Rebasing 'Maddison': new income comparisons and the shape of long-run economic development [1]

Preliminary version – do not cite without contacting the authors

Jutta Bolt

Robert Inklaar

Herman de Jong

Jan Luiten van Zanden

May 2017

Abstract

Economists' understanding of long-run economic development has relied on the historical statistics compiled by the late Angus Maddison. His method for comparing income levels across countries and over time has come under increasing criticism, though, as work on historical price and income comparisons often leads to markedly different outcomes than Maddison's projection (or extrapolation) method based on a single, modern-day relative income benchmark. In this paper, we extend the methods recently introduced for the Penn World Table to exploit data for multiple income benchmarks. By also incorporating historical (pre-1950) income benchmarks and developing new (indirect) income benchmarks, the newly developed Maddison Project Database does greater justice to historical research and provides a fresh impetus for further research efforts.

[1] Acknowledgements: Bart van Ark, John Devereux, Alan Heston, Branko Milanovic, Leandro Prados de la Escosura.

1. Introduction

The late Angus Maddison has greatly contributed to economists' understanding of long-run economic development through his Historical Statistics of the World Economy.[2] By judiciously combining estimates of comparative levels of real GDP per capita in recent periods with long-term time series of growth of GDP per capita, his database provides the broadest coverage of comparative income data and is amongst the most widely used sources of economic data in the world. Especially for the period before 1950, his is the only database that provides systematic and broad cross-country information on comparative income levels.[3] Since his passing, the development of the Maddison Project Database (MPD) has moved a new generation of scholars.[4] In this paper we describe the main changes to the measurement of real GDP per capita in the new version of this database.

Most importantly, we 'rebase' the MPD by incorporating a wealth of historical data on comparative living standards and economic activity, much of which has been developed in reaction to Maddison's pioneering work. The original Maddison series were based on a single modern-day cross-country comparison of relative income levels, in 1990, projected forwards and backwards using data on growth of GDP per capita. Yet extended back over many decades and even centuries, these projections diverged substantially from benchmark comparisons of relative income or living standards for early periods.[5] This is consistent with a recent literature on how differences in real GDP per capita between benchmarks comparisons can diverge from GDP growth from national statistics over the same period.[6] Changing economic structures and measurement error and biases in cross-country price comparisons are important explanations for such differences, but over longer time scales, growth figures also turn unreliable, especially when covering periods of war, rapid inflation or weak-to-non-existent statistical systems. A consequence is that research results can be sensitive to the version of a database that is used in a study.[7] This has been one reason why version 8 and 9 of the Penn World Table (PWT) introduced real GDP series that rely on multiple benchmark comparisons of prices and income; see Feenstra, Inklaar and Timmer (2015).

---

[2] See Maddison (1995, 2001, 2003).
[3] Though Barro and Ursúa (2008) have gone to great lengths to better capture data on economic fluctuations for 42 countries since 1800.
[4] See Bolt and van Zanden (2014) for a first new version.
[5] Prominent examples are Prados de la Escosura 2000 and Lindert and Williamson (2016).
[6] See Deaton (2010), Deaton and Aten (2017) and Inklaar and Rao (2017).
[7] See Johnson, Larson, Papageorgiou and Subramanian (2013) and Ciccone and Jarociński (2010).

In this paper, we implement a multiple benchmark approach for the MPD based, primarily, on (i) post-1950 price benchmarks (as also used in PWT) (ii) pre-1950 real GDP per capita benchmarks based on a variety of historical studies,[8] and (iii) a novel method for incorporating real wage comparisons for estimating comparative income levels.[9] In our new dataset on historical benchmarks we incorporate relative income levels for 29 out of the 78 countries for which there are income estimates available prior to 1950. By integrating comparative income levels for earlier periods, the measurement of long-term relative income developments is more closely related to research covering this historical period. In addition, we incorporate recent estimates of historical national accounts for a range of countries to provide a new version of the MPD that is state-of-the-art and provides a more extensive picture of comparative income levels than had been available thus far, with coverage for over 160 countries and the period from Roman times to the present.

The rest of this paper is organized as follows. In Section 2, we provide a guided tour of the data in the MPD, highlighting the main variables, briefly discussing their construction and indicating areas of research where they can be helpful. As in newer versions of PWT, the MPD distinguishes between a series of real GDP that is useful for comparing income levels across countries and a series that is useful for comparing growth performance over time. Section 3 discusses in greater detail the methodology for comparing income levels, at a point in time, but especially over a (long) period of time. Section 4 discusses the implementation of the multiple-benchmark approach including a discussion of the different types of information that are developed or are available in the different periods. This also includes a discussion of how our chosen approach compares to other methods, such as indirect benchmark estimates.[10] Section 5 discusses a number of applications, highlighting where the new database sheds new light on existing questions, such as the gap between GDP per capita in the United Kingdom and the United States and, more broadly, on the shape of economic development over the past centuries.

2. User guide to the data

The main aim of the MPD is to provide data on GDP per capita for comparisons of relative income levels across countries. This is often called 'real GDP per capita' in the international comparisons literature, where 'real' refers to the series being based on a common set of prices

[8] With Ward and Devereux (2016) as a major contributor.
[9] Such as estimated by Allen (2001) or Lindert and Williamson (2016).
[10] E.g. Prados de la Escosura (2000).

across countries. In the original work by Maddison (1995, 2001, 2003), such data was compiled by starting from a modern-day cross-country income comparison – for the year 1990 – and then using growth rates of GDP per capita from (historical) National Accounts to make comparisons for earlier years. An attractive feature of those data was that the change in real GDP per capita over time matches the growth rate from those National Accounts. However, this internal consistency came at the expense of distorted real GDP per capita comparisons in earlier years; see Section 3 on how, for instance, changing consumption patterns can lead to such distortions. Limitations to data quality also means that estimating the growth of GDP per capita over many decades, or even centuries, is a hazardous undertaking that, despite the best effort of statisticians and researchers, will always be surrounded by a degree of uncertainty. As a result, earlier estimates of relative income levels diverge substantially from standalone benchmark comparisons of relative income for those early periods.

In the new version of the MPD, we therefore introduce a new measure of real GDP per capita based on multiple benchmark comparisons of prices and incomes across countries. The resulting measure of real GDP per capita can best be understood as based on prices that are constant across countries but depend on the current year. In keeping with the terminology used in the Penn World Table (Feenstra et al. 2015), we refer to this measure of real GDP per capita as $CGDPpc$. This variable is expressed in 2011 US dollars by correcting for inflation in the United States to provide magnitudes that are comparable over time, but it is a 'current' measure in the sense that the (implicit) relative prices used for the cross-country comparisons differ over time. As a result, the relative income levels from this exercise more closely reflect direct historical income comparisons. We rely on a number of different types of price or income benchmarks in the construction of the MPD, which will be discussed in more detail in Section 3. We provide labels for all income observations indicating the method used to obtain it.

In addition to the $CGDPpc$ series, we provide a measure of growth of GDP per capita that relies on a single cross-country price comparison, for 2011. This series is also expressed in 2011 US dollars (and $CGDPpc = RGDPNApc$ in 2011), but its defining feature is that it tracks the growth rate of GDP per capita as given in country National Accounts (or their historical reconstructions). Following PWT, we refer to this measure of real GDP per capita as $RGDPNApc$. This series is primarily useful for comparing growth rates of GDP per capita over time. To also allow for a comparison of total GDP, the MPD provides information on population, with variable POP. For the historical (pre-1950) period, data is sometimes available for only population or only for GDP per capita, due to differences in basic data availability.

In compiling this dataset, we set a number of priorities, in line with the earlier work of Maddison. First, the primary goal is to provide measures of GDP per capita, i.e. productive capacity of economies. GDP per capita is a measure that easily diverge from more specific measures of comparative living standards of consumers or laborers,[11] or more comprehensive measures of welfare, that account for differences in health, leisure and inequality.[12] GDP per capita is typically highly correlated with such measures of wellbeing, but important differences can be seen. For example, in oil-rich countries in the Middle East (e.g. Qatar or United Arab Emirates), GDP per capita is considerably higher than household consumption per capita. An important benefit of GDP per capita is that it can be used not only as an (imperfect) indicator of wellbeing or living standards, but can also serve as the basis for productivity comparisons, which have the potential to shed more light on the (proximate) sources of cross-country income differences, such as differences in physical and human capital and productivity.[13]

Another important choice is to maximize the coverage of countries and periods, to provide a broad view on economic development in history. This, again, mirrors the approach of Maddison, but comes at the cost of a sparser set of concepts covered. For example, PWT provides an expenditure-level breakdown of GDP, as well as measures of physical and human capital and productivity for the period since 1950 (Feenstra et al. 2015). In a more historical context, Barro and Ursúa (2008) provide data on consumption per capita, in addition to GDP per capita for a smaller set of countries. While cognizant of this trade-off, we hope that by providing the broadest possible canvas, the MPD can serve as basis for future research to extend it in other directions.

3. Measurement of real GDP per capita

3.1 Measurement at a point in time

In any model of the economy that features non-traded as well as traded products, we can only measure real GDP per capita by measuring and comparing price levels across countries. One could compare real expenditure on *traded* products, using exchange rates to express nominal expenditure in real terms, but only if one is willing to assume that the law-of-one-price (LOP) holds. However, that is a strong assumption, already in modern times (e.g. Burstein and Gopinath, 2014), but even more so in historical periods when barriers to trade and limited

[11] As in e.g. Lindert and Williamson (2016).
[12] See e.g. Jones and Klenow (2016) or Gallardo Albarran (2017).
[13] See e.g. Caselli (2005) or Hsieh and Klenow (2010).

market integration held sway (e.g. Irwin, 2005; O'Rourke, 2007). For non-traded products, there is no mechanism that would push prices towards the LOP and it is amongst the stronger empirical regularities in international economics that prices of non-traded products are systematically lower in low-income economies. This is usually explained using the Balassa-Samuelson hypothesis (Samuelson, 1994), whereby productivity differences between countries are larger in traded goods than in non-traded good. As a country develops and its productivity in the traded sector increases, wages increase across the economy, leading to higher prices of non-traded products. As a result, differences in income levels would be substantially overstated if the comparison would be based on exchange-rate converted expenditure.

So rather than relying on exchange rates, the objective should be to estimate real GDP per capita based on a comparison of prices of traded and non-traded products. Deaton and Heston (2010) provide an extensive overview of the conceptual (as well as practical) challenges in making such comparisons. From a conceptual perspective it might be a desirable goal to compare the cost of living, so that a real expenditure comparison can be interpreted as a comparison of utility across countries. However, in a world of non-homothetic and (quite possibly) non-identical preferences, a true cost-of-living comparison faces substantial conceptual and practical challenges, though see Neary (2004) for an approach of comparing cost-of-living assuming identical but non-homothetic preferences.

A more achievable goal is to compare a weighted average of relative prices across countries, drawing on index number theory. Let $\mathbf{p}_j$ be the vector of prices in country j and let $\mathbf{q}_j$ be the vector of products. Nominal GDP in country j is then $P_j Y_j = \mathbf{p}_j' \mathbf{q}_j$, the sum of spending on (domestic) products.[14] Given these vectors for two countries, we can implement the thought experiment 'what would a person in country k have to spend to purchase the same bundle of products as a person in country j' to arrive at the Laspeyres price index. The Paasche price index is the outcome of the reverse though experiment, switching the bundle of products to that of country k:

$$P_{jk}^L = \frac{\mathbf{p}_k' \mathbf{q}_j}{\mathbf{p}_j' \mathbf{q}_j}, P_{jk}^P = \frac{\mathbf{p}_k' \mathbf{q}_k}{\mathbf{p}_j' \mathbf{q}_k} \quad (1)$$

[14] This implies that imported products enter in $\mathbf{q}_j$ with a negative sign.

Neither of the these thought experiments is inherently preferable as there is no reason why either bundle of products should hold a privileged position. Let, therefore, be the Fisher price index be:

$$P_{jk}^{F} = \left[\frac{\mathbf{p}_k'\mathbf{q}_j}{\mathbf{p}_j'\mathbf{q}_j} \times \frac{\mathbf{p}_k'\mathbf{q}_k}{\mathbf{p}_j'\mathbf{q}_k}\right]^{\frac{1}{2}} \quad (2)$$

The Fisher index has numerous desirable properties, amongst which is that if two countries are compared where the consumer's utility function has a homothetic, quadratic functional form, this index will exactly measure the ratio of utilities u_k/u_j (Diewert, 1976).

In a setting of many countries, a drawback of the Fisher index is that price comparisons are not transitive, i.e. the results depend on the base country, j here. As a result, comparing prices between j and k directly will yield a different outcome than via a third country h: $P_{jk}^{F} \neq P_{jh}^{F} \times P_{hk}^{F}$. To overcome this lack of transitivity we compare prices between j and k as the average across all possible indirect comparisons with country $h = 1, \ldots, \mathrm{C}$ to arrive at the so-called GEKS price index:[15]

$$P_{jk}^{GEKS} = \prod_{h=1}^{C} \left(P_{jh}^{F} P_{hk}^{F}\right)^{\frac{1}{C}} \quad (3)$$

The GEKS index is the most widely-used approach for comparing prices across countries, with it being the main method in the International Comparison Program (ICP) at the World Bank (2014) for computing global relative prices, or purchasing power parities (PPPs). An especially desirable property of the GEKS index is that it does not suffer from substitution bias, i.e. the GEKS index is based on the bundles of products $\mathbf{q}_j$ of all countries rather than relying on some average bundle. In the data provided by Maddison, he relied Geary-Khamis (GK) PPPs for his international comparisons and this index does suffer from substitutions bias. As illustrated by, for instance, Deaton and Heston (2010), this substitution bias causes the GK PPPs to understate prices in low-income countries, thereby overstating their real GDP per capita levels and thus understating the extent of cross-country income differences.

Given a relative price index as defined in equation (3), we can estimate real GDP as:

[15] After Gini, Eltetö, Köves, and Szulc. A modern treatment and references are provided by Balk (2008).

$$Y_k = \frac{P_k Y_k}{P_{jk}^{GEKS}} \quad (4)$$

which allows for comparing GDP or GDP per capita between countries j and k, evaluated at common prices.

3.2 Measuring real GDP per capita over time

The exposition so far has focused on price and comparisons across countries in a given year. Yet the main goals of the MPD (and other databases, such as PWT) is to provide data over time. The simplest approach is the so-called projection or extrapolation approach. In this approach real GDP per capita $y_{jt} \equiv Y_{jt}/N_{jt}$ (with N_{jt} as total population in country j at time t) is estimated as:

$$y_{jt-1} = \frac{y_{jt}}{1 + g_{jt}} (5)$$

where g_{jt} is the growth of GDP per capita in constant national prices. An important consequence of the approach in equation (5) is that the time series of growth in GDP per capita is the same in national prices and in PPP-converted US dollars. Furthermore, the change in the PPP implied by equation (5) is:

$$P_{jkt-1} = P_{jkt}^{GEKS} / \left[\frac{1 + \pi_{jt}}{1 + \pi_{kt}}\right], (6)$$

where $\pi_{jt} = P_{jt}/P_{jt-1} - 1$, the rate of inflation of the GDP deflator.

While straightforward, this extrapolation approach has important conceptual and practical drawbacks. The conceptual argument can be seen by considering the time-series counterpart to equation 2, so where the change in the GDP deflator (in country j) is computed between two time periods:

$$P_{jt,t-1}^{F} = \left[\frac{\mathbf{p}'_{jt}\mathbf{q}_{jt-1}}{\mathbf{p}'_{jt-1}\mathbf{q}_{jt-1}} \times \frac{\mathbf{p}'_{t}\mathbf{q}_{t}}{\mathbf{p}'_{t-1}\mathbf{q}_{t}}\right]^{\frac{1}{2}} \quad (7)$$

Equation (7) makes clear that a price index for national inflation should be computed using the bundle of products in the two periods for country j. Yet as equation (2) makes clear, a good measure of relative prices should take into account the bundle of products in country j *and* in country k. By ignoring country k's bundle in the computation of inflation in country j (and vice versa), the implicit relative price index in $t-1$ is no longer a good measure of relative prices

between countries j and k. Especially if the periods under comparison are far apart, the extrapolation approach of equations (5) and (6) is likely to a poor approximation as the bundle of products will have shifted substantially over time. This is one clear reason why subsequent benchmark estimates of relative prices are (typically) not consistent with relative inflation over the intervening period.

This conceptual problem is compounded by practical concerns. It has long been known that equation (6) does a poor job in predicting changes in PPPs over time,[16] but when the results of the ICP PPP comparison for 2011 were released (World Bank, 2014), the differences with the previous, ICP 2005, results were very large despite the serious global effort that went into both sets of PPPs. As detailed in Deaton and Aten (2017) and Inklaar and Rao (2017), part of the inconsistency was due to biases introduced in the measurement of ICP 2005 PPPs, but even after correcting for theses biases the differences remained substantial. Furthermore, shifts in the bundles of products cannot fully account for these differences, leaving 'measurement error' of some sort as the main (though not very informative) explanation.

This view matches that of Maddison, who argued that the difference between observed PPPs in successive ICP rounds and extrapolations based on relative inflation was more likely due to errors in the ICP estimates than errors in the national growth measures. Reconciling different benchmarks with the time series was in his eyes not the preferred method for long-term comparisons. The basis for this argument was a study by Kravis and Lipsey (1991), who also suggested that estimates of growth rates should be taken from the national accounts, whereas estimates of real GDP per capita should be done by benchmark studies (Maddison, 1995, p. 164).

Yet the approach of Maddison has notable limitations. For one, if any given benchmark comparison of prices and income is imperfect and perturbed by measurement error, relying fully on a single benchmark comparison would mean that the same error would affect real GDP per capita estimates through the decades or centuries. Second, while time series of GDP per capita growth (i.e. g_{jt}) may be considered reliable in modern times for many countries, periods like the World Wars, or periods of economic instability such as in much of Latin America in the 1980s diminish the reliability of statistics. The situation is more problematic in countries with

[16] See Deaton and Heston (2010) for notable contributions to this discussion.

poorly developed statistical systems, such as in many African countries, which can lead to unreliable growth figures.[17]

This was illustrated by Prados de la Escosura (2000), who argued that PPPs based on extrapolations as in equation (6) led to implausible results. His solution was to rely on the regularity of the price-income relationship to estimate what relative prices (and, as result, income levels) would have been if we had been able to observe them historically, see also Klasing and Milionis (2014). Relying heavily on such estimates is less appealing to us, most importantly because there are still important aspects of the price-income relationship that are not fully understood. For example, Hassan (2016) argues that the price-income relationship is non-linear and negative, rather than positive at the lower income levels and Zhang (2017) argues that mismeasured differences in product quality bias the price-income relationship. That said, comparing price levels rather than only income levels can serve as a useful check on relative income estimates derived according to a given methodology, see e.g. appendix E on income levels in oil-rich economies. For the MPD more broadly, we implement a multiple-benchmark approach as detailed in the following section.

4. Implementation

4.1 The MPD measurement approach

In the new version of MPD, we implement a multiple benchmark approach based on post-1950 ICP benchmarks and three types of historical benchmarks, namely independent real GDP per capita benchmarks from historical studies; benchmarks derived from real wage comparisons and data on urbanization; and estimates of living standards as multiples of subsistence.[18] Using the methodology developed for PWT (Feenstra et al. 2015), we subsequently tie the long term income series from the MPD (2013) to the relative income levels, thereby taking into account relative price changes between the different benchmark years. This means the MPD estimates for a particular country and year can be based on direct benchmark estimates, interpolation between benchmarks or extrapolation from the first or last benchmark, following equation (5). To enable users to distinguish between these different types of observations, we introduce clear labeling in the MPD. Furthermore, given the differences in the types of benchmark, we also clearly label which type of benchmark is used to derive a certain estimate.

[17] See e.g. Henderson, Storeygard and Weil (2012), Young (2012) and Jerven (2013).
[18] Additionally, we use estimates of PPPs for 1960 from the study of Braithwaite (1968).

As discussed in the previous section, problematic estimates in benchmarks or time series can have substantial consequences over longer periods of time. Given our stated goal of more closely aligning to our understanding of living standards in history, this requires a degree of judgement when implementing our multiple benchmark approach. In particular, it can be the case that a) benchmark relative price estimates diverge substantially from what might be expected from an estimated price-income relationship using all ICP benchmark PPPs observations; b) income levels can drop below subsistence for sustained periods of time; or c) income levels can remain high, in direct contradiction to the historical record. These observations result in a list of judgmental adjustments, by, for instance, excluding specific ICP PPP benchmarks or cutting short time series; see Appendix B for details. Some of such observations have also prompted greater efforts to develop additional benchmark estimates, as discussed below and in Appendices C and D. Category c observations consist of oil-rich economies whose current high income levels can be understood from large oil earnings, but where high income levels prior to major oil development or prior to high oil prices would run counter to the historical understanding of those countries; see Appendix E.

4.2 Historical benchmarks

Two kinds of historical benchmarks (covering the period before 1950) can be distinguished. Firstly, starting with the pioneering work by Rostas (1948) economists and economic historians have produced benchmarks of the relative income or output levels of economies (or parts of them, such as the manufacturing sector), including the construction of relevant PPPs to make real comparisons. Various methods have been used, making use of the output/value added approach, the income approach, and the expenditure approach. Usually, these studies compare the leading economy (US, UK) with one or more other economies (Germany, France, or Japan) (Broadberry 1998; Fukao et al. 2007). We basically collected all available historical economy-wide benchmarks and used them to re-anchor the historical time series following the PWT methodology described in the previous section; see Appendix A for a an overview of historical benchmarks and studies that we rely upon.

The currently available benchmarks are, however, spread very unequally over the world economy, in particular going back further in time. Therefore, we created an additional benchmark for a relatively large set of countries for 1800, making use of an indirect approach that is often used to estimate trends in GDP per capita in the period before 1850. For the pre-modern period, data constraints often make it impossible to estimate GDP in the usual way, and an indirect approach is used to construct its long-term development. The method used was

pioneered by Malanima (2003, for Northern Italy 1300-1850), and has now been used for many country studies (of, amongst others, Germany, France, Spain, Poland, Mexico, Peru). It is based on a number of assumptions: the demand for agricultural products is determined by real incomes as proxied by real wages, and a set of elasticities assumed to be constant, capturing income and price effects. This demand equation makes it possible to estimate the evolution of the output of foodstuffs per capita. Non-agricultural output is determined by the rate of urbanization (reflecting the relative size of the non-agricultural sector) and the relative productivity of services and industry compared to the agricultural sector (which can be inferred from a benchmark year). This basic set of assumptions makes it possible to estimate the trajectories of GDP growth in the period before 1850 on the basis of series of real wages, estimates of the rate of urbanization, and a set of assumptions about price and income elasticities (moreover, various studies have shown that the precise estimates of the elasticities of demand do not matter much for the estimated series).

We used this indirect approach not to estimate trends in GDP per capita, but to create a benchmark for 1800. In Maddison's dataset the early 19th century benchmark (of 1820) played an important role, as it linked the 19th century estimates with those of the pre-modern period. We created a similar benchmark for 1800 (for which date more data on the rate of urbanization are available), making use of the same set of assumptions. Firstly, we use real wage estimates resulting from the methodology developed by Allen (2001) and Allen et.al. (2011), which results in international comparative estimates of the purchasing power of the wages of unskilled construction workers in a large set of countries. The demand function is simplified to assuming that demand is a function of the real wage and a given income elasticity of demand (of 0.4). The rate of urbanization is based on available datasets (going back to Bairoch ****; Bosker et al. and Buringh). Finally it is assumed (following the similar studies into the long term trends in GDP) that labor productivity in agriculture is half that in services and industry combined. This method was used to create a benchmark for 1800 for Western European countries for which sufficient data were available (Great Britain, The Netherlands, Belgium, France, Germany, Austria, Poland, Italy, Spain, Portugal, Sweden and the Ottoman Empire) and for a number of South American countries (Argentina, Mexico, Peru and Venezuela).

In addition to the direct and indirect historical benchmarks, we follow Maddison's approach to also include estimates of comparative income levels for some of the very earliest (pre-1500) years. As data is increasingly scattered, Maddison used a variety of information to assess to what extent those societies had income levels notably above the level of subsistence, i.e. was

there sufficient surplus beyond subsistence for development. We update that approach by updating the subsistence line to $700 (2011 US dollars), in line with the $1.90/day global poverty line used by the World Bank (Ferreira et al. 2016).

4.3 Updating historical series

This new version of the MPD includes all new historical income estimates that have become available since the previous update (Bolt and Van Zanden, 2014). Such an update is necessary as new work on historical national accounts appears regularly and is important as it provides us new insights in long term global development. One of the goals of the Maddison project is to publish an update every two or three years. Additionally, the previous update included estimates up until 2010, and now estimates are available until 2015 which are integrated in the updated database.

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. This possible inconsistency in the dataset is therefore 'solved' by making use of the new, much improved set of estimates by Wu (2014). Most of the other additions to the Maddison project dataset relate to the period before 1914, as can be seen from table 1.

As is clear from this overview, in particular work on the early modern period (1500-1800) is producing more new time series over per capita GDP, often however making use of indirect methods to estimate its long term development. The 'model' for making such estimates based on the links between real wages, the demand for foodstuffs and agricultural output, which has been developed by Malanima (2003), Alvarez-Nogal and Prados de la Escosura (2013) and others, has now also been applied to Poland (Malinowski and Van Zanden 2016), Spanish America (Arroyo-Abad and Van Zanden 2015), and France (Ridolfi 2016).

Table 1. New Additions to the Maddison Project Database

Country	Period	Source
		Latin America
Mexico	1600-1812	Arroyo-Abad and van Zanden (2015)
Peru	1600-1812	Arroyo-Abad and van Zanden (2015)
Bolivia	1846-1950	Herranz-Loncan and Peres-Cajias (2016)
Panama	1906-1945	De Corso and Kalmanovitz (2015)
		Europe
Poland	1409-1913	Malinowski & Van Zanden (2015)
Finland	1600-1860	Eloranta, Voutilainen, Nummela (2015)
Norway	1820-1930	Grytten (2015) for GDP
Sweden	1300-1560	Krantz (2017)
Sweden	1560-1950	Schon and Krantz (2015).
Portugal	1550-1850	Palma and Reis (2015)
France	1250-1800	Ridolfi (2016)
		Asia
China	1952-2008	Wu (2014)
China	1661-1933	Xu, Shi, van Leeuwen, Ni, Zhang, and Ma (2016)
India	1600-1870	Broadberry, Custodis and Gupta (2015)

Finally, we have extended the national income estimates for all countries in the database to include the most recent years, up until 2015. For this we have used various sources. The Total Economy Database (TED) was used to extend the GDP per capita up to 2016 for all countries included in TED, similar to what has been done for the latest update of the Maddison Project database (Bolt and Van Zanden, 2014). For those countries not present in TED, we have used national accounts estimates from the UN to extend the GDP pc series. We have also used the TED and the International Data Base to extend the population estimates up until 2015.[19] Recently, 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. Lastly, we have extended the series for the former Czechoslovakia, the former Soviet Union and the former Yugoslavia to 2012, based on the series of their successor states.

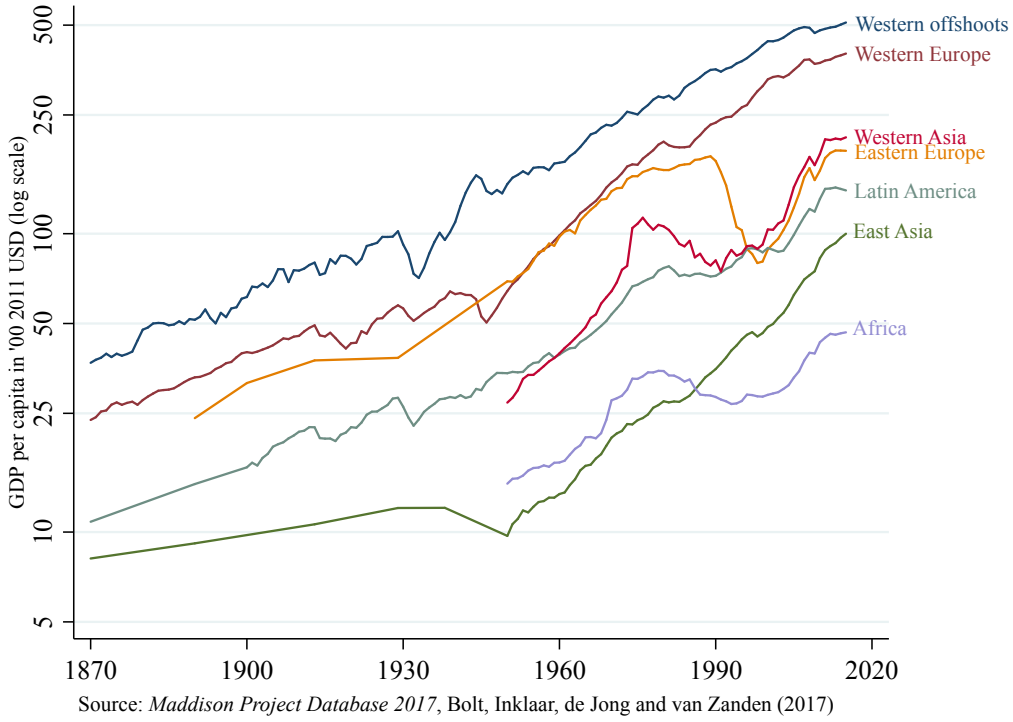
[19] International Data Base: United States Census Bureau. As Palestine and Kosovo are not included in these sources, we used data from the World Development Indicators

5 Applications

5.1 Shape of world income differences: regional perspective

Combining the multiple (historical) benchmarks with the long term series of per capita income from the Maddison Project database changes the pattern of long term income development compared to the original income series. In this section we discuss the major changes between the original series and the updated series presented in the paper on a regional level. It is important to realize though, that sometimes the effect is not only driven by integrating new relative prices but could also be the result of updates of the underlying national account statistics.

Figure 1, Average real GDP per capita across regions, 1870–2015



Source: *Maddison Project Database 2017*, Bolt, Inklaar, de Jong and van Zanden (2017)

For the two poorest regions in the dataset, Africa and East Asia, the pattern of long term development does not change drastically. For Africa, the average level of GDP per capita in 1950 is a bit lower in the new database compared to the original income estimates, while for East Asia it is roughly the same in both databases. However, for individual countries, the change in early income estimates are sometimes quite substantial. For Egypt for example the incomes are much lower in the new database due to the much lower GDP estimates from the 1985 and 1996 than relative to the 2011 estimates. For Africa as a whole, the main difference between

the two series is the more severe drop in incomes during the so called 'lost decades' of the 1980s and 1990s, see the figure and table below.

Table 2. Real GDP per capita by region and major countries for new and previous methodology, in 2011 US dollars

	1870		1910		1950		2011	
	MPD	Original Maddison	MPD	Original Maddison	MPD	Original Maddison	MPD	Original Maddison
Western Offshoots	3692	3749	7517	7610	14666	14881	48582	48582
United States	3736	3736	7586	7586	15241	15241	49675	49675
Western Europe	2375	3600	4545	5811	6425	8144	38040	38040
United Kingdom	3665	5334	5917	7409	9370	10620	34972	34972
France	2373	3086	4552	4878	7442	8531	36691	36691
Germany	2362	3711	5386	6755	5731	7832	43189	43189
Western Asia					2716	4485	20714	20714
Saudi Arabia					5151	9926	48470	48470
Iraq	1973	4812			3743	9125	11484	11484
Eastern Europe					6924	5588	17928	17928
Poland	1481	1844	2410	3295	3148	4771	21837	21837
Latin America	1083	1429	2169	2691	3407	4781	14133	14133
Brazil	790	1478	852	1594	1853	3466	14831	14831
Argentina	3197	2753	7658	7168	9121	9353	20003	20003
Mexico	840	1259	2131	3277	2900	4577	15210	15210
East Asia	815	793			971	999	8813	8813
China	696	717			669	606	10221	10221
India	878	713	1068	932	878	828	4768	4768
Japan	985	1177	1741	2082	2492	3067	34979	34979
Africa					1453	1636	4481	4481
Egypt	589	1709			827	2397	10737	10737
Nigeria					1921	1576	5136	5136
South Africa	1385	1917	2186	2733	5278	6019	11838	11838

Note: 'MPD' is based on the real GDP per capita figures based on the variable $CGDPpc$, column 'Original Maddison' is based on the figures for $RGDPNApc$, which are computed following the methodology as originally employed by Maddison for estimating real GDP per capita over time.

Using multiple benchmarks result in substantially lower average incomes for Latin America, most notably for the mid-20th century, where relative average income drops to 23 percent of the US level, down from 32 percent in the original Maddison database. The new methodology also clearly affects the pattern of average income development in Western Asia and Eastern Europe, again particularly after 1910. Incomes for Eastern Europe are now higher until the mid-1980s, with income levels on par with Western Europe around 1960 (which is also partly due to lower incomes in Western Europe, see below). In Western Asia the effects of using more relative

income estimates translates mainly in much lower incomes up until the mid-1990s after which increasing oil prices result in enormous increases in average incomes.[20]

Western Europe is the region for which most relative income estimates are available. Incorporating this the new information results in substantially lower income estimates for the region compared to US incomes. The original GDP estimates from Maddison indicate that the US and Western Europe were about on par around 1870, after which the US forged ahead of Europe until the end of WW2. Thereafter Europe's economies expanded rapidly, until average incomes reached around 73 percent of the US level during the 1970s. After this, relative incomes remained fairly stable until the present. As a result of integrating the historical benchmarks, Europe seems behind the US already substantially in the 1870. Growth rates of both the US and Europe's economies are then very similar until roughly the Great Depression. Then incomes initially diverge somewhat until the end of WW2, but Europe's incomes grow faster after 1950s to roughly 77 percent of the level of US incomes in 2011.

5.2. Poverty and subsistence

An important implication of using different relative price levels is that the poverty level may change. With the 1990 price levels, the subsistence level income was estimated at between 350 and 400 international dollars per year (Maddison, 2003). The poverty line was equal to around $ 1 per 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 the 1993 PPP's (Ravallion, Datt and van de Walle, 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 price level (in US dollars, the standard used in these calculations) increased by 59% between 1990 and 2011, bringing the poverty line to 636 dollars of 2011. Moreover, The World Bank slightly raised the absolute poverty line to 1,90 US dollars a day or 694 dollars per year, expressed in 2011 prices.

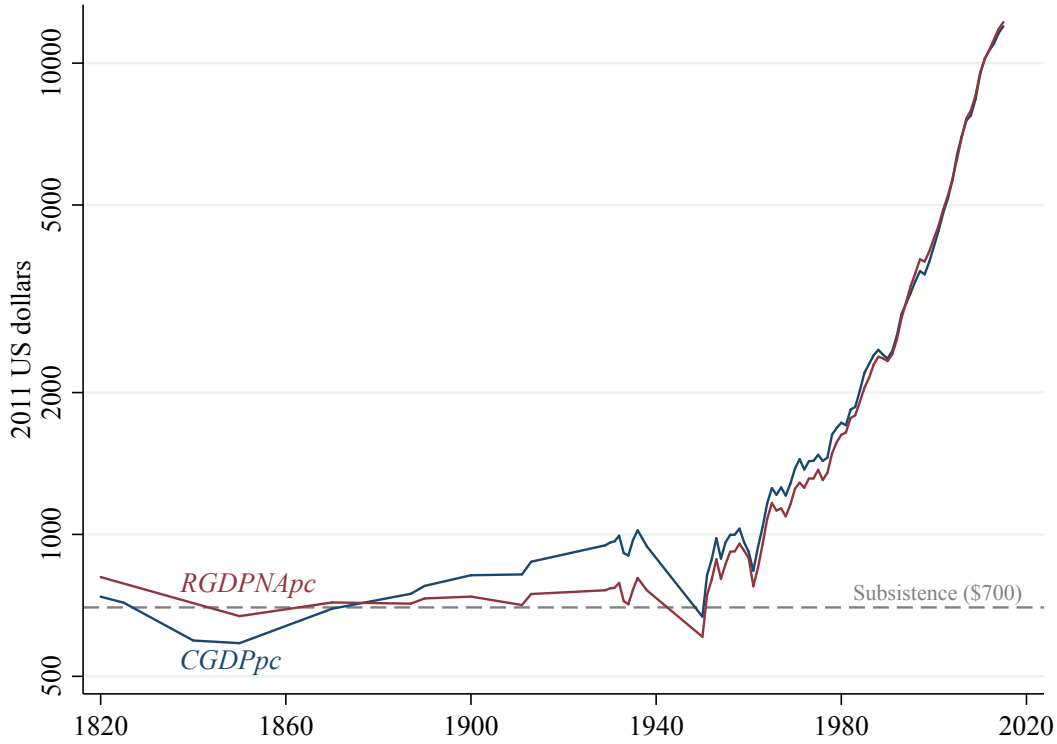
The effects of rebasing the original Maddison estimates has the most notable effects for countries who experienced substantial price changes relative to the US between the benchmarks years. China is an interesting case in this perspective. When the 2005 PPPs were released, the prices for China had increased so much relative to the US, that total GDP per capita came out around 40% lower than China's relative income based on earlier price estimates (Deaton and

[20] For a detailed discussion on income estimates for oil rich countries, see appendix *

Heston, 2010: 3; Feenstra et. al, 2015). 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 original Maddison project income estimates for China based on Wu (2014) which show on average lower growth between 1952 and 2011 than the previous (official) estimates. Combining the updated income series with the 2011 ICP incomes and the available historical benchmarks for China, results in plausible historical income estimates. Income in China never falls substantially below subsistence, also not for earlier periods, except for two decades in the early 20th century.

Figure 2, Historical income series China – *CGDPpc* versus *RGDPNApc*, in 2011 US dollars



Looking more broadly into the subsistence threshold, the original Maddison project dataset includes 123 observations below the subsistence income level of 400 dollars per year (on a total of 15372 that is 0.8%), of which most are countries in times of civil war such as Afghanistan, Somalia, Liberia, Rwanda, and the Democratic Republic of Congo. Using the new, multiple benchmark approach, the number of observations below subsistence increases, and most of

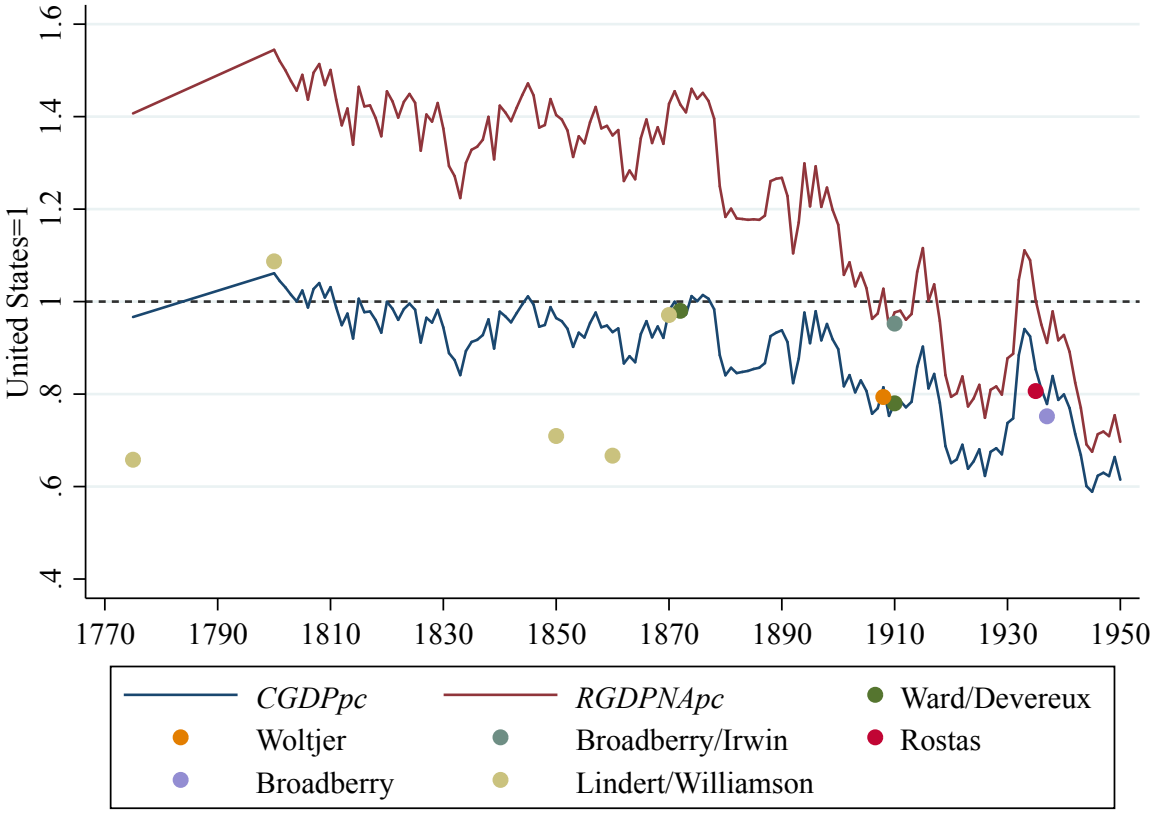
them can be found in Africa, especially prior to 1960 or during periods of civil war. Given that we know still too little about economic development in Africa in earlier years[21], we would need more information about relative incomes for earlier periods to sensibly interpret relative development levels. The best way forward would be to have benchmarks for 1950 for all African countries, as that would provide the best comparison of relative income levels in that year allowing us to 'anchor' the annual series.

5.3 When did the US overtake the UK?

One of the debates in the study of long term economic growth has focused on the relative performance of the UK and the USA, and in particular the question when the USA overtook the UK. Maddison's estimates making use of the backward projection from the 1990 benchmark year, implied that until the 1870s the UK economy was about 40% more productive than that of the USA, and that only after the 1870s the USA gradually overtook the UK (see Figure 3). Broadberry (1998, 2003) came to similar conclusions, based on a benchmark comparison in 1937. These results have however been criticized recently by Ward and Devereux (2003, 2016), who created a set of independent benchmarks for 1872 and 1910 period, and by Lindert and Williamson (2016) who did similar research for the 18th century, indicating that the USA was at least on par with the UK at the time. Our new approach, which makes use of these new Ward/Devereux benchmarks for the 1872-1910 period, also solves the problem of the 18th century, as the new estimates for that period are consistent with the new 1872 benchmark. The figure below shows the new results, which suggest that UK and USA income levels were comparable until the 1870s, at which point the US economy gained a sustained income advantage.

[21] Also the more recent income estimates for many African countries are sometimes of dubious quality (Jerven, 2013; Henderson et al. 2012).

Figure 3. Real GDP per capita in the United Kingdom relative to the United States



6. Concluding remarks

This paper has introduce a new version of the Maddison Project Database (MPD), the successor to Angus Maddison's historical statistics. The main novelty of our approach is to combine the full range of modern-day and historical estimates of relative GDP per capita with improved time series on growth in GDP per capita to provide a comprehensive, global picture of real GDP per capita from Roman times to the present. In contrast to Maddison's last work, which is still widely used, the new MPD matches more closely the best current evidence of living standards and income across countries in the present day and in history – by construction. This feature is of great importance, not just to provide a more historically-grounded perspective of real GDP per capita through history, but also as an invitation to the research community. By incorporating new estimates of historical income levels, the MPD can serve as a platform through which such new research results can be placed in international perspective and provided to a broader audience.

The incorporation of many historical benchmarks has an important effect on our understanding of long-term income trends. For example, the original Maddison statistics showed that it was

not until the early 20th century that real GDP per capita in the United States overtook the level of real GDP per capita in the United Kingdom. But as historical evidence has accumulated, it has become increasingly clear that real GDP per capita in the United States was at comparable levels as in the United Kingdom already at the time of US independence, in the late 18th century. More broadly, we find that in the 19th century, the United States was farther ahead of countries around the world with, in particular, lower levels of real GDP per capita in Western Europe and Latin America.

These new results do not claim to be the final word on these topics. Despite our inclusion and estimation of numerous historical benchmarks, our understanding of comparative income levels becomes based on sparser data as we move farther away from modern day. This is particularly pressing in regions such as Africa and large parts of Asia. Our hope is that our research contributes as a fresh impetus to improving our understanding of historical income differences as this can only sharpen our understanding of why a relatively small set of countries managed to become much richer and to what extent those countries were different. As an example of what research can achieve, take Broadberry and Wallis (2017), who find that avoiding a shrinking economy has been much more important than stimulating growth for reaching higher income levels. More fine-grained information and more comparative studies are crucial to broadening and deepening such understanding.

Finally, we fully recognize that developing estimates of real GDP per capita is but a first step to a broader understanding of wellbeing. A fuller picture of well-being would (at least) distinguish between consumption and investment and, more generally, incorporate additional dimensions of wellbeing, such as health, leisure and inequality. A better understanding of differences in income and living standards would require information on the factors of production – human and physical capital – and productivity. Yet all such subsequent work relies heavily on reliable and informative data on income per capita and we hope that our new data serves as a useful starting point and platform for further research.

Appendices

A. Historical benchmarks

Table A1. Historical benchmark levels of real GDP per capita relative to the United States

Frankema et al. (2013)	1820	1850
Netherlands (original vs. GBR)	0.83	0.86
Van Zanden and Li (2012)	1825	
China (original vs. NLD)	0.32	
Broadberry, Guan and Li (2013)	1840	
China (original vs. GBR)	0.23	
Fremdling (1991)	1860	
Germany (original vs. GBR)	0.46	
Van Zanden (2003)	1860	
Indonesia (original vs. NLD)	0.23	
Heston and Summers (1980)	1870	
India	0.24	
Ward and Devereux (2016)	1872	1910
Australia	1.20	1.07
United States	1.00	1.00
Belgium	0.91	0.92
Canada	0.75	0.83
United Kingdom	0.98	0.78
Switzerland	0.68	0.74
Denmark	0.72	0.71
Netherlands	0.74	0.71
Germany	0.67	0.71
Norway	0.58	0.63
France	0.68	0.60
Sweden	0.49	0.59
Italy	0.37	0.35
Fukao, Ma and Yuan (2007)	1935	
Japan	0.32	
Taiwan	0.23	
China	0.11	
Korea	0.12	
Broadberry (2006)	1935	
Germany (original vs. GBR)	0.65	

Note: whenever a real GDP per capita comparison was based on another reference country, the relative income level is restated relative to the United States. For comparisons originally relative to the United Kingdom, the Ward and Devereux (2016) estimates of relative GDP per capita were combined with the GDP per capita growth rates

for earlier periods to derive a United Kingdom/United States relative income series. Whenever the Netherlands serves as the reference country, the Ward and Devereux (2016) levels are used or, in an additional step, the Frankema et al. (2013) estimates relative to the United Kingdom.

B. Multiple benchmark implementation

Section 4 outlined general principles for excluding benchmarks or (parts of) time series in constructing the MPD. This Appendix will provide an overview of the specific decisions made.

C. Latin America

Providing new estimates of income levels in Latin America has proven to be a particular challenge as the combination of relative price benchmarks with current National Accounts data often leads to notably lower income levels with especially implausible implications over the long run. This appendix discusses how the new estimates relate to those of Maddison (2006, Table A2-g), motivating our main decisions and their drawbacks along the way.

Table C1. GDP per capita in 1980 for 16 Latin American countries based on current market exchange rates

	ICP 1980		Current National Accounts		Current
	USD (XR)	USA=1	USD (XR)	USA=1	NA/ICP 1980
Argentina	7384	0.65	11100	0.89	1.50
Bolivia	942	0.08	631	0.05	0.67
Brazil	2059	0.18	1537	0.12	0.75
Chile	2477	0.22	2615	0.21	1.06
Colombia	1295	0.11	1660	0.13	1.28
Costa Rica	2119	0.19	2616	0.21	1.23
Dominican Republic	1192	0.10	1468	0.12	1.23
Ecuador	1463	0.13	1771	0.14	1.21
El Salvador	743	0.06	774	0.06	1.04
Guatemala	1085	0.09	958	0.08	0.88
Honduras	674	0.06	949	0.08	1.41
Panama	1816	0.16	1919	0.15	1.06
Paraguay	1403	0.12	1255	0.10	0.89
Peru	1196	0.10	926	0.07	0.77
Uruguay	3459	0.30	3438	0.28	0.99
Venezuela	4621	0.40	4304	0.35	0.93
United States	11448	1.00	12468	1.00	1.09

Sources: ICP 1980 from United Nations (1987, Table 1), Current National Accounts based on Maddison GDP series and United Nations National Accounts Main Aggregates Database (https://unstats.un.org/unsd/snaama/), December 2015; except El Salvador (World Bank, World Development Indicators).
Notes: USD (XR) refers to United States dollars, converted using exchange rates. Maddison (2006, Table A2-g) reports that GDP levels were adjusted upwards after the ICP 1980 report to correct for underestimation of the informal sector for Argentina (+36%), Peru (+6.5%) and Venezuela (+17.2%). The reported figures in this table incorporate these adjustments. The 'Current National Accounts' figures rely on the source's 'IMF-based exchange rates' . The 'Current NA/ICP 1980' column divides the two USD (XR) GDP/capita figures.

For most countries in Latin America, Maddison (2006) relied on the ICP 1980 benchmark (United Nations, 1987), which he subsequently extrapolated to 1990.[22] For most countries, this 1980 comparison is also the first benchmark comparison that is incorporated in the new Maddison series and thereby the anchor for earlier years, making it a sensible starting point of this discussion. Table 1 compares the GDP per capita figures used by Maddison and figures based on current vintage data, from the UN National Accounts Main Aggregates Database. Both sets of figures are converted to US dollars using market exchange rates, so the principal difference between the two sets are revisions to National Accounts data.[23,24] Such revisions can relate to changing accounting systems, such as from the System of National Accounts (SNA) of 1968 to SNA 1993 and SNA 2008, but also to changes of the benchmark year and source revisions. These revisions are known to be substantial, see for an overview of the impact of the adoption of SNA 1993 in Latin America, see Olinto Ramos, Pastor and Rivas (2008). They show that the adoption of new accounting rules and the change in benchmark year led to sizeable changes in the level of nominal GDP, ranging from -8.2 to +19.2 percent.

National Accounts revisions are typically considered to lead to an improved perspective on economic activity. However, these revisions pose a challenge for the pre-revision period. Ideally, a country's statistical office would rework its previous set of National Accounts based on new accounting rules and estimation methods. In practice, a typical revised set of National Accounts will include estimates for several years based on updated methodologies. However, extending revisions for longer periods of time is usually not a priority at statistical agencies, so a more common practice is to 'splice' together the old time series and the new, i.e. assume that the new level is a better reflection of economic activity but that economic growth was not mis-measured in the past.[25] Prados de la Escosura (2016) criticizes this 'retropolation' approach, specifically for the case of Spain, and proposes an alternative interpolation method. However,

---

[22] Exceptions are Jamaica and Mexico, neither of which gives rise to particular concerns, so we focus on the 16 countries from ICP 1980.

[23] Note also that several of these countries underwent one or more currency reforms or switched to or from using US dollars. We assume these changes to the unit of account have been appropriately carried out. This is not guaranteed as UN National Accounts data imply that GDP levels in El Salvador (in US dollars) were only one third of the level before the change from Colones to US dollars, while the World Bank's World Development Indicators show a level that is broadly comparable.

[24] Population estimates may also be revised; for instance the current estimate of the population of the United States in 1980 is 229.6 million compared to 227.7 million in United Nations (1987). In most cases these revisions are small, though in some cases they represent a notable part of the overall change in GDP per capita figures. For example, in Colombia population increased from 25.79 to 27.74 million, a revision of 7.6 percent.

[25] One prominent counterexample is the US Bureau of Economic Analysis, which maintains a current database of National Income and Product Accounts from 1929 onwards.

to what extent this approach would be generally applicable is hard to assess. Furthermore, the challenges in the Latin American context are greater because of the numerous episodes of macroeconomic instability in the twentieth century, with large swings in prices and exchange rates and currency reforms or dollarization.

Table C2. GDP price level in 1980 for 16 Latin American countries

	ICP 1980, GK	ICP 1980, GEKS	PWT
Argentina	1.42	1.66	1.71
Bolivia	0.59	0.72	0.67
Brazil	0.62	0.70	0.72
Chile	0.68	0.86	0.92
Colombia	0.47	0.54	0.58
Costa Rica	0.68	0.77	0.81
Dominican Republic	0.59	0.69	0.92
Ecuador	0.57	0.69	0.71
El Salvador	0.52	0.61	0.64
Guatemala	0.47	0.60	0.63
Honduras	0.56	0.67	0.71
Panama	0.56	0.76	0.82
Paraguay	0.67	0.80	0.85
Peru	0.45	0.55	0.57
Uruguay	0.83	0.86	0.91
Venezuela	0.73	0.94	0.93
United States	1.00	1.00	1.00

Sources: ICP 1980, GK from United Nations (1987, Table 1); ICP 1980, GEKS computations based on ICP 1980 data; PWT: price level of GDP^{o} (pl_gdpo) from PWT 9.0, normalized to USA=1, see Feenstra, Inklaar and Timmer (2015).
Notes: GDP price level is the PPP for GDP divided by the exchange rate. GK stands for Geary-Khamis and GEKS for Gini-Elteto-Kovecs-Szulc as two alternative PPP computation methods, see e.g. Diewert (2013).

Beyond revisions to GDP per capita figures, changes to the computation of purchasing power parities (PPPs) is another source of differences between Maddison (2006) and current estimates. Table C2 provides three estimates for the 1980 GDP price level, defined as the PPP for GDP divided by the market exchange rate. The first column, 'ICP 1980, GK', is the main relative price level as reported in United Nations (1987), computed from detailed price and expenditure data using the Geary-Khamis method. This method has come under increased criticism in the period since the ICP 1980 data were released, primarily because the method suffers from

substitution bias, see e.g. Diewert (2013). The alternative GEKS method, does not suffer from this shortcoming and is currently the method of choice for computing PPPs (see e.g. World Bank, 2014).

`The 'ICP 1980, GEKS' column shows that this change in computation method has a substantial effect on price levels, increasing prices of all countries relative to the United States, by an average of 20 percent. The final column, 'PWT', uses the relative price data as provided in the Penn World Table, version 9.0. The main reason for differences with the GEKS column is that PWT includes estimates of PPPs for exports and imports of goods, see Feenstra et al. (2015). This further increases price levels, by an average of 6 percent, though for the Dominican Republic the effect is much larger. These two differences together serve to depress relative income levels in Latin America compared to Maddison's (2006) estimates.

Table C3. GDP per capita in PPP-converted US dollars in 1980 for 16 Latin American countries

	GDP per capita in USD (PPP)		Relative GDP per capita (US=1)	
	Original	Rebased	Original	Rebased
Argentina	5210	6488	0.46	0.52
Bolivia	1591	938	0.14	0.08
Brazil	3337	2144	0.29	0.17
Chile	3622	2836	0.32	0.23
Colombia	2784	2874	0.24	0.23
Costa Rica	3137	3240	0.27	0.26
Dominican Republic	2006	1601	0.18	0.13
Ecuador	2583	2509	0.23	0.20
El Salvador	1418	1204	0.12	0.10
Guatemala	2324	1529	0.20	0.12
Honduras	1204	1334	0.11	0.11
Panama	3220	2345	0.28	0.19
Paraguay	2108	1481	0.18	0.12
Peru	2663	1634	0.23	0.13
Uruguay	4180	3771	0.37	0.30
Venezuela	6317	4625	0.55	0.37
United States	11448	12468	1.00	1.00

Sources: see Tables 1 and 2

Notes: GDP per capita in USD (PPP) is computed as GDP per capita in USD (XR) from Table 1 divided by the GDP price level from Table 2. For example, the 'Original' estimate for Argentina is computed as 7384/1.42=5210, while the 'Rebased' estimate is computed as 11100/1.71=6488.

Table C1 showed the impact of GDP per capita revisions and Table C2 showed the impact of changes in PPP computation methods. Table C3 combines these two factors in estimates of PPP-converted GDP per capita. As the final two columns show, the combination of factors result in lower levels of GDP per capita relative to the United States, with the changes from GK to GEKS PPPs as the most systematic factor. Heterogeneity across countries is substantial, with Colombia's relative income level remaining almost unchanged, while Peru's relative income level declines from 23 percent to 13 percent of the US level. This heterogeneity is primarily due to differences in GDP per capita revisions (Table 1): in Columbia GDP per capita has been revised upwards, while in Peru it has been revised downwards.

Table C4. GDP per capita in PPP-converted 2011 US dollars for 16 Latin American countries – 1800-2011

	1800	1850	1900	1950	1980	2011
Argentina	2312	2835	5887	9121	15409	20003
Bolivia			790	1627	2229	5331
Brazil	758	758	752	1853	5092	14831
Chile	712	1034	2494	4172	6737	19705
Colombia	819	681	946	2984	6825	11788
Costa Rica				2742	7696	12366
Dominican Republic				1593	3802	11679
Ecuador			923	2497	5959	9985
El Salvador				1618	2859	7607
Guatemala				2040	3631	6650
Honduras				2257	3168	4421
Panama				2215	5569	16762
Paraguay				1333	3517	7377
Peru	1013	732	642	1505	3881	10044
Uruguay	1423	1924	2908	6108	8956	17211
Venezuela	1182	1820	1330	5552	10985	17746
United States	1980	2825	6252	15241	29613	49675

Source: Maddison Project Database.
Notes: The column 1980 implies the same relative income levels as the 'rebased' figures from Table C3; e.g. for Argentina 15409/29613=0.52. The difference in figures is because the numbers in Table 3 were 1980 US dollars, while all figures in this table are expressed in 2011 US dollars.

As may be expected, these changes to the 1980 benchmark levels have substantial effects on long-run income levels. Table C4 shows the PPP-converted GDP per capita figures for the 16 Latin American countries since 1800 (or the earliest available year), denominated in 2011 US dollars. Especially the incorporation of indirect, real wage-based benchmarks for Peru (1800) and Venezuela (1830) ensure that real GDP per capita levels do not fall below the $700 subsistence line for prolonged periods of time, though Peru's income level is below that line at the start of the 20th century. This points to the importance of a more extensive set of benchmark comparisons for these countries, to establish whether these patterns, implied by the time series of economic growth and the different benchmarks accords with the historical reality.

D. Africa

The new methodology for extending income levels back in time appear to have limited consequences for the income estimates of the far majority of Sub-Saharan African countries. For a number of African countries the new methodology results in substantially different income levels compared to those originally published by Maddison (2006). This appendix discusses how the new estimates and the original estimates relate to each other.

Maddison (2006) relied on the Penn-World tables 5.6 for his estimates of the 1990 benchmark GDP levels for all African countries except Equatorial Guinea, Sao Tome and Principe, and five very small countries. For those countries Maddison (2006: 221) assumed that the 1990 GDP per capita was equal to the average of the 50 African countries covered by the PWT. When using the multiple benchmarks, for 26 out of 47 countries this in effect means using the 2011 and the 2005 PPPs. For the other countries, various earlier benchmarks are available. Kenya is the first Sub-Saharan African country to participate in an ICP round in 1970, and in 1985 already 15 countries participate in the ICP program. Using the multiple benchmark approach leads to very similar relative income estimates compared to the original Maddison series for about half of the African countries. For most other countries the multiple benchmark approach leads to slightly different but we think still reasonable income estimates. However, for some of the countries, the earlier income estimates deviate away from previous patterns when using multiple benchmarks. This leads to a reshuffling in the order of countries compared to the original series, when we rank them from poor to rich. The best way forward would be to have benchmarks for 1950 for all African countries, as that would provide the best comparison of relative income levels in that year allowing us to 'anchor' the annual series.

Using all benchmarks available to estimate long term income series results in very low income estimates for seven African countries. For some years, average incomes for these countries are below subsistence for different periods in time. The most problematic results are for Mozambique, Liberia, Burundi and Mali. The trend in Mozambique's current income series clearly show the effect of the civil war[26] on its economy. Income drops significantly between a few years prior to the war and the end of the war (1973 – 1992). But the new

[26] The civil war in Mozambique officially begun in 1977 and lasted until 1992. The trend is similar to the original estimates but due to different PPPs (2011 and 2005), the new income levels using the multiple benchmark approach is much lower.

income estimates show incomes below subsistence from 1950 until 2003, with an exception of 1969 – 1973. The original estimates in contrast never fall below subsistence.

Liberia's income is also much lower using multiple benchmarks. The only benchmarks available for Liberia are 2005 and 2011, hence the new income estimates are the result of much higher prices in in the 2000s compared to 1990. Despite the much lower income levels, average incomes always remain above subsistence with the exception of the civil war period (1989 until 1997. For the Burundi the new methodology results in income estimates on or around subsistence for most of the period after 1975. Prior to 1975, average income is declining and always below the subsistence level. A new earlier benchmark could provide new insides in earlier income developments. From the four most problematic countries, Mali is the only one with more benchmarks available than the most recent two. Mali also participated in the 1985 and the 1996 ICP round. However, including these earlier benchmarks lead to substantially lower income estimates, and below subsistence GDP pc estimates prior to 1985. From the lower income levels during the 1980s and 1990s, Mali has experienced more rapid economic growth since the mid-1990s compared to what earlier income estimate suggest. This could reflect the fact that previous income estimates for many African countries actually underestimated the recent economic expansion of the continent (Young, 2012).

Finally, for Sierra Leone, Burkina Faso and Lesotho, income estimates are also lower using the new methodology, but incomes only fall below subsistence prior to 1960. For these countries especially, a benchmark for 1950 could aid in providing a sensible anchor to the earliest income estimates available.

E. Oil-rich countries

Some of the largest differences between the original Maddison 1990 benchmark figures and more recent figures can found in oil-exporting economies. Table 1 illustrates this for 13 countries with data for 1990. Not all these countries were major oil exporters already in 1990, but for most of these countries substantial differences can be seen between GDP per capita based on the original Maddison 1990 figures and the new estimates based on multiple (ICP) benchmarks (MBM). The most extreme case is United Arab Emirates, whose income level was 61 percent of the US level in 1990 according to Maddison, but 343 percent of the US level according to the multiple benchmarks estimates.

Table E 1, Relative income levels and oil dependence in 1990 for selected countries

Country	ISO code	Oil share	GDP per capita (USA=100)			ICP coverage
			Maddison	MPD	Difference	
Equatorial Guinea	GNQ	0	7	5	-2	2005, 2011
Iran	IRN	18	15	13	-2	All except 1980
Iraq	IRQ	3	11	20	9	2005, 2011
Venezuela	VEN	20	35	20	-15	1980, 1996, 2005, 2011
Algeria	DZA	12	12	22	10	2011
Trinidad and Tobago	TTO	29	40	26	-14	1985, 1996, 2011
Oman	OMN	45	28	30	2	1996, 2005, 2011
Bahrain	BHR	15	18	39	21	1996, 2005, 2011
Saudi Arabia	SAU	32	38	48	10	2005, 2011
Kuwait	KWT	40	28	63	36	2005, 2011
Libya	LBY	n.a.	13	66	52	none
Qatar	QAT	59	30	94	64	1996, 2005, 2011
United Arab Emirates	ARE	4	61	343	282	2011

Notes: 'oil share' is the share of fuels and lubricants exports in real GDP (source: PWT 9.0). Maddison GDP per capita figures are computed based on the Maddison (2008) database, MPD is the new Maddison Project Database figures.

Since many of these countries did not participate in ICP benchmarks before 2005, the MPD estimates for 1990 are predominantly based on extrapolations using National Accounts figures. The heavy reliance on oil exports means that swings in terms of trade will have a substantial effect on GDP and places a substantial burden on statistical offices in the countries to produce accurate price and volume estimates. The required statistical capacity is not uniformly available

in this set of countries. It is therefore sensible to also present the 2011 relative income and price levels, in Table E2.

The table illustrates that even in this benchmark year, several countries show levels of GDP per capita that are close to or (substantially) exceed US levels. The relative price levels, in turn, are comparatively low as income levels increasing towards those in the US would normally imply relative price levels rising to similar levels as well.

Table E 2, Relative income and price levels and oil dependence in 2011 for selected countries

Country	ISO code	Oil share	GDP/capita (USA=100)	Price level (USA=100)
Equatorial Guinea	GNQ	82	88	55
Iran	IRN	15	36	44
Iraq	IRQ	35	23	43
Venezuela	VEN	30	36	61
Algeria	DZA	25	27	41
Trinidad and Tobago	TTO	38	60	61
Oman	OMN	44	89	48
Bahrain	BHR	0	73	61
Saudi Arabia	SAU	39	98	48
Kuwait	KWT	53	155	62
Libya	LBY	n.a.	27	50
Qatar	QAT	60	313	57
United Arab Emirates	ARE	41	131	61

Notes: see Table D1. Price level refers to the price level of GDP^{o} from PWT 9.0.

To illustrate the peculiarity of GDP per capita figures in this set of countries, it is helpful to also compare the level of domestic absorption per capita (consumption plus investment, i.e. GDP excluding net exports) and consumption per capita (including household and government consumption). Table 3 shows this comparison for 2011. Especially relative consumption levels are substantially lower than relative GDP, with Equatorial Guinea as a striking example, where GDP/capita was 88 percent of US level, but consumption/capita only 13 percent. All countries in this comparison have consumption/capita levels well below those in the US.

So a first conclusion has to be that for this set of countries in particular, GDP per capita is a poor measure of current living standards, with so much of income not consumed or invested

domestically, but invested abroad. “Squaring the circle”, whereby GDP per capita would accurately reflect current living standards would, in turn, require unrealistically high price levels of GDP.[27] A second conclusion, based on the comparison between Tables 1 and 2, is that some of the National Accounts time series are suspect. For instance, why would the United Arab Emirates have a GDP/capita level in 1990 that is even higher than in 2011 while oil exports where only one-tenth as large (relative to GDP)?

Table E 3, Relative levels of GDP, domestic absorption and consumption in 2011

Country	ISO code	GDP	Dom. Absorp.	Consumption
Equatorial Guinea	GNQ	88	45	13
Iran	IRN	36	31	28
Iraq	IRQ	23	19	19
Venezuela	VEN	36	29	27
Algeria	DZA	27	23	20
Trinidad and Tobago	TTO	60	44	43
Oman	OMN	89	61	46
Bahrain	BHR	73	55	50
Saudi Arabia	SAU	98	72	52
Kuwait	KWT	155	76	64
Libya	LBY	27	n.a.	n.a.
Qatar	QAT	313	164	61
United Arab Emirates	ARE	131	99	76

Source: PWT 9.0

Note: 'Dom. Absorp.’ is domestic absorption.

For the final comparison, it is helpful to contrast the Maddison 1990 income levels with the relative price levels implied by these income levels. This requires additional data on GDP per capita converted to US dollars using market exchange rates. To put the magnitudes of these

[27] To be more precise: for the relative level of GDP/capita in (for example) Equatorial Guinea to be equal to its relative consumption/capita level would imply a GDP price level of 372 percent of the US, a level never observed in international price comparisons.

relative price levels in context, the final column presents standardized residuals, based on comparing the implied price levels to predicted price levels from a Balassa-Samuelson price-income regression for (the log of) all benchmark price level observations and (the log of) GDP per capita converted to US dollars using market exchange rates. It is clear that the price levels required to arrive at the Maddison 1990 relative income levels are often unrealistically high and sometimes unrealistically low, using a standardized residual of approximately ±2 as a standard for realism. So the third conclusion would be that Maddison 1990 income estimates are not a plausible alternative to the MBM income estimates.

Table E 4, Relative income and price level in 1990 based on Maddison and price residual

Country	ISO code	GDP/capita (USA=100)	Price level (USA=100)	Standardized Residual
Equatorial Guinea	GNQ	7	22	-1.97
Iran	IRN	15	47	0.03
Iraq	IRQ	11	489	7.10
Venezuela	VEN	35	28	-2.39
Algeria	DZA	12	81	1.98
Trinidad and Tobago	TTO	40	44	-1.21
Oman	OMN	28	98	1.49
Bahrain	BHR	18	232	4.42
Saudi Arabia	SAU	38	79	0.43
Kuwait	KWT	28	137	2.38
Libya	LBY	13	236	4.93
Qatar	QAT	30	218	3.40
United Arab Emirates	ARE	61	193	1.75

Note: 'Standardized residual' is the difference between the relative price level implied by Maddison's relative income estimates and the price level predicted by the Balassa-Samuelson price-income relationship, standardized by the standard deviation.

References

Arroyo Abad, L., E. Davies and J. L. van Zanden (2012), "Between conquest and independence: real wages and demographic changes in Spanish America, 1530–1820" *Explorations in Economic History* 49(2): 149–166.

Allen, R.C. (2001), "The great divergence in European wages and prices from the Middle Ages to the First World War" *Explorations in Economic History*, 38(4), pp. 411–447.

Allen R.C., J.P. Bassino, D. Ma, C. Moll-Murata and J.L. van Zanden (2011), "Wages, Prices, and Livings Standards in China, Japan, and Europe, 1738-1925" *Economic History Review* 64(1): 8–38.

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.

Asian Development Bank (2014).

Balk, Bert M (2008), *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference* Cambridge, UK: Cambridge University Press.

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

Broadberry, S.N. (1998), "How Did the United States and Germany Overtake Britain? A Sectoral Analysis of Comparative Productivity Levels, 1870-1990" *Journal of Economic History* 58(2): 375-407.

Broadberrry, S.N. (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., 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

Burstein, Ariel and Gita Gopinath (2014), "International prices and exchange rates" In Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff (eds.), *Handbook of International Economics,* Vol. IV, pp. 391–451. Amsterdam: Elsevier.

Ciccone, A. and M. Jarociński (2010), "Determinants of Economic Growth: Will Data Tell?" *American Economic Journal: Macroeconomics* 2(4): 226–246.

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

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. 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. (201), 'Understanding PPPs and PPP-based National Accounts' *American Economic Journal: Macroeconomics*, 2: 1–35.

Diewert, W. Erwin. (1976) "Exact and Superlative Index Numbers." *Journal of Econometrics* 4 (2): 115–45.

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

Feenstra, R.C., Hong Ma, J. Peter Neary and D.S. Prasada 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).

Fukao, K., D. Ma, and T. Yuan (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(3): 503–537.

Gallardo-Albarrán, D. (2017), "Missed opportunities? The development of human welfare in Western Europe, 1913-1950" *GGDC Research Memorandum* no. 166.

Grytten, O.H. (2015). "Norwegian Gross Domestic Product by industry 1830 – 1930" *Norges Bank Working paper* no. 19.

Hassan, F. (2016), 'The Price of Development: the Penn-Balassa-Samuelson effect revisited' *Journal of International Economics,* volume 102, pp 291-309.

Henderson, J.V., A. Storeygard and D.N. Weil (2012), "Measuring economic growth from outer space" *American Economic Review* 102(2): 994–1028.

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

Heston, A. (2010), "What can be learned about the Economies of China and India from Purchasing Power Comparisons", in Barry Eichengreen, Poonam Gupta and Rajiv Kumar, eds, Emerging Giants, China and India in the World Economy, Cambridge University Press, pp. 3-31.

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

Irwin, Douglas A., (2005) "The Welfare Cost of Autarky: Evidence from the Jeffersonian Trade Embargo, 1807–09" *Review of International Economics* 13, 631–645.

Jerven, M. (2013), "Comparability of GDP estimates in Sub-Saharan Africa: the effect of revisions in sources and methods since structural adjustment" *Review of Income and Wealth* 69(S1): S16–S36.

Johnson S., W. Larson, C. Papageorgiou and A. Subramanian (2013), "Is newer better? Penn World Table Revisions and their impact on growth estimates" *Journal of Monetary Economic*s 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 (1991) "The International Comparison Program: Current Status and Problems", in P.E. Hooper and J.D. Richardson, International Economic Transactions: Issues in Measurement and Empirical Research, Chicago, University of Chicago Press.

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.

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

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

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

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

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

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

Maddison, A. (2007a), *Chinese Economic Performance in the Long Run, Second Edition, Revised and Updated, 960-2030 AD*, Paris, OECD Development Center.

Malinowski, M. and Van Zanden (2015), "National income and its distribution in preindustrial Poland in a global perspective" *EHES working paper* no. 76.

O'Rourke, Kevin H. (2007), "War and Welfare: Britain, France and the United States, 1807–14" *Oxford Economic Papers* 59: i8–i30.

Olinto Ramos, R., G. Pastor and L. Rivas (2008), "Latin America: Highlights from the implementation of the *System of National Accounts 1993 (1993 SNA)*" *IMF Working Paper*, no. WP/08/239.

Paige, D. and Bombach, G. (1959) *A Comparison of National Output and Productivity of the United Kingdom and the United States* Paris: Organisation for European Economic Co-operation.

Palma, N. and Reis, J. (2015), "Portuguese demography and economic growth, 1500-1850" Unpublished manuscript

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

Prados de la Escosura, L. (2016), "Mismeasuring long run growth: the bias from splicing national accounts – the case of Spain" *Cliometrica* 10(3): 251–275.

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

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

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.

Schon, 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 (2015).

Subramanian, A. (2011), "Eclipse: Living in the Shadow of China's Economic Dominance" Peterson Institute for International Economics, Washington D.C..

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

United Nations (1987), *World Comparisons of Purchasing Power and Real Product for 1980. Phase IV of the International Comparison Project. Part Two: Detailed Results for 60 countries*, United Nations: New York.

Ward, M. and Devereux, J. (2016). New Perspectives on Productivity and Living Standards in the late 19th century. *mimeo*

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

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): 1–22.

World Bank (2008), *Global Purchasing Power Parities and Real Expenditures: 2005 International Comparison Program*, World Bank: 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. (2011), "Accounting for China's Growth in 1952-2008 - China's Growth Performance Debate Revisited with a Newly Constructed Data Set," RIETI Discussion Paper Series 11-E-003.

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* (forthcoming)

Young, A. (2012), "The African Growth Miracle" *Journal of Political Economy* 120(4): 696–739.

Zhang, Q. (2017), 'The Balassa–Samuelson relationship: Services, manufacturing and product quality' *Journal of International Economics* 106: 55–82.