

Documentation for

GDP Per Capita by Purchasing Power Parities

for countries and territories

Gapminder Documentation constitutes work in stepwise progress.
We welcome all sorts of comments, corrections and suggestions through e-mail to the author.

Gapminder Documentation 001

Version 8

Uploaded: 2010-04-26

Published by: The Gapminder Foundation,
Sweden, Stockholm, 2008

Author: Mattias Lindgren

E-mail: [mattias.lindgren \(at\) Gapminder.org](mailto:mattias.lindgren@gapminder.org)

Gapminder is a non-profit foundation
promoting sustainable global development
by increased use and understanding of statistics.
www.gapminder.org



ABBREVIATIONS

GDP	Gross Domestic Product
GNI	Gross National Income
ICP	the International Comparison Program
PPP	Purchasing Power Parities
UNSTAT	United Nations Statistical Division
WDI	World Development Indicators

1. Introduction

This is the documentation for the Gapminder compilation of Gross Domestic Product (GDP) per capita by Purchasing Power Parities (PPPs). The data is used in the interactive graph *Gapminder World* available at:

www.gapminder.org/world

This written documentation is accompanied by an Excel-file, which includes both the detailed meta-data, as well as the actual observations. This Excel-file will henceforth be referred to as the “Excel-file”. The Excel file, as well as this document, are both available at:

www.gapminder.org/downloads/documentation/#gd001

The goal has been to include as many countries and territories as possible. A guiding principle has been to use the most transparent and well-documented source whenever possible; however, preference was also given to sources with as wide coverage of countries and territories as possible.

The main purpose of the data is to produce graphical presentations that display the magnitude of income disparities in the world over time. Therefore, we have also included very rough estimates for countries and territories for which reliable data were not available. These estimates can only be taken as an indication of the order of magnitude for the indicator. Furthermore, we have not been able to make sure that every single observation is based on the best estimates available. Hence we discourage the use of this data set for statistical analysis and advise those who require more exact data to investigate the available data more carefully and look for additional sources, when appropriate.

!

NOTE: The observations for the period before 1950 are, in the majority of cases, based on rough estimates within a range of likely values. In many cases we have no information, what-so-ever, on the relative ranking of countries.

There are at least two purposes for including very rough estimates:

- a) Rough estimates give a starting point for a discussion of the data. Unrealistic estimates become more apparent if they are displayed in a format that is both accessible to many as well as easy to assess. This, we hope, might elicit comments and criticism from others, which would help us to improve the data step by step.
- b) We want to display the broad patterns of development of the world. Hence, the observations before 1900 (or even 1950) are less concerned with describing the relative position of individual countries. Rather, we want to display our best guesses for the development of the world as a whole.

1.1 The countries included in the dataset

For a discussion on what countries and territories we try to cover, and how we try to handle border changes and the like, see the document “Countries and Territories in Gapminder World”. The basic principle, however, is to get estimates for the geographical areas corresponding to the present borders.

To be succinct, we will hereafter jointly refer to all countries and other types of geographical entities and territories as “countries”, irrespectively of their statehood status. The inclusion of any area in this data set does not, in any way, imply a stated opinion of Gapminder on the legal status of the area.

It is not always clear to what extent certain semi-autonomous or disputed territories also are included in the data for their “mother country”, so there may be some “double counting” in these instances.

1.2 Basic principles

The data in this dataset is based on GDP per capita, in fixed 2005 prices, and is adjusted for Purchasing Power Parities (PPPs), as calculated in the 2005 round of the International Comparison Program (ICP).

The PPP rates could be considered as an estimate of what the exchange rates would have been if one dollar would buy the same amount of goods and services in every country. The PPP rates are based on price comparisons between countries. The estimations of the PPP rates for 2005 are described in International Comparison Program (2008).

Our starting point is the relative income levels for 2005, as implied by the ICP data. We then link these levels to the national real growth rate for each country, taken from other sources. This means that cross-country comparisons for years far away from 2005 contain large compounded errors. Hence, we make a couple of adjustments to iron out the most blatant inconsistencies for the earlier years.

The unit of measurement for GDP per capita is quite abstract in the sense that (international) dollars mainly have a meaning in terms of what they can buy. Hence, our main interest is not the absolute *level* of incomes as such, but rather the *relative* incomes of countries.

We have two principal ways of making all the observations comparable. The first option is to use the real growth rates of each country as they are given by the national sources, and link these series to cross-country data for one benchmark year, as they are given by some international comparison. This is illustrated in Figure 1 below, where 2005 has been chosen as the benchmark year. Y is the GDP per capita. In this case the relative positions of countries are a residual for all years except the benchmark year. This method is the method used by Maddison (2003) and several others.

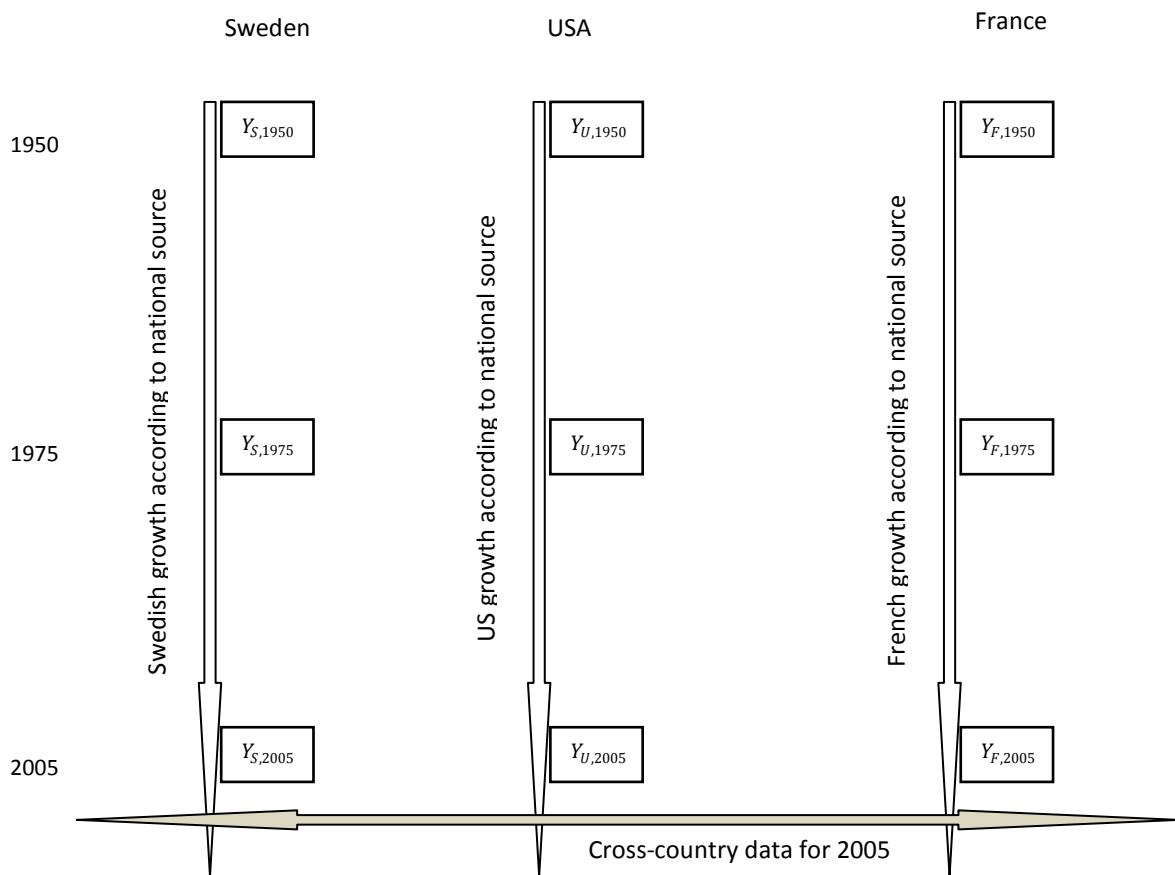


Figure 1. Illustration of the use of one benchmark year.

The second option is to use the relative positions of the countries for each year as given by a sequence of cross-country comparisons. We then link the real growth rate of one chosen country to these cross-country comparisons. This is illustrated in Figure 2 below, where the U.S. has been chosen as a benchmark country. In this case the real growth rates are a residual for all countries except the benchmark country (e.g. the U.S. in the figure below).

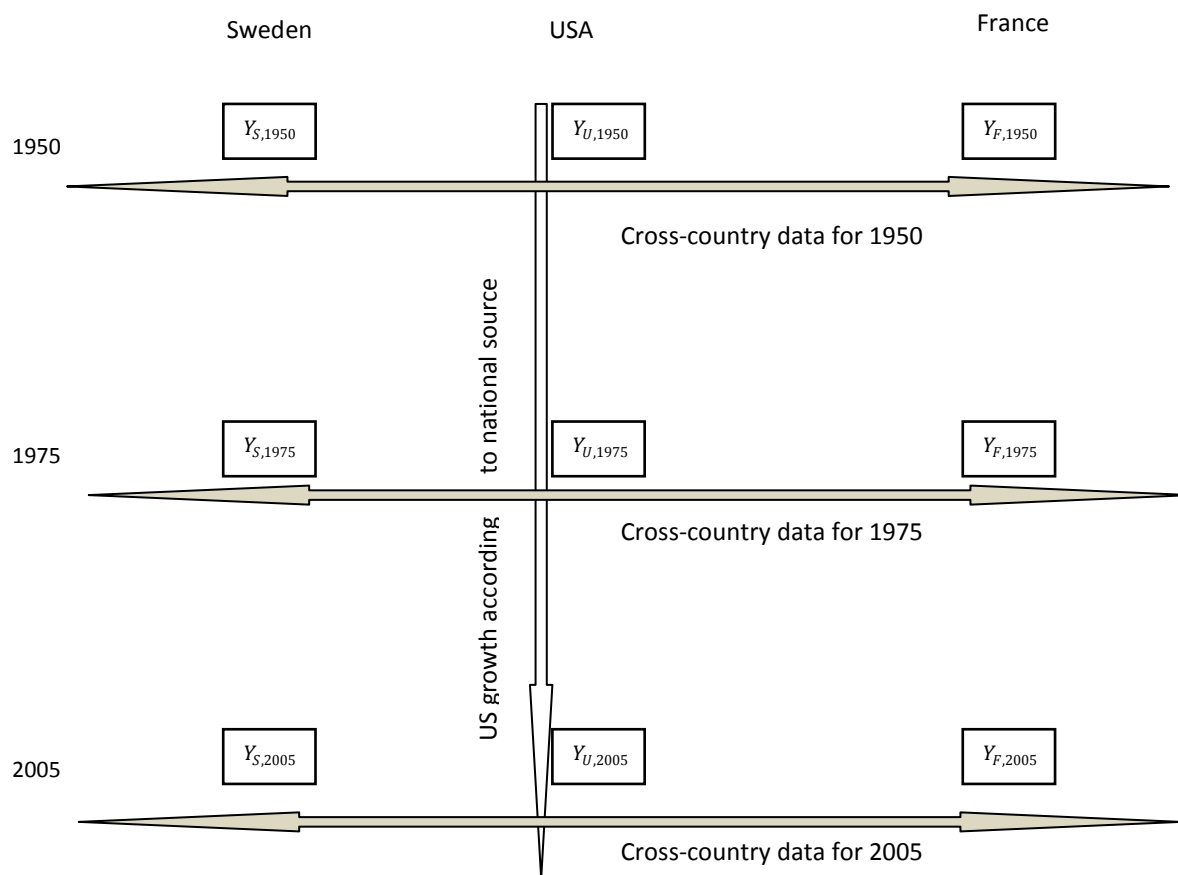


Figure 2. Illustration of the use of one benchmark country.

The choice of method does not make any difference to the results if all the observations are calculated in the same way and if all the relative prices are the same. However, since this almost never is the case, the choice of method matters for the outcome. Furthermore, when we use a fixed benchmark year, i.e. the first method, we are likely to compound the errors as we go back in time from the benchmark year. One alternative solution is to combine the two methods, e.g. by taking some average. The Penn World Tables (Heston et al, 2006) integrate aspects of this method, although they too mainly use the national growth rates.

The relative positions of countries for a given year are more visible than the magnitudes of the growth rates in the graphical presentations we use. However, since the data availability for cross-country comparisons is exceptional for one specific year, 2005, we choose to rely on that year as a benchmark year. Furthermore, international comparisons of GDP per capita, based on PPP rates, are only available from 1970, and we want to go back much further. At this stage we have not been able to find or integrate much information based on direct comparisons between countries for earlier years.

However, even with our limited information we judged that some of the cross-country comparisons for the earlier years were unrealistic. In those cases we accepted some ad-hoc adjustments to the growth rates.

1.3 Overview of the work process

Our starting point is the ICP data for GDP per capita for 2005 recalculated with the ICP PPP rates. Other sources provide national real growth rates. In many cases we have to link several such national sources together.

In a case where two national sources, Z and W , have been linked at year s , the GDP per capita for country i in year t , becomes:

$$Y_{i,t} = I_{i,2005} * \frac{Z_{i,s}}{Z_{i,2005}} * \frac{W_{i,t}}{W_{i,s}}$$

$I_{i,2005}$ is the GDP per capita by PPP provided by the ICP. It is calculated by ICP as the GDP per capita in local prices, $n_{i,2005}$, multiplied by the price-ratio $PPP_{i,2005}$, so that $I_{i,2005} = n_{i,2005} * PPP_{i,2005}$. $Z_{i,t}$ is the real GDP per capita as provided by the source Z , and $W_{i,t}$ is the real GDP per capita as provided by the source W .

Our work consisted of the following steps:

1. First we used the GDP per capita by PPPs for 2005, as provided by the ICP (including their estimated figures).
2. We extended this cross-section for 2005 to an additional number of countries, using our own estimates.
3. We linked a number of sources of national growth to each other and to the 2005 benchmark data.
4. Then, when needed and possible, we extended the national growth rates with the growth rates of country groups, e.g. the growth rates of former states, or with the growth rates of neighbouring countries.
5. Then, when needed, we used adjusted estimates of the regional averages, to get coverage for most countries back to 1820.
6. For a couple of countries we deemed it necessary to adjust the growth rates for the whole time series. This was done to avoid overly absurd consequences for the earlier years.
7. In a few cases we had to do additional adjustments, as commented on below.
8. Finally we did some rough extrapolations back to 1800.

The process is described in more detail below.

2. Compiling data for the benchmark year 2005

The first step is to find GDP per capita data for the benchmark year 2005, using the PPP of 2005. ICP provides GDP per capita data for 144 countries. These 144 observations are the “official” ICP data, based on direct price information. The 144 countries exclude Burundi which had price data, but no GDP data.

The ICP also supplies estimates for a number of countries, in addition to the 144 official observations. These ICP estimates are based on a regression analysis that uses the GDP per capita of ICP as the dependent variable. The independent variables were Gross national Income (GNI) per capita, by exchange rates (the Atlas method) and Gross enrolment in secondary school. They refer to this model as “model 6”, and the results were:

$$\begin{aligned} \text{ICP model 6: } \quad \text{Ln}(\text{GDP per capita PPP}) &= 0,3404 + \\ &+ 0,6994 * \text{Ln}(\text{GNI per capita, exchange rate}) \\ &+ 0,2292 * \text{Ln}(\text{Secondary School Gross Enrolment}) \end{aligned}$$

The predicted values from this model were used for countries lacking official ICP data, but which had observations for the two independent variables. Some countries lacked GNI per capita. For those countries an alternative model was used, using GDP per capita, by exchange rates instead, referred to as “model 7”:

$$\begin{aligned} \text{ICP model 7: } \quad \text{Ln}(\text{GDP per capita PPP}) &= 0,1987 + \\ &+ 0,7147 * \text{Ln}(\text{GDP per capita, exchange rate}) \\ &+ 0,2422 * \text{Ln}(\text{Secondary School Gross Enrolment}) \end{aligned}$$

For more details see International Comparison Program (2008), p. 164 ff. A more ambitious attempt of modelling is described in World Bank (2008).

The official ICP figures are likely to vary a-lot in quality. However, the only specific note we found so far is that the data for China is less robust. This will be discussed in greater detail further on in this document.

Nevertheless, several other figures seem problematic, e.g. since they imply levels below what could reasonably be considered the starvation level. In one case, Congo Dem. Rep., we made an upward adjustment of the 2005 value with 25%. The reasons and details for this adjustment are discussed in section 5.4 below.

2.1 Extending the 2005 data to more countries

For countries that did not even have estimates from the ICP we did our own, even rougher, estimates, based on a number of alternative sources: (a) The (old) homepage of UNSTAT; (b) Maddison on-line; (c) CIA world fact book (most of their data had to be extrapolated or interpolated to 2005); (d) Rough estimates provided by the World Bank.

Each of these sets of data had to be adjusted to be comparable to the ICP data. We deal with each of the data sources below in order of priority. The order of priority is mainly based on whether the data is based in the prices of 2005 and on the degree of coverage. Hence, Maddison only comes second despite the fact that he uses data based on PPPs rather than exchange rates, as UNSTAT does.

a) UNSTAT

The homepage of the UN statistical division, unstats.un.org, provide data on GNI per capita by exchange rate in prices of 2005. The indicator is called "GNI, per capita GNI - US Dollars" in the list on the homepage. Note that we use a slightly different data set from UNSTAT for the time series. To keep them apart we call the data-set we use here for "UNSTAT I" and the dataset we use for the time series "UNSTAT II".

There were other possible sources of GNI per capita by exchange rate that we also could have used. One such source is the World Development Indicator (WDI). Another related source is a special PDF from the World Bank that had included slightly more countries than the WDI.

The World Bank data was not fully consistent with the UNSTAT data. For some countries the difference was larger than 30%. However, a couple of rough regressions did not reveal any systematic differences between the sources. We choose to use the UNSTAT data, since they had the best coverage.

The unit of measurement in the UNSTAT data is quite different from the ICP data, i.e. GNI instead of GDP and exchange rate instead of PPPs. Hence we need to adjust the data to make them (somewhat) comparable, essentially by doing a simpler version of the ICP model:¹

$$\begin{aligned} \text{Gapminder model (1): } \ln(\text{GDP per capita, by PPP}) &= \\ &= \alpha + \beta * \ln(\text{GNI per capita, by exchange rate}) + \text{unmeasured factors} \end{aligned}$$

In this model we use the observations for which we had both ICP (non-estimated) data and UNSTAT data. This gives us 144 observations in total. The results are shown in Table 1 below. The observations and predicted values are shown in Figure 3 below.

	<i>Coefficients</i>	<i>p-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
α	2,51	0,00	2,33	2,70
β	0,77	0,00	0,75	0,79
R^2	98,48			

Table 1: Estimation results for "Gapminder model 1". Number of observations: 144.

¹ Some of the models used by ICP are "constrained" in the sense that they force the predicted value for the U.S. to be a certain value. This is done by indexing the variables and forcing the constant to be zero. We have not done that. The difference between our unconstrained and a constrained model is, in any case, not very big.

There was a systematic difference between the two series in the expected way, i.e. using exchange rates systematically implies a lower relative position for poorer countries. A few of the predicted values differed from the actual values with a factor of two, i.e. the predicted values were almost twice the actual values (as for Iran) or almost half the actual values, as for (the unadjusted values for) Congo, Dem. Rep. In the majority of cases the deviation were less than +/- 25%.

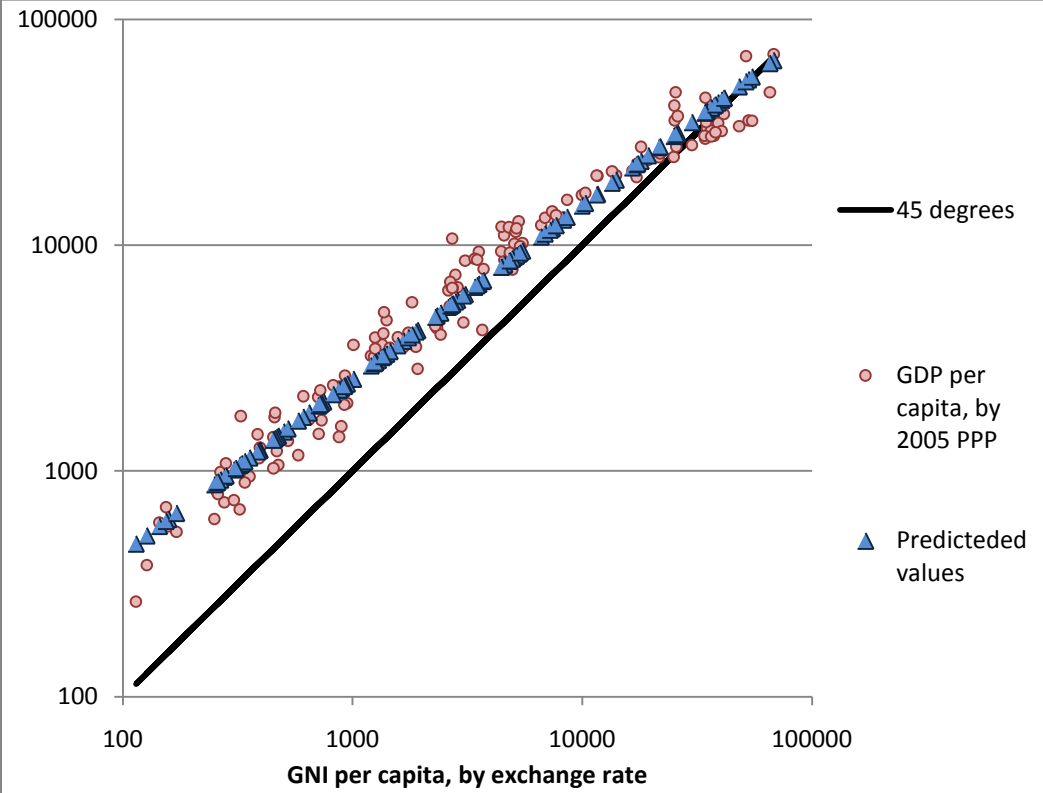


Figure 3: Actual and predicted values for “Gapminder model 1”.

The final step was to use the model above to adjust the data of UNSTAT. This adjusted data could be used to fill in the gaps for the countries which had UNSTAT data, but no ICP data.

b) Maddison

Maddison (2008 I), which we henceforth refer to as “Maddison online”, provides data for GDP per capita, by PPP, in the prices of 1990, and using PPP data from 1990. The main difference from the ICP data is accordingly that Maddison use PPP data from the 1990 benchmark and expresses his data in 1990 prices.

To adjust the data we use the following regression:

$$\begin{aligned}
 \text{Gapminder model (2): } \ln(\text{GDP per capita, by PPP, ICP}) &= \\
 &= \alpha + \beta * \ln(\text{GDP per capita, by PPP, Maddison}) \\
 &+ \text{unmeasured factors}
 \end{aligned}$$

In this model we use the observations for which we had both ICP (non-estimated) data and the data of Maddison. We also include a number of countries which are not included in the final dataset (e.g. former countries). This gives us 133 observations in total. The results are shown in Table 2 below. The observations and predicted values are shown in Figure 4 below.

	Coefficients	p -value	Lower 95%	Upper 95%
α	-0,06	0,82	-0,56	0,44
β	1,04	0,00	0,98	1,10
R^2	94,94			

Table 2: Estimation results for “Gapminder model 2”. Number of observations: 133.

The model does not reveal any systematic differences between Maddison and ICP, which is not that surprising since both series use PPPs. We could perhaps have expected that the 15 years of inflation between 1990 and 2005 would have meant higher nominal values in Maddison in a more systematic way. This “inflation effect” seems to have been drowned in the general noise stemming from very different PPP ratios in 2005. However, the direction of the change is in the expected direction, i.e. the 2005 data is on average higher.

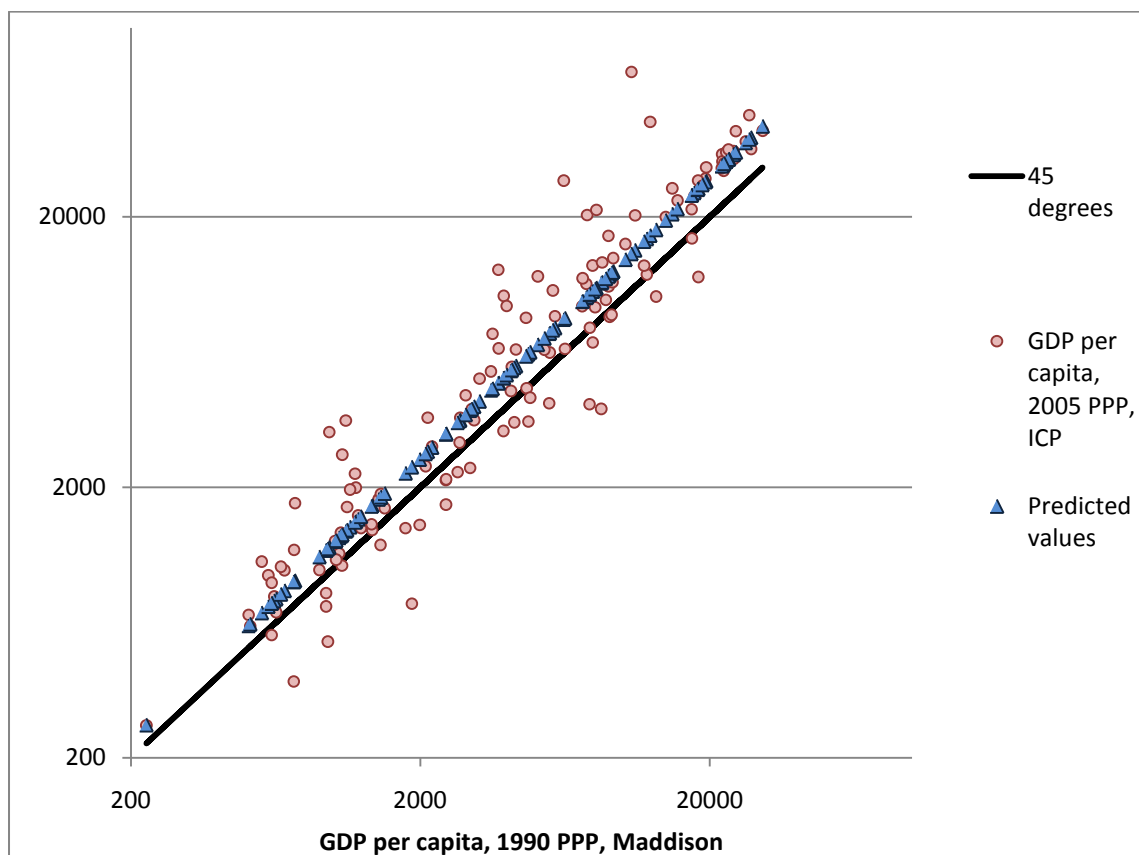


Figure 4: Actual and predicted values for “Gapminder model 2”.

Figure 4 illustrates the size of the revisions in the PPP rates for 2005 in the ICP, compared to the earlier PPP rates used by Maddison and others. The biggest differences are found among major oil-exporters, e.g. Qatar for which the 2005 data is more than 6 times as high as the 1990 data and almost 5 times as high as the predicted values. This, however, is the biggest outlier (it is represented by the dot most high up in the graph).

We use the model, despite all these shortcomings, due to a lack of better alternatives. Hence, we used the model above to adjust the data of Maddison. This adjusted data could be used to fill in the gaps for the countries which had Maddison data, but no ICP or UNSTAT data.

c) *CIA World Fact Book*

The CIA World Fact Book 2008 has data for GDP per capita (PPP), although we were unable to find the exact definition and sources for these data. Accordingly, we do not know the extent by which the figures are comparable across countries or with the ICP data.

Furthermore, for many countries the data referred to years other than 2005. Sometimes we were able to find data for 2005 in older issues of the fact book (2006 and 2007), that are available from the homepage of the University of Missouri University Libraries. For the remaining countries we had to interpolate or extrapolate to get data for 2005. Occasionally we had national growth rates which we could use to extrapolate to 2005. In most cases we simply had to assume that the level was unchanged.

Even though we had no information on how the CIA data differed from ICP we did a regression similar to the earlier ones²:

$$\begin{aligned} \text{Gapminder model (4): } \ln(\text{GDP per capita, by PPP, ICP}) &= \\ &= \alpha + \beta * \ln(\text{GDP per capita, by PPP, CIA}) \\ &+ \text{unmeasured factors} \end{aligned}$$

² The observant reader will note that there is no "Gapminder model 3". The present numbering only reflects a certain degree of "path dependency" when it comes to the labelling of our Excel sheets, it does not reflect any deeper mysteries than that.

We include all the overlapping observations, which gives us 142 observations. The results are shown in Table 3 below. The observations and predicted values are shown in Figure 5 below.

	Coefficients	p-value	Lower 95%	Upper 95%
α	-0,461	0,05	-0,927	0,0049
β	1,0498	0,00	0,9969	1,1027
R^2	95,74			

Table 3: Estimation results for “Gapminder model 4”. Number of observations: 142.

The regression does not reveal any systematic differences; although, again, there are substantial differences for many countries, as can be seen below. We use the model anyway out of principle, to get estimated values for the missing countries.

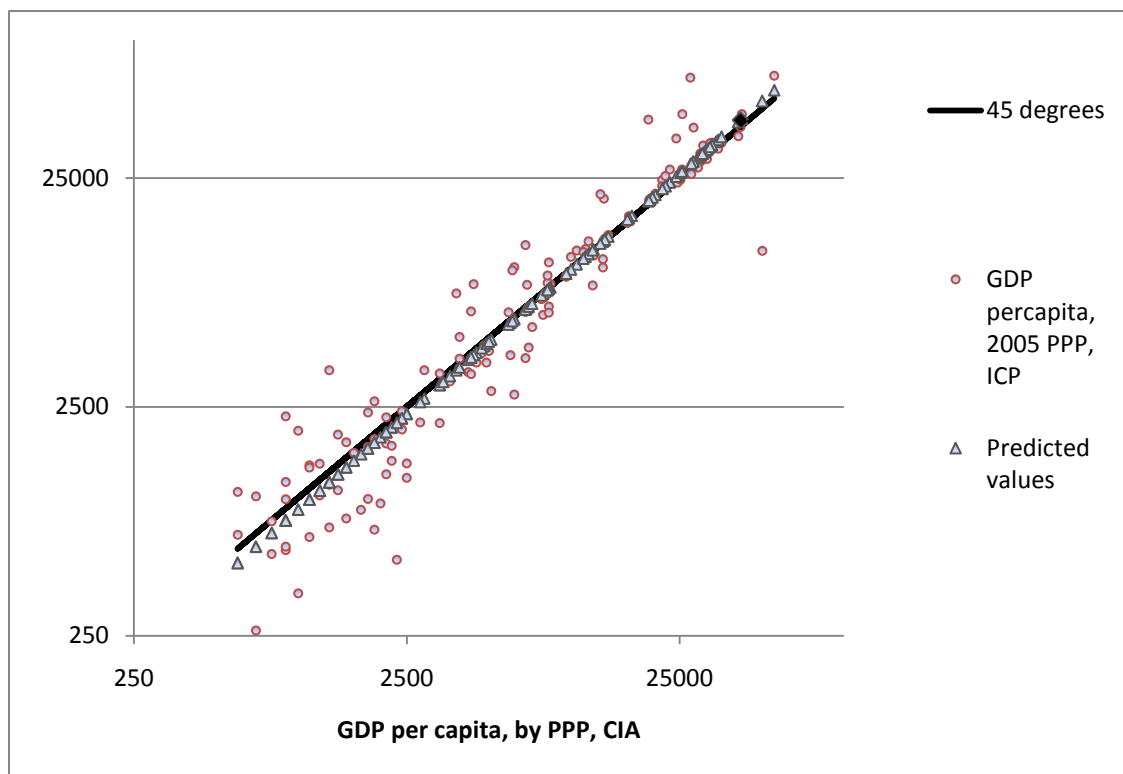


Figure 5: Actual and predicted values for “Gapminder model 4”.

d) *Rough estimates by the World Bank*

For the remaining countries we used the rough estimates available in the PDF of the World Bank (World Bank, 2005). Similar to the UNSTAT data above, it was expressed in GNI per capita, in exchange rates (Atlas method), in the prices of 2005.

In the case of Isle of Man the source provided a rough estimate of the relative income ranking of the country. In that case we used the geometric average of the country with one rank higher and the country with one rank lower.

In a few remaining countries the World Bank only provided an estimate of which income group the country belonged to, as displayed in Table 4 below. The World Bank also provided an average income level for each income group. Hence, we used the average level for the corresponding country.

Income group	Definition	Estimated average income
Low income	\$875 or less	\$580
Lower middle income	\$876 to \$3 465	\$1 918
Upper middle income	\$3 466 to \$10 725	\$5 625
High income	\$10 726 or more	\$35 131

Table 4: The income groups used by World Bank (2005). All incomes are in US \$ per year, by exchange rates.

The World Bank estimates use in principle the same measurement as the UNSTAT data above, i.e. GNI per capita by exchange rate in 2005 prices. Hence, we re-used Gapminder model 1 to adjust the data, even though the estimates in question are so rough so that it really does not make much difference.

2.2 The quality grading of the PPP rates

The quality of the PPP rates (i.e. the price-ratios) is one of several factors that affect the overall quality of a GDP per capita observation. Hence, we assign a quality grade to each PPP-observation (i.e. one grade per country for 2005). These grades should be considered as an upper bound of the quality. The grades are: Very Good; Good; Fair; Poor; Very Poor. We assign them according to the following principles:

- Good: Official ICP data, as long as there are no footnotes that indicate extraordinarily bad (or good) data quality.
- Fair: Official ICP data, with caveats about the data quality
- Fair: Official ICP estimates (ICP model 6 and 7)
- Poor: Official ICP estimates adjusted by gapminder (e.g. Congo Dem. Rep.)
- Very Poor: Gapminder estimates based on the World Bank assessment of what income group the country belongs to
- Poor: All other Gapminder estimates

Note that this data quality only refers to the quality of the PPP rates. A country could in principle have good PPP data, but bad income data, which implies a bad final quality of the observation.

3. Linking real growth series to the values of 2005

The next step is to link national real growth in GDP per capita to the levels we have calculated for 2005. We use a variety of sources for these time series, which we will describe below. In many cases we had to link different sources to each other.

When data at the country level were unavailable, we used the growth rate of a larger group of countries to which the country belonged. This often was the case for new countries that had emerged from a larger former state, such as Croatia emerging from Yugoslavia. Another regional growth rate that we occasionally used was the aggregated regions in Maddison on-line. When we use the same regional growth rates for several countries we preserve the relative positions between of these countries as they were at the first year for which we have data.

When we did not even have regional data, we used the growth rate of a neighbouring country that was judged to have had a similar development. In some cases we simply assumed that the GDP per capita had been exactly the same as in the neighbouring country.

To indicate the existence of longer gaps in the data we have added a special “source type” in the Excel sheet, called “no data”. The Gapminder World software automatically interpolates linearly between two observations. However, when the gap in the data is too long the automatic linear interpolation creates unreasonable results, i.e. the growth rate for early years will be unrealistically high. To avoid this, we, in some instances, included a couple of observations based on an exponential interpolation.

Countries with large oil incomes create special problems, which will be discussed in detail further on. If there was no other solution at hand we simply added a rough guesstimate for a year sometime before the year of the oil discovery. We mention it here because the meta-data for “arbitrary guess” is included in columns concerning the real growth. Finally, for a few countries we found no growth data at all, even though we had GDP data for the benchmark year from the ICP. In these cases we only include data for 2005.

We checked whether there were any major discrepancies in the growth patterns of the different sources, before linking sources to each other. When the discrepancies were large we had to make a judgement on what growth pattern made most sense. In most cases this judgement had to be of a rather ad-hoc character. These decisions, as well as some other minor corrections or adjustments, is documented in the Excel file, in the column “Other Footnotes”.

3.1 Sources for the growth rates

We use a variety of sources for the national growth rates. Many of these sources are regularly revised. Hence, when data for new years became available it is common that also previous years are adjusted. Ideally we would like to revise our whole dataset as these adjustments become available. However, for practical reasons, we normally adjust only the latest years in our dataset.

The datasources are described on the next page, in their approximate order of priority.

- a) Maddison on-line is our main source for long-term growth.
- b) Barro & Ursúa (2008) have updated the data of Maddison, using more recent estimates. They revise the figures of Maddison in some cases and they also filled some of the gaps or interpolations made by him. Whenever there were a major difference between them and Maddison, we used Barro and Ursúa, otherwise we continued using Maddison.
- c) From the homepage of UNSTAT we used the per capita GDP in current U.S. dollars, deflated with the implicit price deflators in national currency. We call this data-set “UNSTAT II”, to separate it from the one we used for the cross-country data for 2005. We have used two different revisions of the UNSTAT data, one from 2008 and one from 2009.
- d) From the Penn World Table, mark 6.2, we used real GDP per capita, in Laspeyres Constant Prices (RGDPL).
- e) The IMF provides the “World Economic Outlook database”, which also includes projections for the GDP per capita data.
- f) From the World Development Indicators 2008 we used the GDP per capita, PPP, in constant 2005 international \$.
- g) From the World Development Indicators 2008 we also used data for GDP per capita, in constant local currency units. These were used when WDI lacked data expressed in PPPs, but had data available in local currencies. The *level* of these observations differed, of course, from the WDI data expressed in constant PPPs, but the *growth rates* were the same. Since we are only using the growth rates, the difference in level is of no importance.
- h) The homepage of Eurostat (the statistical agency of the EU) provided “GDP per inhabitant in purchasing power standards”, which we used for a few countries. Purchasing power standards is essentially the same as PPPs.
- i) For a few countries we took data directly from their national sources (e.g. the home pages of the national statistical offices or some other specific publication). The national sources include the St. Helena Development Agency (2008), FEPS (2007), the Isle of Man Treasury (2008) and INSEE (2008). Sometimes the data was expressed in current prices, in which case we deflated the figures with the available price index.
- j) Maddison (2003) provides some additional data that are not found in his on-line data base. Former countries and aggregated regions are cases in point where there was only data in the printed publications. Sometimes the tables only provided data for total GDP and population, so we had to do the divisions ourselves.

In many cases the printed tables in Maddison only provided data for a few benchmark years, e.g. for 1950, 1973, 1990 and 2001. However, the source that we liked to link this data with went back to somewhere between 1950 and 1973. In those cases we linked the data from

Maddison (2003) and the subsequent source at 1973, but included the earliest year available in the subsequent source. As usual we checked that the two sources were not totally inconsistent with each other for the overlapping years. In some cases they were indeed totally inconsistent, which mean that we did not use Maddison (2003).

4. Pre-1950 estimates

Many countries do not have any data at all for the earlier years. For other countries Maddison provides some data, but these should, to our understanding, mainly be considered as estimates of *growth rates*, i.e. they give information on how the incomes have developed from year to year for a given country.

4.1 Our assumptions for the data

Since we want to provide rough estimates for this period we need to make at least some assumptions. The assumptions we make are not based on any detailed literature review, so there are likely to be adjustments in future updates of the dataset. Our assumptions are:

- a) *A lower bound at \$275 for all periods.* We assume that no country fall below \$275 per capita for any sustained period. Pritchett (1997) assumes a lower bound of GDP per capita of \$250 in prices of 1985. He bases this figure on a number of arguments. First, if we look at the data we actually have, there is no country, at any point in time, that has had a lower income than that for any sustained period of time. Second, this corresponds to a level below the extreme poverty line. Third, a country where a majority had such an income would, in all likelihood, face a significant population decrease.³ We do not observe this to happen for any sustained period of time (although there are some historical examples of this).

\$250 in 1985 prices would be even higher if expressed in 2005 prices, which is what we use. Hence, we use \$275 as our lower bound. Still, our dataset contains observations that fall below this limit for a couple of years. Guinea Bissau and Lesotho do exactly this around 1950.⁴

- b) *A higher bound at \$2000 for the early 19th century.* We assume that in the early 19th century no country had yet reached the level of a middle income country, as we define it today. The poorest middle income country in 2005 in our dataset is Vietnam, which had an income of \$2142 in that year. Hence, we set the upper bound for the early 19th century at \$2000.
- c) *Industrialisation had started to lift the incomes of some countries in the early 19th century.* A number of countries had started to industrialise in the early 19th century. These countries displayed rising incomes and they started to surge ahead. Hence, they were at least somewhat richer than the rest of the world even at this point in time. United Kingdom and Holland were the leading early examples of this.

³ Pritchett bases this proposition on a demographic model and an estimation including incomes and infant mortality.

⁴ The unadjusted figures for Congo, Dem. Rep., are below 275\$ for the years 1999-2007. With our adjustments the figures are just above 275\$.

- d) *There was no sharp drop in incomes during the 19th century.* We assume that there was no general long-term collapse in incomes during the 19th century, for any region. Some specific countries, for which we have national growth data, display a drop in income for some periods.

However, the data and information we have reviewed so far gives us no reason to assume that long-term income collapse was a persistent or general pattern in the 19th century, for any regions. Hence, we will assume that there was no sharp drop in incomes in the countries for which we have no country data. There are exceptions in a couple of individual cases, which are either based on country-specific data or was something we had to assume in order to avoid unreasonable values in 1800.

- e) *China was not much poorer than Europe in the early 19th century...* We assume that China was roughly at the same income level as Western Europe at the eve of the Industrial Revolution. This is in contrast to the assumptions of Maddison. The reasons for our assumption will be discussed in more details below.
- f) *... but lost ground thereafter.* We assume a drop in incomes in China between 1820 and 1929. This will also be discussed in more detail below.

4.2 The data used for the 19th century

The assumptions above are, in principle, the only things we would like to display in the graph. Ideally, we would like to show a “cloud” of some sort that illustrates the range of likely values for this early period. However, the software does not permit that at this stage. Hence, we still need to assign an actual figure to each individual country. Furthermore, if we assign the same number to a large group of countries they will not be visible in the graph, since they would be in the exact same place in the graph. Such a picture would also, in all likelihood, communicate a too equal distribution of incomes.⁵

Hence, we need some principle for assigning actual numbers to each country, even if by necessary it is a very arbitrary principle. We use the growth rates of Maddison when these are available for individual countries or small groups of countries. However, in a few cases we adjust these growth rates since they imply overly absurd relative incomes in 1820. This will be discussed below.

Maddison also provides rough estimates for all regions in the world back to at least 1820. One way to extrapolate backwards would be to link the growth rates of a relevant region to the earliest observation of a country. However, the resulting relative positions in 1820 for many countries become absurd when we use this method.

Hence, rather than using the *growth rates* of the relevant regions we base our figures on the regional average *level* for the relevant region. However, we do two adjustments to these regional levels.

⁵ Even though, as Pritchett, 1997, has argued, the gaps between the richest and the poorest countries at this time were reasonably much smaller than today.

Firstly, we need to adjust for the fact that the data of Maddison is based on PPP data from the 1990 benchmark-round and that he express his data in 1990 prices. Hence, we use the predicted values from the “Gapminder model 2”, i.e. we do the same adjustments as we did with the Maddison data we used for 2005. These adjustments are likely to be immensely much smaller than the “uncertainty range”, but in the name of consistency we do them nevertheless.

Secondly, we multiply each observation with an arbitrary “spread-out factor” that adjusts the regional average up or down with an arbitrarily assumed percentage. This is done to avoid having a large number of bubbles in exactly the same place. These spread-out factors are mostly assigned in a totally arbitrary way. Sometimes we roughly base the spread-out factor on the relative position within the region at some later year for which we have data. This is just a convenient way to assign the spread-out factor. It does not mean that we believe that a fixed ranking of countries is a reasonable assumption.

Some other times we base the spread-out factor on some quick reading of the historical or geographical condition of the country, e.g. a land-locked area with a desert climate that, to our understanding, lacked any centralised political structure, was assigned a lower value. Again, this was just a convenient way of assigning values. A more careful literature review in the future might give us a somewhat more firm standing in coming updates. We document all these considerations in the Excel sheet.

Furthermore, the assumptions we made above, in combination with the fact that the position of *some* countries are based on national growth rates, implies specific constraints for the spread-out factors. These constraints are discussed below.

We use adjusted regional averages for 1820, 1913 and 1950, depending on the data availability for the individual countries.⁶

In a number of cases the adjusted average for 1913 implied a negative growth between 1913 and 1950 (the first available national observation). Since we had no reason to assume that there actually was negative growth,⁷ we instead choose to not include that imputed value. Instead we used an interpolated value, assuming exponential growth between 1820 and 1950. We also used interpolated values for 1913 in the few cases where the implied growth rates between 1820 and 1950 actually were negative (see Section 4.3). For some oil rich countries we also added an additional observation, for the year of the first oil-exploration, based on regional averages. That will be discussed in more detail below.

To summarise, what we want to end up with is $Y_{t,i}$, the adjusted value for country i , in the year t , and it is calculated in the following way:

$$Y_{t,i} = S_i * e^{\alpha} * X_{t,r}^{\beta}$$

$X_{t,r}$ is the regional average for region r , in year t , taken from Maddison. α and β are the estimated coefficients in “Gapminder model 2”. S_i is the “spread-out factor” for country i .

⁶ 1950 was only used for the republics of the former USSR.

⁷ Since these values for 1913 only were based on regional averages, which in themselves were based on guesstimates, the growth rate implied from these values carries no real weight.

4.3 Specific constraints for the “spread-out factor”

Maddison provide growth for South Africa back to 1820.⁸ The unadjusted figures for 1820 by Maddison put South Africa on approximately the same level as his estimate for the African average for that year. If we turn to 2005, we find that the ICP data for 2005 gives South Africa a higher relative position than in Maddison.

Since we link the growth rate of Maddison to the ICP data in 2005, it means that our South African figure for 1820 is much higher than the African average for 1820 of Maddison. If we assume that South Africa was at the wealthy end of Sub-Saharan distribution, we must assign spread-out factors that put the “richest” Sub-Saharan countries at the approximate level of South-Africa.

An alternative to this approach is to assign our own values to South-Africa, since Maddison’s figure for 1820 for South Africa is also just a guesstimate. However, we stick to the growth rates of Maddison.

1950 is the first available year in Maddison for many countries. Several Sub-Saharan countries display figures for 1950 that are close to, or even slightly below, our assumed threshold level of \$275.

In particular, the poorest country in 1950, Guinea Bissau, had a GDP per capita of only \$269 according to our adjusted data.⁹ The unadjusted data of Maddison also display a very low level. We link his data to the ICP figures of 2005, which puts Guinea Bissau at an even lower level than Maddison for this year. Hence, our figure for 1950 is even lower than the one Maddison provides.

The low level of Guinea Bissau (and other similarly poor countries) implies a lower range when we assign our figures for 1800. Since we do not want to assume any systematic decline in incomes during 1800-1950, we would in principle have to assign values of maximum \$269 in 1800 to the poorest countries. That would set the lower bound of our range of possible values in 1800, at least for Sub-Saharan Africa.

However, that level would imply that at least Guinea Bissau had an income below \$275 for over a century, which we also believe is unreasonable. Hence, we compromise a bit with the assumption for long term growth, and allow the growth rate to be somewhat negative in a few cases.

5. Adjustments made to some national growth series

The combination of the growth rates of Maddison and the levels of ICP sometimes puts countries way off the reasonable range in 1800, even in relation to our extremely weak assumptions for this period. Furthermore, in the specific case of China we do indeed make some quite specific assumptions.

At the same time, just changing the figures for 1800 and 1820 might create unreasonable “jumps” in the time series of the individual countries. One quick and extremely rough way of handling this is to

⁸ He also provides growth data from 1820 for Ghana. However, we choose not to use them because his data for the early 1800s seems to be guesstimates and the relative position would, in any case, be different since we are using the ICP figures for 2005.

⁹ The only other country to ever get so low is the unadjusted figures for Congo, Dem. Rep, around 2001.

still use the implied growth rates of Maddison, but adjust it proportionally so that if we extrapolate backwards from 2005 we end up in the value we have assumed for 1820. This could be thought of as if we pivot the whole time series by moving 1820 upwards or downwards, while fixing the series in 2005. In at least one case we only “pivot” the time series from 1950 and onwards. In these cases the time series for earlier years were linked to the values of 1950.

What we do, to be more precise, is the following: $X_{t,i}$ is the original value for year t for country i and $Y_{t,i}$ is the new value for year t , for country i , that we want to calculate. First we decide on what we want the new value for 1820 to be (i.e. the value for $Y_{1820,i}$). Then we calculate an adjustment factor: $K_i = \frac{Y_{1820,i}}{X_{1820,i}}$. We then use this adjustment factor to calculate the new values for all subsequent years:

$$Y_{t,i} = X_{t,i} * K_i * \left(\frac{1}{K_i}\right)^{\left(\frac{t-1820}{2005-1820}\right)}$$

This means that $Y_{2005,i} = X_{2005,i}$, as we wanted. When we pivoted the series from 1950 and onwards we just replaced 1820 with 1950 in the formula above. “Pivoted” data are noted in the sources in the Excel file as “National data, adjusted”.

The observation for 1800 was normally interpolated or extrapolated in a separate step, but any methodological problem related to that adjustment is certainly smaller than the overall uncertainty range for these observations, which are bound to be enormous.

Note that the adjustment of the growth is not based on any assessment of the growth rate itself, but only comes from the need to adjust the levels in the 19th century. Given the graphical format of Gapminder we judged it to be more important to assess the relative positions of countries at a given time, as well as avoiding too inconsistent “jumps” from one year to another, rather than to preserve the precise level of average growth, which is hard to see in the graph.

In future updates we hope to find a more reasonable way of doing these adjustments, without doing unfounded changes to the growth rates.

We do these changes for a small number of countries. Which these are, what the adjustment factor K has been, and the principle behind the individual adjustments can be found in the Excel file (in the column “1820 ad-hoc adjustments”). However, two countries, Myanmar and China, call for some additional comments.

5.1 Myanmar

The new PPP data for Myanmar pose especially large problems. The new ICP GDP per capita for Myanmar is only about a third of what it was in Maddison. If we combine that with the growth rates of Maddison we get extremely low figures for earlier years, below what we consider the subsistence level. The levels get especially low around 1950, when the country makes a sharp dip. However, the GDP per capita is also very low at other points in time, including all observations in the 19th century.

If we “pivot” all the data for 1820 and onwards we would have obtained unrealistically high values for 1931 and earlier, since the income levels drop sharply between 1931 and 1950. Hence, we adjust the 1950 figure upwards, and “tilt” the subsequent time series. The unadjusted growth rates for the period before 1950 were linked to the adjusted value of 1950.

5.2 China in the 19th century

The relative position of China and Western Europe on the eve of the Industrial Revolution has been a hotly debated issue. A traditional assumption has been that the living standards in Europe were higher than in China long before Europe became industrialised. The assumptions of Maddison have been in this tradition, i.e. he assumes the following (Zanden, 2005):

- In the year 1000 China had a higher GDP per capita than Europe. Even though there are obviously no direct data for GDP per capita, the general idea that China was more advanced in the year 1000 is widely accepted.
- In the early 19th century the Chinese GDP per capita was about 1/3 of the UK level.
- The Chinese incomes did not display any significant decline in 1000-1800, so that could not explain the reversal of positions.
- Rather, the reversal is explained by a non-negligible positive growth in Western Europe from 1000-1800. This assumption seems mainly to be based on the previous assumptions, i.e. the growth rate is assumed to be what it has to be to explain the reversal of positions.

However, criticism has been raised against the traditional assumption of a much lower income in China in 1800, as compared to Western Europe. Zanden (2005) has estimated the growth in Western Europe from 1000-1800 based on models using wage data and the share of the labour force in agriculture. According to the estimates, the growth rates in Western Europe were much smaller than what Maddison assumes: positive in dynamic regions such as England and the Netherlands, although slower than what Maddison assumes, while stagnant or even negative in more peripheral countries such as Sweden.

Furthermore, Maddison seems to base his 1820 figure for China on the growth rates 1820-1990 and the PPP comparisons of 1990. However, the Chinese growth rates before 1949 are not so reliable (Zanden). Hence, the basis for the relative poverty of China in 1800, in term of hard figures, is not that strong, according to this line of argument.

Zanden has also made a direct comparison between the Netherlands and Java (which could be a reasonable proxy for Indonesia) during the early 19th century. The result is quite in line with the results of Maddison. Furthermore, Maddison assumes that China had lower incomes than Indonesia at this time. However, Zanden believes, for a number of reasons, that China was more developed than Indonesia at this time. Hence, it would be reasonable to assume a higher income level in China for this period.

In addition, Zanden used a model based on wage data and the share of labour force in agriculture to estimate the incomes in China relative to the UK in 1800. According to the estimate, the Chinese GDP per capita in 1750-1800 was about 53% of the UK level, i.e. quite a bit higher than the 1/3 assumed by Maddison.

Allen, Bengtsson & Dribe (2005) also refer to similar conclusions. Using data on prices and wages they argue that the living standards in China and Western Europe were relatively similar in the 18th century (while there were significant differences within Europe and within China). They also suggest that there could have been a deterioration in living standards in China from the 19th century until the beginning of the 20th century.

If we were to use the Chinese growth rates of Maddison, and combine them with the 2005 level of China of ICP, then we would get an *even lower* relative position of China in 1800 than what Maddison assumes. This is an outcome of the fact that the ICP display a lower level for China in 2005 than what Maddison implies.

To compensate for that, we adjust China in 1820 upward so that it gets the same level as the lower half of the European countries in the same year. This means that we assume that the Chinese GDP per capita was about 43% of the UK level in 1820. This is lower than the 53% of the Zanden estimate for 1750-1800, but it is higher than the assumption of Maddison, and even higher than what we had gotten if we would have used the growth rates of Maddison unadjusted.

As a consequence of our “tilting” of the series (i.e. downward adjustment of the growth rate) we get a negative growth for the period 1820-1929. However, this is in line with the pattern suggested by Allen, Bengtsson & Dribe.

5.3 Some further notes on Chinese GDP per capita

Also the modern growth and income levels of China have also been debated. Maddison assumes, for a number of reasons, a slower growth rate for the post-war period than the World Bank does.¹⁰

Furthermore, the ICP data implies a much lower relative position of China in 2005, than in previous data sets, including the data of Maddison. However, it is likely that the PPP rate of China is based on a sample that under-represents rural China, as argued by Chen & Ravallion (2008). This implies that the relative income level of China is underestimated in the ICP data, although it would still be lower than implied by the older data sets.

Maddison (2008 II) has also argued, for a number of reasons, that the ICP figure for China is too low. One argument is simply that we would get an unreasonable low figure for China in 1950 if we combine the ICP figure for 2005 with the growth rate for 1950 to 2005. Table 5 below charts the implied GDP per capita in 1950 if we were to combine the income levels for 2005 and growth rates for 1950-2005 from the different sources.

¹⁰ The Penn World Table 6.2 display a growth rate that is somewhere between Maddison and the World Bank. (Deaton & Heston, 2008)

	Growth 1950-2005 in Maddison (12.5 fold):	Growth 1950- 2005 in WB: (21.0 fold):	Growth 1950- 2005 in Gapminder (9.8 fold):
Value in 2005 in Maddison (I\$ 5578):	446	266	570
Value in 2005 in WB & Gapminder (I\$ 4091):	327	195	418

Table 5: Implied values for GDP per capita, by PPP, in China in 1950, under various combinations of assumptions.

Combining the 2005 level from ICP, with the growth rate of the World Bank, we get a GDP per capita for 1950 of only 195, which is below our lower bound level, and much below the lower bound level suggested by Pritchett. If we combine the ICP data for 2005 with the lower growth rate of Maddison we get a somewhat higher figure, but still a very low one, which still is highly debatable. If we would accept the lower growth rate of Maddison and inflate the 2005 figure from ICP upward a bit, to adjust for the suggested underestimation for 2005, then we would get a number somewhere between 327 and 446.

We choose to use the ICP figure for 2005 unadjusted and combine that with the growth rates of Maddison, which we adjusted downward a bit further. Our post-war growth rate hence becomes lower than the growth rates of all other sources. The reason for the downward adjustment of the growth rate was based on our assumption of a higher level in 1800 (compared to Maddison), not on any assessment of the growth rate itself. Given the graphical format of Gapminder we judged it to be more important to assess the relative position of countries at a given time, as well as avoiding excessively inconsistent “jumps” from one year to another, rather than the precise level of average growth, which is hard to see in the graph.

5.4 Congo, Dem. Rep.

It is believed that the ICP figure for Congo, Dem. Rep., for 2005 is underestimated.¹¹ The level of 264 international dollars (PPP) is below the level we assumed was the lowest possible (i.e. 275\$, which is below the 1\$ per day level). Furthermore, the official growth rates during the 1990s imply that the country had incomes below this very low level during the seven years 2000-2006. In figure 6 below the GDP per capita of Congo, Dem. Rep. is displayed together with the second poorest country in the world, Liberia. The pink line is the assumed lowest level possible (275\$).

Such figures seem quite implausible, since that implies that even if the incomes were distributed perfectly equal within the country, everyone in the country were subject to famine conditions for a decade. The conditions in the country are without doubts horrendous, but incomes on that level would, in principle, starve the entire population. That has not been observed.

¹¹ Jerven (in press).

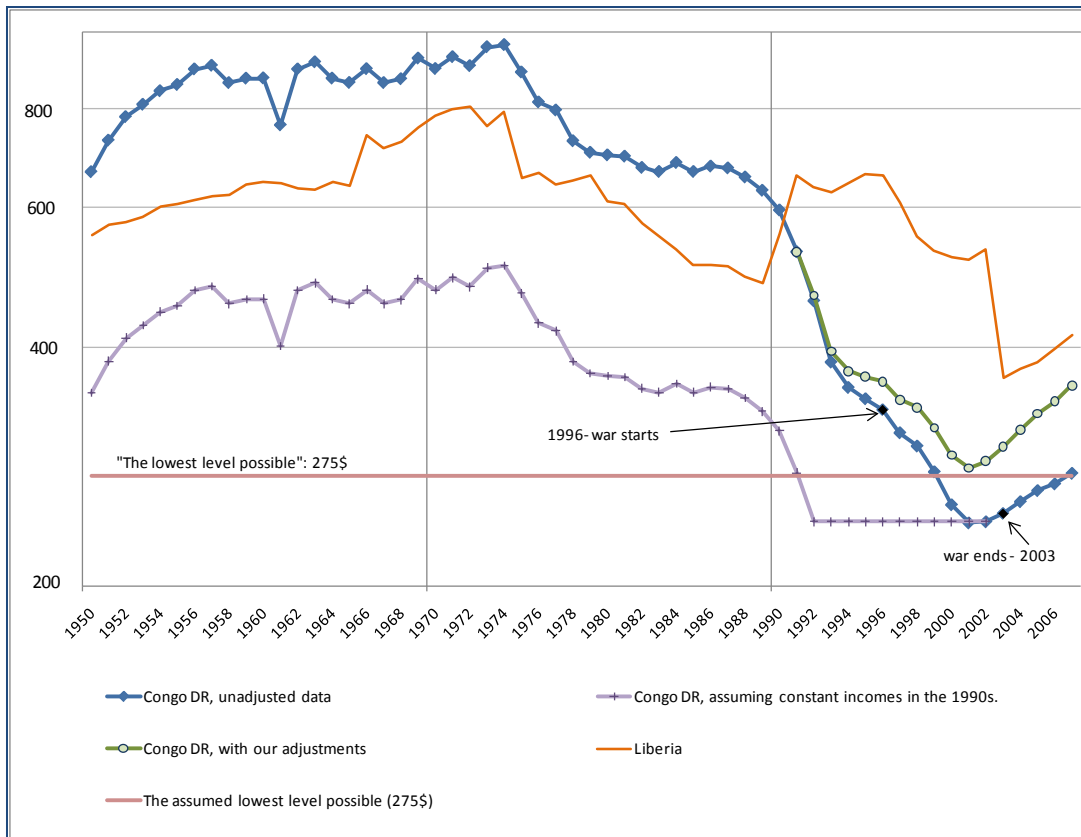


Figure 6. GDP per capita for Congo, Dem. Rep., and Liberia (the second poorest country)

In addition, a recent paper (Henderson et al, 2009) calls for a reinterpretation of the growth rates of Congo DR in the 1990s. They estimate the growth rates in the 1990s from changes in night lights as measured from space. Their results, taken at face value, imply that the incomes were stagnant 1992-2002, rather than the sharp fall implied by the official data. Since the growth rates we use are linked to the level of 2005, this would imply that the country were subject to catastrophic famine conditions for almost two decades. To illustrate this we added the GDP per capita for Congo in figure 6, as it would have been if the incomes were constant in the 1990s.

The assessment in Henderson et al, though, is that the truth is somewhere between what the estimates based on night lights say and the official growth rates. That would mean that there was a fall in the 90s, but that this was less dramatic than implied by the official data. This drop in incomes is also plausible given the devastation from the war 1996-2003.

Considering the problems with accepting the low levels and the evidence that the collapse in income might have been smaller, we decided to adjust the ICP figure for 2005 upwards with 25%. We also decided to adjust the growth rate from 1991 to 2007, rather than to adjust the whole series upwards with 25%. What we did was to keep the values until 1991 as they would have been with the unadjusted ICP figure for 2005 and the national growth rates. From 1991 to 2007 we used the same adjustment formula as described in the beginning of this section.

The function used is:

$$Y_t = X_t * 1,25 * \left(\frac{1}{1,25}\right)^{\left(\frac{t-2005}{1991-2005}\right)}$$

X_t is the original value for year t and Y_t is the new value for year t . The new adjusted data is also included in figure 6.

After 2007 we used unadjusted growth rates. The choice of 2007 for the ending of the growth adjustments was an arbitrary choice. However, it has little effect on the relative position of the country.

6. Adjustments of oil rich economies

There are a few high income economies that are largely based on their oil incomes. Hence, their growth pattern is highly linked to this one income source. It is accordingly reasonable to assume that the income level before their oil discovery was significantly lower.

However, in many instances our data does not go back to the year of the oil discovery. If we use the adjusted regional average for 1820 and 1913 (as we do for several other countries), then the graph software would automatically interpolate linearly between 1913 and the first year with data. The graph would then display a very sharp surge in growth after 1913, which is very unrealistic.

To avoid this we add an observation, based on the regional average, for the year of the first oil exploitation. Since Maddison does not provide regional averages for these years we use the adjusted regional average for 1913, extrapolated with the regional growth rate from 1820 to 1913. With “adjusted regional average” we mean, as before, the regional average adjusted with Gapminder model 2, multiplied with the “spread out factor” for the country.

The “year of oil discovery” was based on a quick reading of on-line country guides e.g. the Country Guides of the Library of Congress (2008). A more careful reading of the history of these countries will probably lead to adjustments in future updates of this dataset.

There is also a general problem with oil rich countries. The different sources are often very inconsistent with each other for these countries. It is easy to see a number for reasons for this. Oil-incomes could be accounted for in very different ways. Furthermore, most of these countries have a very large foreign work force. This foreign work force might, or might not, be included in the total population that is used to calculate the per capita production. All this taken together means that oil-rich countries should be treated with particular care.

7. Data for 1800

Maddison only provided data for 1820 and occasionally for years before 1800. For completeness we wanted some rough estimates for 1800. We interpolated, with constant growth, between 1820 and

earlier years, for countries or regions where Maddison provided data for years before 1800. When we interpolated for a whole region we used the same spread-out factor as for 1820 to get the country figure. Note that we do not normally include the data for the years before 1800, even though there are figures in Maddison. Hence 1800 might be interpolated even though it is the first year in our dataset.

For some countries 1820 is the first available year in Maddison. In these cases we simply extrapolate from 1820 by using the same values, i.e. we assume zero growth from 1800-1820.

8. Data quality – possible maximum quality

Our long term ambition is to assign a quality rating to each observation. These can be displayed in the graph with colours. As for other indicators compiled by Gapminder, we consider 5 quality grades: Very Good, Good, Fair, Poor and Very Poor.

However, we have not yet been able to compile the necessary meta-data for the actual GDP figures, only for other aspects of the data, such as the PPP rate. Even if the PPP rate is of good quality, and there are no other observed problems, we cannot exclude that the underlying GDP data is of poor quality, something that implies a low overall quality.

The meta-data we do have is as follows:

- a) Information on the price comparisons (the PPP rates) for 2005 (the benchmark year)
- b) Whether we have done any rough extrapolations or other adjustments ourselves, something that would decrease the quality.
- c) Additional problems can be noted in some cases (e.g. large inconsistencies between sources)
- d) We can also assume that the quality deteriorates the farther away we get from the benchmark year.

The problems above, taken together, imply a “Very Poor” data quality for many observations. For the other observations we do not know the data quality. We accordingly calculate a *maximum possible* quality grade, rather than an actual quality grade. In the graph we display only two values: “Very Poor” (given the value 5 in the graph) and “data quality not known yet” (given the value 0 in the graph).

We assign the “maximum quality grades” in the following way. First, we start with the data quality implied by the PPP rate, as outlined earlier:

- Good: Official ICP data, as long as there are no footnotes that indicate extraordinarily bad (or good) data quality.
- Fair: Official ICP data, with caveats about the data quality
- Fair: Official ICP estimates (ICP model 6 and 7)

- Very Poor: Gapminder estimates based on the World Bank assessment of what income group the country belongs to
- Poor: All other Gapminder estimates

Second, for observations other than for 2005, we decrease the grading of the country according to the following:

- Observations before 1975: we decrease the grading one step.
- Observations before 1950: we decrease the grading two steps (i.e. one additional step from 1975).
- Observations before 1900: we decrease the grading three steps (i.e. one additional step from 1950)
- If we use the regional growth rate or the growth rate of a neighbouring country: we decrease the grading one step
- If there are some other problems, e.g. inconsistencies between sources, or the country is dependent on oil incomes: we decrease the quality grading one step or more.

Finally, if we used the adjusted regional average, arbitrary guesses or something similar, the quality grading is always Very Poor.

However, we should keep in mind that it is not reasonable to simply count the number of quality problems. Several small sources of errors, that are uncorrelated with each other, are not the same as one big source of error. Accordingly, the rules above should only be considered as a rough set of rules of thumb, not a formula that could be applied mechanically. We hope to refine our approach in future updates of the dataset.

The grade “Very Poor” is, in a way, not the lowest possible quality grading. There are after all observations that we have not included at all. The information for these observations could accordingly be considered even poorer than for the ones graded as “Very Poor”. Hence, the observations graded as Very Poor could still be considered to contain some kind of information. One way of regarding the message of such an observation is that we have at least some reason to assign the country to one of the three income groups: low, middle or high income countries.

Our present quality gradings can be viewed in the Gapminder World. The data quality indicators can be found in the menu under “for advanced users” -> “data quality”. The best way to view it is to select the data quality for the colours of the bubbles. An example can be viewed at:

<http://tinyurl.com/dxzfmq>

In the graph in the link above the red bubbles are based on observations with “Very Poor” data quality for GDP per capita, while blue bubbles have an unknown data quality.

10. The Excel file

Detailed meta-data is provided in the Excel file. Each row in the excel-file contains data for a specific country. Each column contains some specific type of meta-data, or actual data for a specific year.

Table 6 below charts the included columns:

Column	Title	Contents
A	Area	The name of the country or area

Two miscellaneous columns for 1800 & 1820.

C	1800 - type of estimate	This column describe how we estimated the figure for 1800 (e.g. whether interpolated or extrapolated).
D	1820 ad-hoc adjustment: subsequent levels all the years up to 2005 were also adjusted proportionally.	In some cases we have adjusted the data of Maddison for 1820. We also adjusted the subsequent growth so that the figure for 2005 remains unchanged. This column indicates if such an adjustment was done and how it was done.

Regional averages. Some observations were based on an adjusted figure based on regional averages provided by Maddison. The following columns describe how we did that.

F	Years for which we use the average level of the Maddison country groups	A year.
G	Maddison country group - For some countries, for some years, we used the average for the relevant Maddison aggregated country group. This data was then adjusted with Gapminder model 2 and then multiplied with an arbitrary "spread out factor".	A region or a country group, as they are called in Maddison.
H	"Spread out factor" - we multiply this number to the adjusted average for the relevant Maddison country group.	A number. The number corresponds to S_i in the equation in section 4.2.
I	"Spread out factor" explanation - what ad hoc principle was used to assign the spread out factor	A short text describing the principles we used to assign a value to S_i .
J	"Spread out factor" - explanation 2	A short text giving more details of the principles we used to assign a value to S_i , such as the use of "sub-groups".

Oil-rich countries		
L	Oil-rich countries: year of oil-discovery & what data we used for that year	A year and a short text.

Growth rates: sources and other notes. These columns contain information for the growth rates. The time series for a country might be based on several sources linked to each other. The columns are organised in "blocks" that each contains 4 columns. Each "block" contains information for one specific individual source. The columns in each block are:

N to AH	Year	The first year covered by the source in question. It is also the last year for the preceding source, if any, and the year used for linking the two sources.
N to AH	Type of data (national growth or other)	This could be: (1) National growth; (2) National growth, adjusted; (3) Regional growth; (4) Neighbouring country growth; (5) Interpolation; (6) Arbitrary guess; (7) No data
N to AH	Basis for estimate (if other than national growth)	Here you find (1) the name of the region (if we used "regional growth"); or (2) the name of the neighbouring country (if we used "neighbouring country growth"); or (3) the years used for interpolation (if we interpolated).
N to AH	Source	Here you find a short name for the specific source used. See Table 7 below for details on each source.

PPP rates (2005). Here you find information for the benchmark year 2005.

AJ	PPP benchmark year (2005) - data source	Here you find a short-name of the source for the cross-country data for the benchmark year 2005. See Table 7 below for details on each source. We adjusted the data for sources other than the ICP. The adjustments are specified in the column "2005 adjustment model".
AK	PPP benchmark year (2005) - data footnotes	Here you find additional information about the source for 2005.
AL	2005 adjustment model. (We adjusted the data when we used other sources than ICP. We did this to make the data more comparable.)	In this column we specify which model, if any, we used to adjust the data for 2005.

AM	2005: PPP quality	Here you find our assessment of the quality of the PPP rate. These rates are mainly based on the source for the PPP rate. The grades only refer to the quality of the price-comparisons. They do not say anything about the quality of the GDP per capita itself. The grades are: Very Good; Good; Fair; Poor; Very Poor.
----	-------------------	---

Other footnotes		
AO	Other footnotes	Here you find other information about the data, e.g. some specific problems, some minor corrections we did, or explanations why we chose a specific source. The notes sometimes refer to all data for the country, and sometimes they only refer to a specific year.

Data quality - The quality grades are based on the combination of a number of factors, e.g. the quality of the PPP rate; the quality of the raw data for GDP per capita growth (presently unknown); the time-distance from the benchmark year 2005; whether the observation is an estimate by Gapminder; other problems noted with the data.		
AQ	Data quality - "very poor" until this year	All observations before this year are considered to be very poor.
AR	Data quality - "data quality unknown" from this year	All observations, in principle, after this year are considered to have unknown data quality. In some cases there are observations after 2005 that are considered very poor.
AS	Data quality - "very poor" from this year	All data after this year are considered to be very poor
AT	Data quality - summary explanation for the grading	Here you find an abbreviated summary of the chain of quality factors that motivated a "very poor" grading.

Hidden working notes + Data		
AU-BA:	(Hidden)	These columns are hidden and contain some internal working notes
BB-JR:	(Years)	These columns contain the actual data.

Table 6. Explanations for the columns in the "Excel file"

11. Explanations of short names used for the sources

The Excel file use short-names for the sources, both the sources for the cross-sectional data for 2005, as well as the sources for the real growth rate of each country. Table 7 below explains the details for each short-name:

Short name (used in the Excel file)	Source (see the reference list)	Notes
Barro & Ursúa	Barro & Ursúa (2008)	On-line table "GDP pc"
CIA	CIA (2006), (2007) or (2008).	CIA World Factbook, "Field Listing" of "GDP - per capita (PPP)". The footnotes in the Excel file indicate which factbook we used for the individual observation.
Eurostat	Eurostat (2008)	From the homepage of Eurostat. Regional GDP (PPS per inhabitant), by NUTS 2 regions
FEPS (2007)	FEPS (2007)	The data was "extracted" from the figure "Diagram 2: GDP per capita at 2006 values"
ICP	International Comparison Program (2008)	"Gross domestic product per capita – PPP" from the "Summary Table".
ICP - adjusted	International Comparison Program (2008)	"Gross domestic product per capita – PPP" from the "Summary Table". Adjusted by gapminder.
ICP estimate - model 6.	International Comparison Program (2008)	"Estimated GDP per capita (PPP)" from Table 8 in the "technical notes of the 2005 ICP" (page 165).
ICP estimate - model 7.	International Comparison Program (2008)	"Estimated GDP per capita (PPP)" from Table 8 in the "technical notes of the 2005 ICP" (page 165).
IMF (2009)	IMF (2009)	Gross domestic product per capita, constant prices
IMF (2010)	IMF (2010)	Gross domestic product per capita, constant prices
INSEE (2008), adjusted	INSEE (2008)	"PIB par habitant (à prix courant)" in page 155, deflated with the inflation rate from page 159.
Isle of Man Treasury (2008), adjusted	Isle of Man Treasury (2008)	"Isle of Man GDP per capita" (in current prices) taken from Table 7. An implicit GDP deflator was calculated from Table 2.
Krantz & Schön	Krantz & Schön (2007)	GDP per capita in market prices, taken from Table IV.
Maddison (2003)	Maddison (2003)	Sometimes we had to take total GDP and divide it by the total

		population.
Maddison on-line	Maddison (2008 I)	GDP per capita from the “horizontal file” at his homepage
PWT	Heston; Summers; Aten (2006)	Penn World Table 6.2, real GDP per capita, in Laspeyres Constant Prices (RGDPL).
St. Helena Development Agency (2008)	St. Helena Development Agency (2008)	“Per Capita GDP (constant 2002 prices) in £” from Table 1.
UNSTAT I	UNSTAT (2008 I)	GNI per capita by exchange rate in prices of 2005. In the list on the homepage the indicator is called "GNI, per capita GNI - US Dollars".
UNSTAT II - adjusted	UNSTAT (2008 II)	“Per capita GDP in current US dollars”, deflated with the “implicit price deflators in national currency and US dollars”.
UNSTAT II (2009) -adjusted	UNSTAT (2009 II)	“Per capita GDP in current US dollars”, deflated with the “implicit price deflators in national currency and US dollars”.
WDI (lcu)	World Development Indicators 2008 on-line	WDI GDP per capita (constant LCU)
WDI (ppp)	World Development Indicators 2008 on-line	GDP per capita, PPP (constant 2005 international \$)
World Bank (rough estimate)	World Bank (2005)	We used the data based on the Atlas methodology

Table 7: Explanations for the short names used in the Excel file

References

Allen, Robert; Bengtsson, Tommy; Dribe, Martin (2005); "Introduction" in "Living Standards in the Past – New perspectives on well-being in Asia and Europe"; edited by Allen, Robert; Bengtsson, Tommy; Dribe, Martin

Barro, Robert & Ursúa, José (2008); "Macroeconomic crises since 1870", downloaded April 2009 from:

www.economics.harvard.edu/faculty/barro/files/MacroeconomicCrisesSince1870_Brookings_08_05_07.pdf The data in the paper was downloaded from:

www.economics.harvard.edu/faculty/barro/data_sets_barro

Chen, Shaohua & Ravallion, Martin (2008), "China is poorer than we thought , but no less successful in the fight against poverty", The World Bank Policy Research Paper 4621

CIA (2008), "CIA World factbook 2008", Downloaded 2008-11-04 from:

<https://www.cia.gov/library/publications/the-world-factbook/fields/2004.html> The page was last updated on 23 October 2008.

CIA (2007), "CIA World factbook 2008", Downloaded 2008-11-04 from:

<http://www.umsl.edu/services/govdocs/wofact2007/fields/2004.html> The page was last updated on 15 May, 2007.

CIA (2006), "CIA World factbook 2006", Downloaded 2008-11-04 from:

<http://www.umsl.edu/services/govdocs/wofact2006/fields/2004.html> The page was last updated on 13 June, 2006.

Deaton, Angus & Heston, Alan (2008), "Understanding PPPs and PPP-based national accounts", NBER Working Paper 14499

Eurostat (2008), "Regional gross domestic product (PPS per inhabitant), by NUTS 2 regions ", downloaded 2008-11-20 from:

http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1996,45323734&_dad=portal&_schema=PORTAL&screen=welcomeref&open=/t_reg/t_reg_eco&language=en&product=REF_TB_regional&root=REF_TB_regional&scrollto=0

FEPS (Fiscal and Economic Policy Steering Group, States of Guernsey) (2007), "Strategic Economic Plan produced by the Fiscal and Economic Policy Steering Group for inclusion in the Government Business Plan (Billet XVIII, 2007)". Downloaded November 2008 from <http://www.gov.gg/ccm/navigation/government/government-business-plan/>

Henderson, Storeygard and Weil (2009), "Measuring Economic Growth from Outer space", Working Paper, July 21

Heston, Alan; Summers, Robert and Aten, Bettina (2006), "Penn World Table Version 6.2", Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006. Data downloaded 2008-10-21 from http://pwt.econ.upenn.edu/php_site/pwt62/pwt62_form.php

INSEE (Institut national de la statistique et des études économiques) (2008), "Tableau Économique de La Réunion". All pdf-files marked with "INSEE-RÉUNION - TER 2008-2009"; downloaded 08-11-12 from:
www.insee.fr/fr/insee_regions/reunion/themes/dossiers/ter/ter2008_resultats_economiques.pdf
and http://www.insee.fr/fr/insee_regions/reunion/themes/dossiers/ter/ter2008_evolution_prix.pdf
Available at this homepage:
http://www.insee.fr/fr/regions/reunion/default.asp?page=themes/dossiers/ter/ter_economie.htm

International Monetary Fund (2009), "World Economic Outlook database. October 2009"; downloaded 2009-12-16 from:
<http://www.imf.org/external/pubs/ft/weo/2009/02/weodata/download.aspx>

International Monetary Fund (2010), "World Economic Outlook database. April 2010"; accessed 2010-04-21 from: <http://www.imf.org/external/pubs/ft/weo/2010/01/weodata/download.aspx>

International Comparison Program (2008), "Global Purchasing Power Parities and Real Expenditures. 2005 International Comparison Program", <http://siteresources.worldbank.org/ICPINT/Resources/icp-final.pdf>

Isle of Man Treasury (2008), "Isle of Man National Income Accounts 2006/07". Downloaded 08-11-18 from: www.gov.im/lib/docs/treasury/economic/income/nationalincome200607.pdf Available at: www.gov.im/treasury/economic/data/income.xml

Jerven, Morten (in press), "The African growth evidence", African Affairs

Krantz, Olle & Schön, Lennart (2007), "Swedish Historical National Account 1800-2000", Lund Studies in Economic History 41. The data was downloaded 2009-04-17 from <http://www.ehl.lu.se/database/LU-MADD/National%20Accounts/default.htm>

Library of Congress (2008), "Library of Congress country studies". Downloaded November 2008 from: <http://memory.loc.gov/frd/cs/>

Maddison, Angus (2003), "The World Economy – Historical Statistics"

Maddison, Angus (2008 I), "Historical Statistics for the World Economy: 1-2006 AD", Downloaded 2008-10-21 from www.ggdc.net/maddison/ Last update according to homepage: October 2008

Maddison, Angus (2008 II), "Explanatory Background Note on Historical Statistics", http://www.ggdc.net/maddison/Historical_Statistics/BackgroundHistoricalStatistics_09-2008.pdf
Downloaded: 2008-10-21 from <http://www.ggdc.net/maddison/>

Pritchett, Lant (1997), "Divergence, Big Time", Journal of Economic Perspectives; Vol. 11; No. 3

St. Helena Development Agency (2008) "St Helena Sustainable Development Plan 2007 – 2010". Downloaded 2008-11-04 from www.shda.co.sh/pages/st_helena_facts_and_figures.html.

UNSTAT (2008 I), "Basic data selection", data downloaded 2008-11-02 from: <http://unstats.un.org/unsd/snaama/selectionbasicFast.asp>

UNSTAT (2008 II), "Downloads", data downloaded 2008-11-10 from: <http://unstats.un.org/unsd/snaama/dnllist.asp>

UNSTAT (2009 II), "Downloads", data downloaded 2009-12-17 from: <http://unstats.un.org/unsd/snaama/dnllist.asp>

Zanden, Jan Luiten van (2005?), "Simulating early modern economic growth", Working Paper, downloaded 2009-04-17 from: <http://www.iisg.nl/research/jvz-simulating.pdf>

World Bank (2005), "GNI per capita 2005, Atlas method and PPP", summary note from the WDI, downloaded November 2008 from:

http://siteresources.worldbank.org/ICPINT/Resources/Atlas_2005.pdf

World Bank (2008), "Estimations of PPPs for non-benchmark economies for the 2005 ICP round" Downloaded November 2008 from: <http://siteresources.worldbank.org/ICPINT/Resources/non-benchmark.pdf>

World Development Indicators 2008 on-line (2008). Downloaded 2008-10-23 from: <http://publications.worldbank.org/WDI/>