We have used this benchmark to correct the GDP series of the two countries. Projecting this benchmark into the 19th century with the series of GDP per capita of both countries results in the two countries achieving parity in 1880, see figure 2.5. This is close to Prados de la Escosura's conjecture based on his short-cut method (Prados de la Escosura, 2000[53]) and even closer to the (Lindert and Williamson, 2015[56]) results

Figure 9. GDP per capita in the United States relative to the United Kingdom, 1850-1950



Note: Based on the new relative income levels included in the MPD 2020 update

Changing the US/UK ratio on the basis of the new research by (Woltjer, 2015) raises the question of which country's GDP estimates should be adapted. In the current PWT approach, the growth of GDP per capita in the United States is the anchor for the entire system. For the 19th century, however, it is more logical to take the United Kingdom as the anchor, because it was the productivity leader, and because most research focused on creating historical benchmarks takes the United Kingdom as reference point. We have therefore adapted the US series for the period 1908-1950 to fit the 1907-09 (Woltjer, 2015) benchmark. The reason for selecting this period is that there are doubts about the accuracy of price changes and deflators

for this period, which was characterized by two big waves of inflation (during the two World Wars) and by large swings in relative prices and exchange rates (as documented in the detailed analysis by (Stohr, 2016) for Switzerland). Future research will have to assess whether this choice is justified.

A similar problem arises with the pre 1800 series for Italy. It is now possible to link the Malanima series for Northern Italy between 1310 and 1871 with the series for the country as a whole after 1871 (thanks to the work by Felice 2018 on relative levels of GDP per capita in 1871). This 'lifts' the North Italian series however, as average GDP per capita in the north was higher than the national average, and results in estimates of GDP per capita for the period before 1800 which seem much too high. For example, according to these estimates, Northern Italy in 1500 would be more than twice as rich as Spain and Portugal, and 60% more wealthy than France and England. There is only one, rather tentative, benchmark for the pre 1500 period, related to the year 1427 for which detailed estimates of Tuscan GDP per capita can be made (Van Zanden and Felice, 2017). These authors also found a much smaller gap in GDP per capita between Tuscany (which is more or less representative of Northern Italy) on the one hand, and Holland and England on the other hand. The Holland and England estimates were consistent with each other and with their backward projections. We therefore considered correcting the Italian GDP series (by Malanima 2010) to fit the 1427 – 1871 trend. After consultation of a number of Italy experts, we decided not to make such changes (or, as Malanima suggested, to also adapt the 1871 level) (but see the discussion of the pre 1850-time series in Van Zanden and Bolt 2020).

5. **Conclusion**

In this paper we presented the new update of the Maddison project database. The Maddison project is an ongoing research project aimed at standardizing and updating the academic work in the field of historical national accounting in the tradition of the synthesis produced by Angus Maddison (1995: 2003). This paper complements the update presented in 2018 (Bolt et al., 2018), and is meant to explain the choices made in the 2018 update, and explore the implications of these choices for charting long run economic growth.

The Maddison database, like the original global database constructed by Maddison (1995; 2001) has always used the 1990 PPPs as a benchmark for its long-term income series. The primary aim of the 2018 revision was to incorporate the newly published and arguably better

new benchmark of 2011 PPPs. However, the switch to the new benchmark had substantial and often unexpected consequences for the picture of long run global development. We therefore felt it important to in this paper take a step back and explore in a more fundamental way the consequences of using alternative benchmarks, the 1990 benchmark, the 2011 benchmark and the multiple benchmark method following the recent PWT methodology, for especially historical developments paths.

This paper focusses on the pre-1950 period and develops a methodology to find out how large the biases are of the various ways in which the relative levels of GDP per capita in the more distant past can be estimated. For this purpose we collected all available direct and indirect estimates of relative income levels for groups of countries in the 19th and early 20th century from the literature. We then compare the relative income levels using the different PPPs, with the relative income levels as indicated by the independent (direct) and indirect benchmarks. We subsequently analyse whether a revision based on the 2011 PPPs or on the PWT approach would produce better estimates of relative levels of GDP per capita in the period before 1940 than the 'traditional' Maddison approach that uses his 1990 benchmark.

The conclusion of this analysis is that simply moving from the 1990 to the 2011 benchmark is not improving estimates for the historical period, which is probably due to the greater distance in years between 2011 and the historical time series under review. However, the PWT approach, which uses all post 1950 benchmark estimates does not suffer from this bias, but does result in a rather high share of countries with below subsistence levels of GDP per capita, which is also considered problematic. An explanation of these results is that the quality of the benchmarks is probably subject to two changes: the quality of the PPPs and the coverage of the various ICP-rounds is increasing over time, whereas at the same time the distance to the 'historical' period is increasing as well, resulting in more biases due to changes in relative prices.

This assessment therefore results in the conclusion that the best way forward is to stick to the 1990 benchmark for the overall architecture of the dataset. At the same time, our analysis also leads us to conclude that a number of departures from the original Maddison approach are preferable. There is, to begin with, no doubt that the 2011 PPPs and the related estimates of GDP per capita reflect the relative levels of GDP per capita in the world economy better than forward extrapolations based on national account growth rates from the 1990 benchmark. We therefore adapted the growth rates of GDP per capita in the period 1990-2011 to get a close fit between the two (1990 and 2011) benchmarks. Growth after 2011 is in the current update again exclusively based on the growth rates of GDP according to national accounts. We also

incorporated those historical benchmarks estimates into the dataset that resulted in major changes in levels of GDP per capita.

An important lesson on the use of PPPs for historical analysis, such as the Maddison Project, is that there is no single PPP concept, method or dataset that can be seen as optimal to satisfy all uses by economic historians or any researcher. Not only can PPPs differ for the reference year chosen, but they can also be dependent on whether they are obtained in a bilateral (comparing two countries) or multilateral (more than two countries) framework, whether the country price comparisons need to be weighted or unweighted for their share in the global economy, etc. The perspectives of the research from an expenditure or industry point-of-view can also make a difference for the choice of the PPP, as even at aggregate level it makes a difference whether one is looking for comparisons of living standards (requiring an expenditure view) or productivity capacity (Feenstra et al, 2013). In sum, the choice of the best set of PPPs for comparative analysis of economic performance over time ultimately depends on the question one wants to answer.

Appendix 1

Benchmark comparison	Year	Benchmark result	Maddison (1990)	Maddison (ICP 2011)	MBM	Source benchmark
Indonesia/Netherlands	1820	0.30	0.28	0.28	0.28	Van Zanden (2003)
	1860	0.28	0.21	0.21	0.22	Van Zanden (2003)
Netherlands/UK	1820	0.83	0.95	1.11	0.98	Frankema, Woltjer and Smits (2013)
	1850	0.89	0.91	1.07	0.94	Frankema, Woltjer and Smits (2013)
	1870	0.78	0.79	0.93	0.81	Frankema, Woltjer and Smits (2013)
	1890	0.70	0.74	0.87	0.76	Frankema, Woltjer and Smits (2013)
	1910	0.71	0.78	0.92	0.81	Frankema, Woltjer and Smits (2013)
Germany/UK	1860	0.50	0.48	0.63	0.52	Fremdling (1991)
	1870	0.61	0.47	0.61	0.51	Fremdling (1991)
	1880	0.69	0.47	0.62	0.51	Fremdling (1991)
	1890	0.63	0.50	0.66	0.55	Fremdling (1991)
	1900	0.69	0.55	0.73	0.60	Fremdling (1991)
	1910	0.81	0.61	0.80	0.67	Fremdling (1991)
	1935	0.76	0.63	0.83	0.69	Broadberry (2006)
India/UK	1950/51	0.09	0.09	0.08	0.09	Broadberry and Gupta (2010)
US/UK	1873	1.18	0.68	0.71	0.81	Ward and Devereux (2003)
	1878	1.27	0.69	0.72	0.82	Ward and Devereux (2003)
	1884	1.30	0.82	0.85	0.97	Ward and Devereux (2003)
	1891	1.22	0.78	0.81	0.93	Ward and Devereux (2003)
	1905	1.22	0.93	0.97	1.11	Ward and Devereux (2003)
	1908	1.26	0.93	0.97	1.11	Woltjer (2015)
	1910	1.05	0.98	1.02	1.17	Broadberry and Irwin (2006)
	1930	1.33	1.09	1.14	1.30	Ward and Devereux (2003)
	1935	1.24	0.96	1.00	1.14	Rostas (1948)
	1937	1.33	1.05	1.10	1.26	Broadberry (2006)

Australia/UK	1891	1.36	1.10	1.27	1.25	Haig (1989)
	1900	1.03	0.81	0.94	0.92	Haig (1989)
France/UK	1820	0.88	0.57	0.62	0.61	O'Brien and Keyder (1978)
	1830	0.98	0.56	0.60	0.60	O'Brien and Keyder (1978)
	1840	0.95	0.59	0.64	0.63	O'Brien and Keyder (1978)
	1850	0.87	0.62	0.66	0.66	O'Brien and Keyder (1978)
	1860	1.21	0.62	0.67	0.66	O'Brien and Keyder (1978)
	1870	1.20	0.54	0.58	0.57	O'Brien and Keyder (1978)
	1880	1.06	0.56	0.61	0.60	O'Brien and Keyder (1978)
	1890	1.19	0.55	0.59	0.59	O'Brien and Keyder (1978)
	1900	1.26	0.60	0.65	0.64	O'Brien and Keyder (1978)
	1910	1.01	0.61	0.66	0.65	O'Brien and Keyder (1978)
China/Netherlands	1825	0.38	0.35	0.24	0.34	Li and Van Zanden (2012)
China/UK	1840	0.24	0.27	0.22	0.28	Broadberry Guan and Li (2013)
India/US	1870	0.24	0.23	0.19	0.20	Heston and Summers (1980)
China/US	1912	0.11	0.12	0.09	0.10	Ma and de Jong (2019)
Japan/US	1935	0.32	0.38	0.38	0.31	Fukao, Ma and Yuan (2007)
Taiwan/US	1935	0.23	0.25	0.24	0.26	Fukao, Ma and Yuan (2007)
Korea/US	1935	0.12	0.13	0.12	0.09	Fukao, Ma and Yuan (2007)
China/US	1935	0.11	0.11	0.09	0.09	Fukao, Ma and Yuan (2007)
Burma/Japan	1913	0.87	0.49	0.31	0.46	Bassino and Van der Eng (2020)
	1938	0.61	0.30	0.19	0.28	Bassino and Van der Eng (2020)
Ceylon/Japan	1922	0.52	0.59	0.58	1.10	Bassino and Van der Eng (2020)
	1938	0.73	0.50	0.49	0.93	Bassino and Van der Eng (2020)
China/Japan	1913	0.65	0.43	0.34	0.46	Bassino and Van der Eng (2020)
	1938	0.18	0.25	0.19	0.26	Bassino and Van der Eng (2020)
India/Japan	1913	1.18	0.48	0.41	0.53	Bassino and Van der Eng (2020)
	1922	0.79	0.38	0.32	0.42	Bassino and Van der Eng (2020)

	1938	0.56	0.27	0.23	0.30	Bassino and Van der Eng (2020)
Indonesia/Japan	1913	1.04	0.62	0.69	0.66	Bassino and Van der Eng (2020)
	1922	0.53	0.49	0.54	0.52	Bassino and Van der Eng (2020)
	1938	0.39	0.42	0.47	0.46	Bassino and Van der Eng (2020)
Korea/Japan	1913	0.80	0.35	0.30	0.28	Bassino and Van der Eng (2020)
	1922	0.38	0.32	0.28	0.26	Bassino and Van der Eng (2020)
	1938	0.40	0.37	0.32	0.30	Bassino and Van der Eng (2020)
Philippines/Japan	1913	1.68	0.71	0.84	0.67	Bassino and Van der Eng (2020)
	1922	1.08	0.74	0.87	0.70	Bassino and Van der Eng (2020)
	1938	0.69	0.59	0.69	0.55	Bassino and Van der Eng (2020)
Taiwan/Japan	1913	0.95	0.57	0.57	0.76	Bassino and Van der Eng (2020)
	1922	0.82	0.59	0.59	0.79	Bassino and Van der Eng (2020)
	1938	0.74	0.59	0.59	0.79	Bassino and Van der Eng (2020)
Thailand/Japan	1938	0.40	0.34	0.29	0.32	Bassino and Van der Eng (2020)
Vietnam/Japan	1913	0.76	0.52	0.48	0.55	Bassino and Van der Eng (2020)

References

Alvarez-Nogal, C. and L. Prados de la Escosura (2013). The rise and fall of Spain (1270 – 1850). The Economic History Review, Vol. 66 (1): 1 - 37

Arroyo Abad, L. and J.L. van Zanden (2016), "Growth under Extractive Institutions? Latin American Per Capita GDP in Colonial Times" *Journal of Economic History* 76(4): 1182–1215.

Axenciuc, V. (2012). Produsul intern brut al Romaniei, Vol. 1, Institutl de Economie Nationala, Bucarest

Barro, R.J. and J.F. Ursua, (2008). "Macroeconomic Crises since 1870" *Brookings Papers on Economic Activity*, Economic Studies Program, The Brookings Institution, vol. 39(1 (Spring), pages 255-350

Bassino, J.-P. and van der Eng, P. (2020), Asia's 'little divergence' in the twentieth century: evidence from PPP-based direct estimates of GDP per capita, 1913–69. The Economic History Review, 73: 185-208. doi:10.1111/ehr.12880

Bassino, Jean-Pascal & Broadberry, Stephen & Fukao, Kyoji & Gupta, Bishnupriya & Takashima, Masanori, (2018). "Japan and the Great Divergence, 730-1874," CEI Working Paper Series 2018-13, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.

Bèrtola, L. (2016). El PIB per Capita de Uruguay 1870 - 2016: una reconstruccion. PHES working paper No 48

Bèrtola, L. and Ocampo, J. A. (2012). *The economic development of Latin America since independence* (Oxford: Oxford University Press).

Bolt, J. and J. L. van Zanden (2014). The Maddison Project: collaborative research on historical national accounts. The Economic History Review, 67 (3): 627–651; Maddison Project Database, version 2013.

Bolt, Jutta, Robert Inklaar, Herman de Jong and Jan Luiten van Zanden (2018), "Rebasing 'Maddison': new income comparisons and the shape of long-run economic development", Maddison Project Working paper 10; Maddison Project Database, version 2018

Broadberrry, S. (2003). “Relative per capita Income Levels in the United Kingdom and the United States since 1870: Reconciling Time Series Projections and Direct Benchmark Estimates”, *Journal of Economic History*, 63: 852-863

Broadberry, S. and B. Gupta, The historical roots of India’s service-led development: A sectoral analysis of Anglo-Indian productivity differences, 1870–2000, Explorations in Economic History, Volume 47, Issue 3, Pages 264-278,

Broadberry, S. and D. Irwin (2006). "Labor productivity in the United States and the United Kingdom during the nineteenth century," Explorations in Economic History, Elsevier, vol. 43(2), pages 257-279, April.

Broadberry, S. Guan and Li (2013a), “China, Europe and the Great Divergence: A Study in Historical National Accounting”, LSE. Heston and Summers (1980)

Broadberry, S.N. (2006), Market Services and the Productivity Race, 1850-2000: Britain in International Perspective, page 219, table 9.3. Cambridge: Cambridge University Press.

Broadberry, S.N., B. Campbell, A. Klein, M. Overton and B. van Leeuwen (2015), *British Economic Growth 1270-1870* Cambridge: Cambridge University Press.

Broadberry, S.N., Custodis, J. and Gupta, B. (2015), “India and the great divergence: an Anglo-Indian comparison of GDP per capita, 1600–1871” *Explorations in Economic History*, 55: 58-75.

Broadberry, S.N., Guan, Hanhui and David Daokui Li (2018), “China, Europe and the Great Divergence: a Study in Historical National Accounting, 980-1850”, Journal of Economic History, 78, 4 (2018) 955-1000.

Cha, M.S., Kim, N.N., Park, K.-J., Park, Y. (Eds.) (2020), Historical Statistics of Korea. Studies in Economic History, New York: Springer Publishing

Chen, S. and M. Ravallion (2001), “How Did the World’s Poor Fare in the 1990s?” *Review of Income and Wealth* 47(3): 283–300.

De Corso, G. (2013). El crecimiento economico de Venuzuela, Desde la Oligarquia Conservadora Hasta La Revolucion Bolivariana: 1830-2012. Uno Vision cuantitativa *: Venezuelan Economic Growth From The Conservative Oligarchy To The Bolivarian Revolution (1830-2012), Revista De Historia Económica / Journal of Iberian and Latin American Economic History, 31(3), 321-357. doi:10.1017/S0212610913000190

De Corso, G. and S. Kalmanovitz (2016). Una estimación del PIB de Panamá: 1906-1945, *Tiempo&Economía* 3 (1).

Deaton, A. (2010) "Price indexes, inequality, and the measurement of world poverty" *American Economic Review*100(1): 5–34.

Deaton, A. (2011), Reshaping the world: The 2005 round of the International Comparison Program, forthcoming, in Prasada Rao and Fred Vogel (eds.) Measuring the size of the world economy: the framework, methodology, and results from the International Comparison Program, Washington, DC., World Bank

Deaton, A. and Aten, B. (2017). 'Trying to Understand the PPPs in ICP 2011: Why Are the Results So Different?' *American Economic Journal: Macroeconomics* 9(1).

Deaton, A. and Heston, A., 'Understanding PPPs and PPP-based National Accounts' *American Economic Journal: Macroeconomics*, 2 (2010), pp. 1–35.

Díaz, J.B. Lüders, R. and Wagner, G. (2007) Economia Chilena 1810 - 2000, Producto total y sectorial una nueva mirada, Pontificia universidad Catolica de Chile, Insituto de Economia, *Documeno de Trabajo* no. 315

Eloranta, J., Voutilainen, M. and Nummela, I. (2016). "Estimating Finnish Economic Growth Before 1860" mimeo.

Feenstra, R.C., H. Ma, J.P. Neary and D.S.P. Rao (2013) "Who Shrunk China? Puzzles in the Measurement of Real GDP" *Economic Journal* 123(573): 1100–1129.

Feenstra, R.C., H. Ma, J.P. Neary and D.S.P. Rao (2013) "Who Shrunk China? Puzzles in the Measurement of Real GDP" *Economic Journal* 123(573): 1100–1129.

Feenstra, Robert C, Robert Inklaar and Marcel Timmer, (2015), 'The Next Generation of the Penn World Table' *American Economic Review* 105(10).

Felice, E., (2018) The roots of a dual equilibrium: GDP, productivity, and structural change in the Italian regions in the long run (1871–2011), *European Review of Economic History*, 19, 1-30. https://doi.org/10.1093/ereh/hey018

Fourie, J. and Van Zanden, J.L. (2013). GDP in the Dutch Cape Colony: the Nationals Accounts of a Slave-Based Society, South African Journal of Economics, vol. 81 (4): 467 - 490

Frankema, E., Woltjer, P. and Smits, J.P., 2013. Changing Economic Leadership. A new benchmark of sector productivity in the United States and Western Europe, ca. 1910. *TSEG/ Low Countries Journal of Social and Economic History*, 10(3), pp.80–114. DOI: http://doi.org/10.18352/tseg.219

Fremdling, Rainer. (1991). Productivity Comparison between Great Britain and Germany, 1855-1913. Scandinavian Economic History Review. 39. 28-42. 10.1080/03585522.1991.10408198.

Fukao, K., Bassino, J.-P., Makino, T., Paprzycki, R., Settsu, T., Takashima, M., and Tokui, J. (2015) Regional Inequality and Industrial Structure in Japan: 1874-2008, Tokyo: Maruzen Publishing.

Fukao, K., Ma, D. and Yuan, T. (2007), REAL GDP IN PRE-WAR EAST ASIA: A 1934–36 BENCHMARK PURCHASING POWER PARITY COMPARISON WITH THE U.S.. Review of Income and Wealth, 53: 503-537. doi:10.1111/j.1475-4991.2007.00243.x

Grytten, O.H. (2015). Norwegian gross domestic product by industry 1830 - 1930, Norges Bank Working paper 19/2015. Population from Maddison (2006).

Haig, B. (1989), Ineternational comparison of Australian GDP in the 19th century. Review of Income and Wealth, 35: 151-162. doi:10.1111/j.1475-4991.1989.tb00587.x

Herranz-Loncán, A. and Peres-Cajías (2016). "Bolivian GDP per capita since the mid-nineteenth century" *Cliometrica* 10: 99-128

Inklaar, R. and D.S.P. Rao (2017). 'Cross-Country Income Levels over Time: Did the Developing World Suddenly Become Much Richer?' *American Economic Journal: Macroeconomics* 9(1): 265–290.

Johnson Simon, Arvind Subramanian, Will Larson and Chris Papageorgiou (2009), "Is newer better? The Penn World Table growth estimates" VoxEU.org, 7 December.

Johnson Simon, Will Larson, Chris Papageorgiou and Arvind Subramanian(2013), "Is newer better? Penn World Table Revisions and their impact on growth estimates", Journal of Monetary Economics 60(2), 255-274.

Klasing, M. J. and Milionis, P. (2014). "Quantifying the evolution of world trade, 1870–1949," *Journal of International Economics* 92(1): 185-197.

Krantz, O. (2017) “Swedish GDP 1300-1560 A Tentative Estimate” Lund Papers in Economic History: General Issues; No. 152.

Kravis, I.B., and R.E. Lipsey, “The International Comparison Program: Current Status and Problems”, in P.E. Hooper and J.D. Richardson, International Economic Transactions: Issuesin Measurement and Empirical Research, (Chicago, University of Chicago Press, 1991.)

Kravis, I.B., R. Summers and A. Heston (1982), “Comments on: ‘Selected Problems of the Intercountry Comparisons on the Basis of the Experience of the EEC.’” *Review of Income and Wealth* 28(4): 407-10.

Li, B. and J.L. Van Zanden. “Before the Great Divergence? Comparing the Yangzi Delta and the Netherlands at the Beginning of the Nineteenth Century.” The Journal of Economic History, vol. 72, no. 4, 2012, pp. 956–989., www.jstor.org/stable/41811360. Accessed 21 Apr. 2020.

Lindert, P. H. and Williamson, J. G. (2015), ‘American colonial incomes, 1650–1774’. The Economic History Review. doi: 10.1111/ehr.12106

Ma, Y. and de Jong, H. (2019), Unfolding the Turbulent Century: A Reconstruction of China's Historical National Accounts, 1840–19121. Review of Income and Wealth, 65: 75-98. doi:10.1111/roiw.12314

Maddison , A., Contours of the World Economy, 1-2030 AD, (Oxford University Press, 2007).

Maddison, A. Dynamic Forces in Capitalist Development. A Long-Run Comparative View, (Oxford: Oxford University Press, 1991)

Maddison, A. Phases of Capitalist Development, (Oxford, Oxford University Press, 1982).

Maddison, A. The World Economy: Historical Statistics, (Paris: Organisation for Economic Cooperation and Development, 2006).

Maddison, A., Monitoring the World Economy 1820–1992, (Paris: Organisation for Economic Cooperation and Development, 1995).

Maddison, A., The World Economy: a Millennial Perspective, (Paris: Organisation for Economic Cooperation and Development, 2001).

Malanima, P. (2011), "The long decline of a leading economy: GDP in central and northern Italy, 1300–1913" *European Review of Economic History* 15: 169–219.

Malinowski, M. and Van Zanden (2017), "National income and its distribution in preindustrial Poland in a global perspective" *Cliometrica, Volume 11, Issue 3, pp 375–40*

Nazrin, S. (2016) *Economic history of Malaya* (Kuala Lumpur, Oxford UP), data online at https://www.ehm.my/data/, accessed at March 2020.

O'Brien, P. and C. Keyder (1978). Economic Growth in Britain and France 1780–1914: Two Paths to the Twentieth Century, Boston: George Allen and Unwin

Palma, N., & Reis, J. (2019). From Convergence to Divergence: Portuguese Economic Growth, 1527–1850. *The Journal of Economic History, 79*(2), 477-506. doi:10.1017/S0022050719000056

Pamuk, S. (2006), Estimating Economic Growth in the Middle East since 1820, The Journal of Economic History, vol 66, no. 3, pp. 809 - 828

Pamuk, S. (2009). "Estimating GDP per capita for the Ottoman Empire in a European Comparative Framework, 1500-1820", paper presented at the XVth World Economic History Congress, August 2009, Utrecht.

Prados de la Escosura, L. (2017), Spanish Economic Growth, 1850-2015 (London: Palgrave Macmillan)

Prados de la Escosura, L., 'International Comparisons of Real Product, 1820–1990', Explorations in Economic History, 37 (2000), pp. 1–41.

Prados de la Escosura, L., 'Lost decades? Economic performance in post-independence Latin America', Journal of Latin America Studies, 41 (2009), pp. 279–307.

PWT (Penn World Tables) (2011), Description of PWT 7.0 (http://pwt.econ.upenn.edu/Downloads/pwt70/PWT%207%20Web2011a.doc)

Ravallion, Datt and van de Walle (1991), "Quantifying absolute poverty in the developing world" *Review of Income and Wealth* 37(4).

Ren, R. China's Economic Performance in International Perspective, (Paris: Organisation for Economic Cooperation and Development, 1997).

Ridolfi, L. (2016) “The French economy in the longue durée. A study on real wages, working days and economic performance from Louis IX to the Revolution (1250-1789)” Dissertation IMT School for Advanced Studies, Lucca, available at http://e-theses.imtlucca.it/211/1/Ridolfi_phdthesis.pdf

Rostas, L. (1948). Comparative Productivity in British and American Industry,Cambridge, Cambridge University Press.

Santamaria Garcia, A. (2005). Las Cuentas nacionales de Cuba, 1960 - 2005', mimeo

Schön, L., and O. Krantz (2015) “New Swedish Historical National Accounts since the 16th Century in Constant and Current Prices” Lund Papers in Economic History no. 140.

Seminario, B. (2015). El Desarrallo de la Economía Peruana en la Era Moderna, Universidad de Pacifico, Lima

Stohr, Christian (2016), Trading Gains: new estimates of Swiss GDP, 1851-2008. Economic History Working Papers, 245/2016. London School of Economics and Political Science, Economic History Department, London, UK

Sugimoto, I. (2011), *Economic growth of Singapore in the twentieth century: historical GDP estimates and empirical investigations*, Economic Growth Centre Research Monograph ser., 2, http://www.worldscibooks.com/economics/7858.html (accessed on 30 Jan. 2013).

Summers, R. and Heston, A. (1991), ‘The Penn World Table(Mark 5): An Expanded Set of International Comparisons, 1950–1988’, The Quarterly Journal of Economics, 106, pp. 327–368.

Van Zanden 'Rich and Poor before the Industrial Revolution: a comparison between Java and the Netherlands at the beginning of the 19th century', Explorations in Economic History 40 (2003), 1-23.

van Zanden, J. L. and van Leeuwen, B. (2012), ‘Persistent but not consistent: the growth of national income in Holland 1347–1807’, Explorations in Economic History, 49 (2012), pp. 119–30.

van Zanden, J.L. and Emanuele Felice, (2017). ‘Benchmarking the Middle Ages. XV century Tuscany in European Perspective, Working Papers 0081, Utrecht University, Centre for Global Economic History.

Van Zanden, J.L. and J. Bolt, (2020). Two concerns about the interpretation of the estimates of historical national accounts before 1850. Submitted Journal of Global History.

Ward M. and J. Devereux (2003). Measuring British Decline: Direct Versus Long-Span Income Measures. *The Journal of Economic History, 63*(3), 826-851. doi:10.1017/S0022050703542000

Ward, M. and Devereux, J. (2003). Measuring British Decline: Direct Versus Long-Span Income Measures. The Journal of Economic History, 63: 826-851. doi:10.1017/S0022050703542000.

Ward, M. and Devereux, J. (2012), "The Road Not Taken: Pre-Revolutionary Cuban Living Standards in Comparative Perspective" *Journal of Economic History*, 72(1): 104–132.

Woltjer, P.J., (2015), Taking over: a new appraisal of the Anglo-American Productivity gap and the nature of American economic leadership ca. 1910, Scandinavian Economic History Review 63:3, 280-301, DOI: 10.1080/03585522.2015.1034766

World Bank (2008), Global Purchasing Power Parities and Real Expenditures: 2005 International Comparison Program, Washington, DC.

World Bank (2014). *Purchasing Power Parities and Real Expenditures of World Economies Summary of Results and Findings of the 2011 International Comparison Program* World Bank: Washington DC.

Wu, Harry X. (2014), "China's growth and productivity performance debate revisited – Accounting for China's sources of growth with a new data set" *The Conference Board Economics Program Working Paper Series* EWP#14-01.

Xu, Y. Z. Shi, B. van Leeuwen, Y Ni, Z Zhang, and Y Ma, (2016) 'Chinese National Income, ca. 1661-1933', Australian Economic History Review 57(3), 368–393.