

Documentation for

GDP Per Capita by Purchasing Power Parities

for sub-national units

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

Gapminder Documentation 007

Version 1

Uploaded: 2009-05-28

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



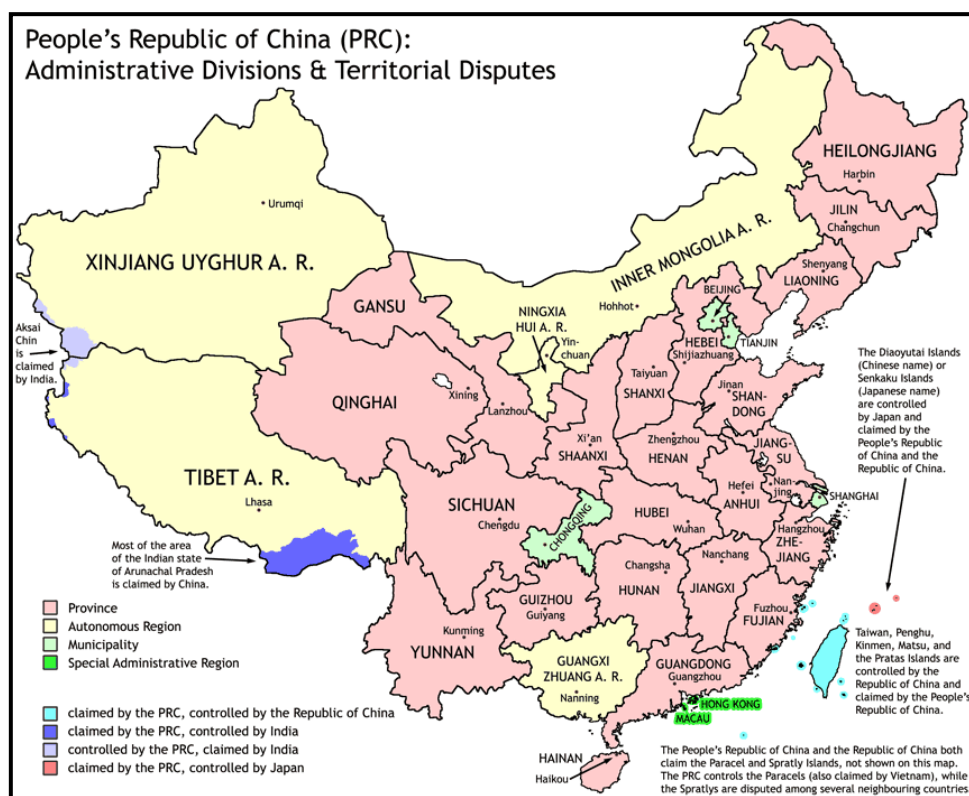
Background

This is the documentation for the Gapminder compilation of sub-national GDP per capita by Purchasing Power Parity (PPP) of China, India and the U.S.; the data is used in the interactive graph *Gapminder 4 Big* at: http://graphs.gapminder.org/world/four_big.php

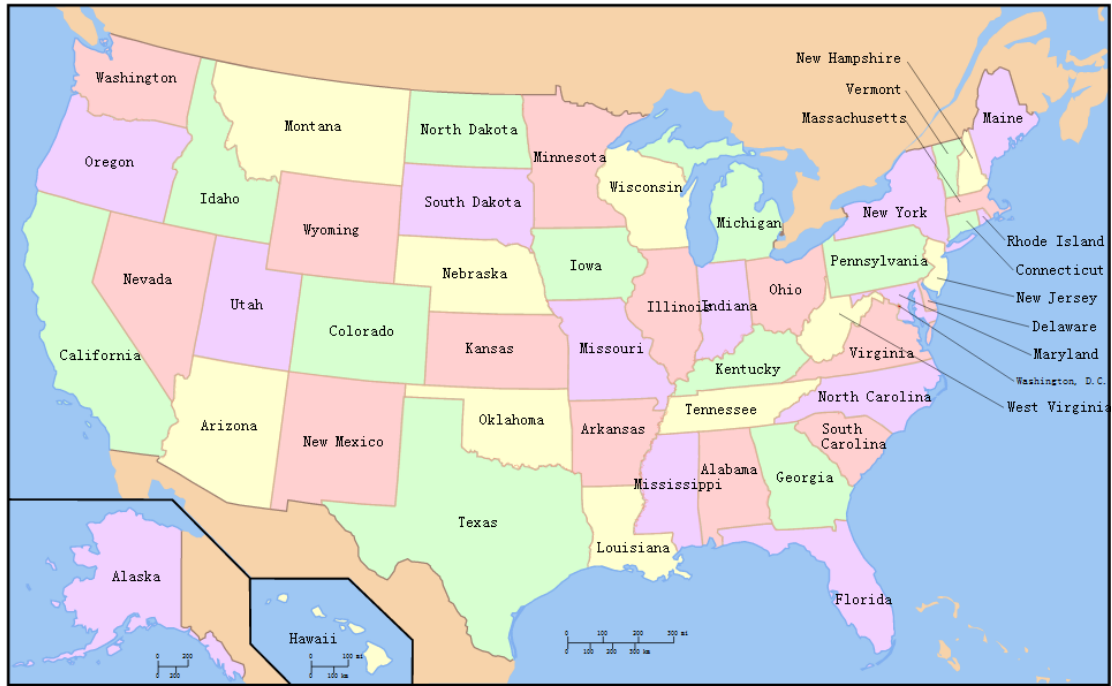
The goal of data compilation for sub-national GDP per capita by PPP is to make all sub-national units of China, India and the U.S. comparable to the rest of the world over time. Therefore, we have tried different combinations of methodologies and balanced their advantages against their shortcomings, and finally decided to apply a standardized compilation methodology for the time series data of all sub-national units.

1. Introduction

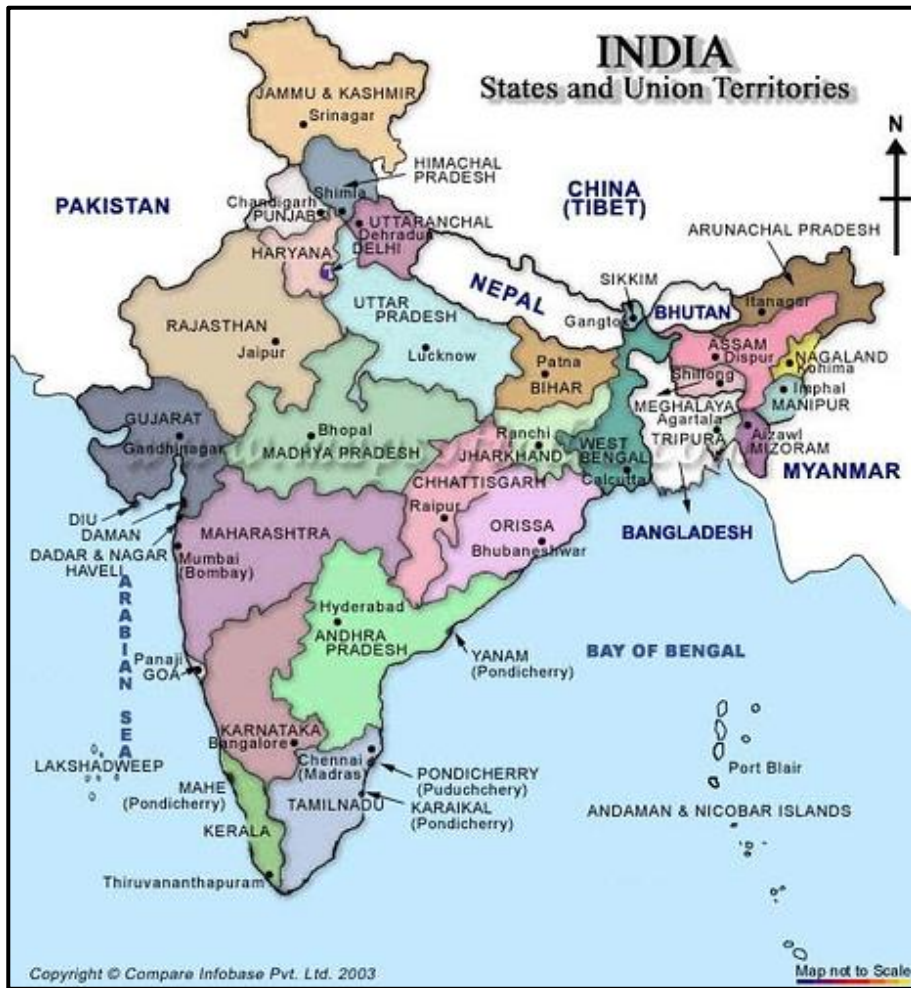
Nearly half of the world's population is now living in China, India and the U.S., which makes these 3 countries the biggest ones in the world. Therefore, the national data for these countries may not be sufficient or satisfactory to represent the individual situation and development of each sub-national unit, and it would also be interesting to compare each sub-national unit with other countries in the world, which is what triggered the original inspiration of the meta-data compilation recorded in this documentation.



Graph 1 Map of China sub-national units [Wikipedia website for China]



Graph 2 Map of U.S. sub-national units [Wikipedia website for U.S.]



Graph 3 Map of India sub-national units [Flickr website for India]

2. Data sources

In order to compile a comparable time series data set for GDP per capita of sub-national units in China, India and the U.S., we need two levels of raw data: the *national GDP per capita*, which is comparable to the rest of the world and the *sub-national GDP per capita*, which is consistent over time.

Gapminder (2009) provides a compiled database for *GDP per capita* of all countries in the world including China, India and the U.S., and the value is measured in *fixed PPP₂₀₀₅ dollars*. This time series database covers the period from 1800 to 2007, which has been compiled by *Lindgren (2009)* and more details can be found here: www.gapminder.org/downloads/documentation/#gd001

More details about the sub-national level of raw data have been explained below.

2.1 Data for sub-national units of China

There are 2 main sources for the raw data of sub-national units' *GDP per capita* of China. One is the official website of the National Bureau of Statistics of China, from which we could download both the national and divisional *GDP total* and also *population total* from 1995 to 2006. Then we calculated the *GDP per capita* for both the national and sub-national level by dividing *GDP total* by *population total*. The value of *GDP per capita* is measured in local currency "Yuan" and is calculated in *current price* of the given year. We coded this data source as "A"; the links to the raw data can be found here: <http://spreadsheets.google.com/pub?key=pp59adS3CHWfKPVb7dEexFA&gid=2>

The other source is the published report from the National Bureau of Statistics of China - "*China compendium of statistics 1949-2004*" (2005), from which we were able to obtain both the national and divisional *GDP per capita* covering 1949-2004, measured in local currency "Yuan" and is calculated in *current price* of the given year. We coded this data source as "B".

Both data source A and B have been compiled by the National Bureau of Statistics of China using the same counting units: "Yuan" in *current price*. However, there is a slight difference for the same observations for overlapping years (1995-2004), which is possibly due to the calculating process for data source A. In addition, most sub-China divisions have economy data since 1952; we deleted a handful of observations before 1952 and kept the rest.

2.2 Data for sub-national units of India

"National Human Development Report 2001" of India is the only data source we have found so far for the *GDP per capita* of India sub-national units. However, the raw data is not GDP per capita, but GNP per capita. Since we could not find a better source representing a consistent time series data for the sub-national India economy, we have to use this source for now. The counting unit of the raw data is in *Rupees at 1980-1981 prices*.

Some sub-national units of India have no data for certain years in this source; therefore we have to apply extrapolation methods for these missing observations. More details can be found here:

<http://spreadsheets.google.com/pub?key=pp59adS3CHWfKPVb7dEexFA&gid=2>

2.3 Data for sub-national units of the U.S.

There are two data sources for the *GDP per capita* of sub-national U.S.: the data before 1970 comes from a book "National Income and Wealth" of the United States; and the data (1970-2006) comes from the Bureau of Economic Analysis, Regional Economic Accounts.

In the book "National Income and Wealth" of the United States, only personal income per capita is available before 1970, so we assumed that GDP per capita of each state is to the same extent related to the personal income of that state, and the compilation methodology we have applied (*Chapter 3.2*) could quantify the level of the relationship between personal income and GDP per capita, and therefore could lead to the final results of each state's GDP per capita before 1970.

The raw data (1970-2006) of GDP per capita provided by Bureau of Economic Analysis, Regional Economic Accounts, has already been deflated by Bureau of Economic Analysis and the data is measured by *constant 2000 price* in *US Dollars*. More details can be found here: <http://www.bea.gov/regional/gsp/>

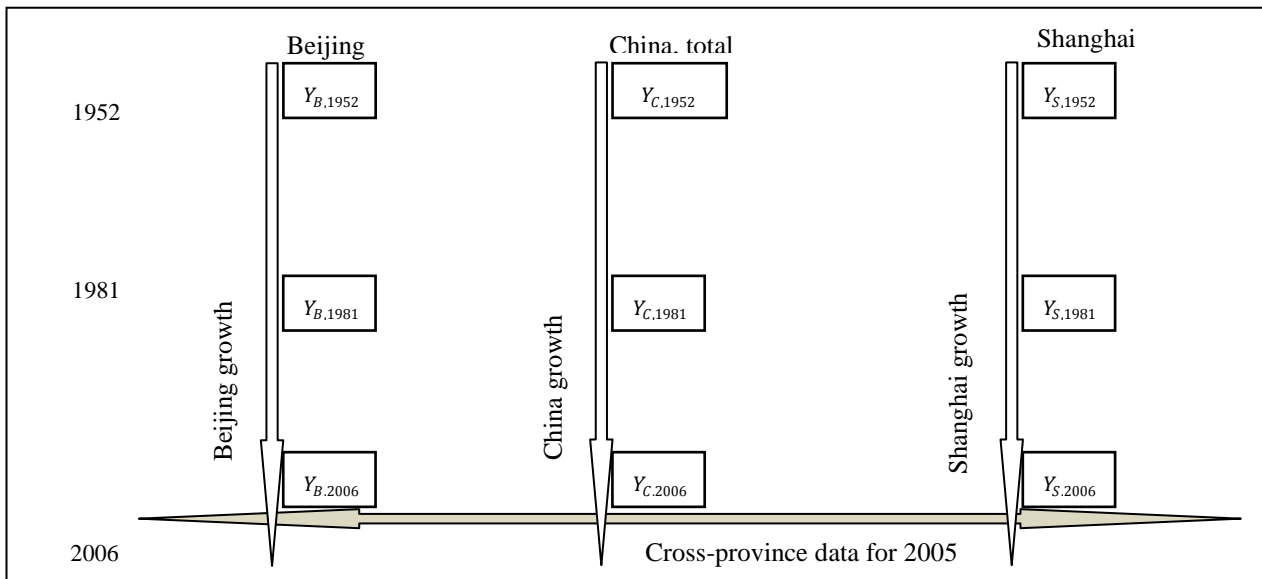
3. Data compilation methodology

We have two principal ways of making all the observations comparable, which are described below.

3.1 Compilation methodology (not used)

The first option, as displayed in *Graph 4* and *Equation 1*, is to use the real growth rate of each sub-national unit, and link the data series to the relative “position” of that unit in the world for one benchmark year, e.g. 2005.

(In order to explain the methodology, we use the sub-national units of China as examples.)



Graph 4 Compilation methodology (not used)

$$Y_{i,t} = \left(\frac{X_{i,t}}{X_{i,2005}} \right) * \left(\frac{X_{i,2005}}{X_{C,2005}} \times G_{C,2005} \right) = X_{i,t} \times \left(\frac{G_{C,2005}}{X_{C,2005}} \right)$$

Equation 1 Compilation methodology (not used)

$Y_{i,t}$ is provincial *GDP per capita* in *fixed PPP 2005*, for province i in year t .

$X_{i,t}$ is the provincial *GDP per capita* in *current price Yuan*, for province i in year t , as given by National Bureau of Statistics.

$X_{i,2005}$ is the provincial *GDP per capita* in *current price Yuan*, for province i in year 2005, as given by National Bureau of Statistics.

$X_{C,2005}$ is the *GDP per capita* in *current price Yuan*, for *total China* in year 2005, as given by National Bureau of Statistics.

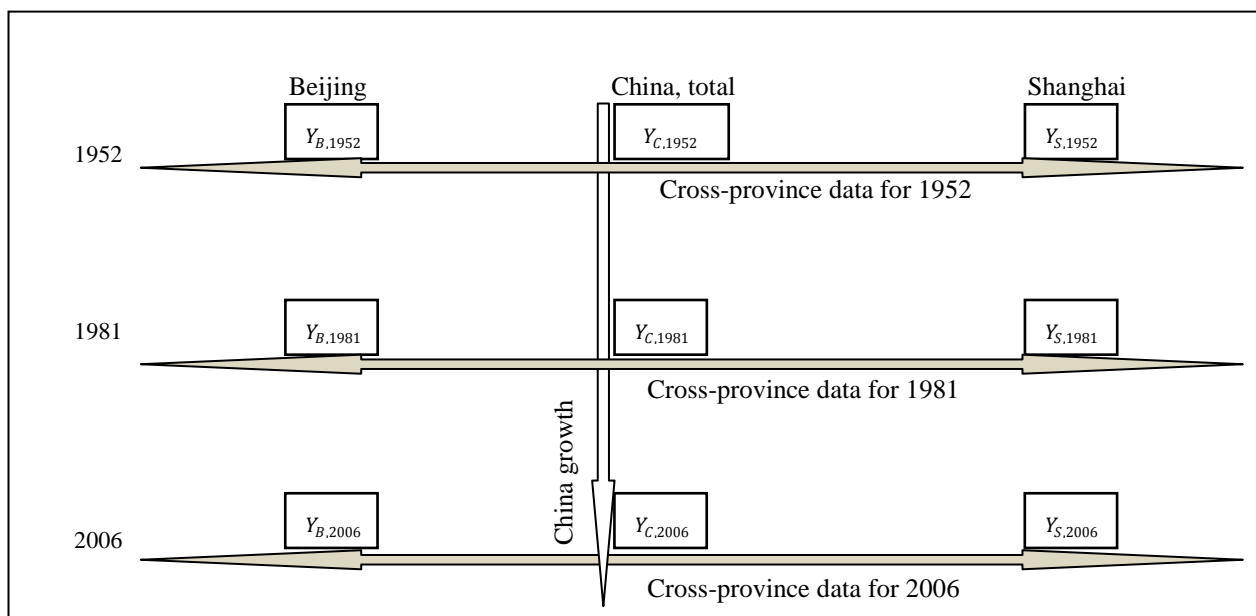
$G_{C,2005}$ is the *GDP per capita* in *fixed PPP 2005*, for *total China* in year 2005, as given by Gapminder (2008).

As described in the figure and formula above, this methodology could calculate out the real growth rate of each province over time; however, considering the comparability of the provinces to other countries, this method could generate a reliable result only in year 2005, which has been used as the benchmark year. For other years, the relative positions of Chinese economy in the world are not involved in this calculation, and the result will subsequently bear a high bias in comparability between Chinese provinces with other countries.

3.2 Compilation methodology (applied)

The other option is to calculate the relative economical “position” of each sub-national unit, comparing it against the whole nation for each year, and link it to the “position” of the nation in the whole world during that year, and then the relative “positions” of the sub-national units in the world for that year could be figured out, which is described in *Graph 5* and *Equation 2*.

(In order to explain the methodology, we use the sub-national units of China as examples.)



Graph 5 Compilation methodology (applied)

$$Y_{i,t} = \left(\frac{X_{i,t}}{X_{C,t}} \right) \times G_{C,t}$$

Equation 2 Compilation methodology (applied)

$Y_{i,t}$ is provincial *GDP per capita* in *fixed PPP 2005*, for province i in year t .

$X_{i,t}$ is the provincial *GDP per capita* in *current price Yuan*, for province i in year t , as given by National Bureau of Statistics.

$X_{C,t}$ is the *GDP per capita in current price Yuan*, for total China in year t , as given by National Bureau of Statistics.

$G_{C,t}$ is the *GDP per capita in fixed PPP 2005*, for total China in year t , as given by Gapminder (2008).

As described in the figure and formula above, this methodology has taken into account the relative position of the Chinese economy in the world for each year, and therefore could maximize the comparability between Chinese provinces with other countries.

In addition, this method has another advantage - as displayed in the equation below, the relative “positions” of the provinces (i and j) after our data compilation still remain the same as the original data from the National Bureau of Statistics, when $X_{C,t}$ and $G_{C,t}$ have been ruled out by calculation.

$$\frac{Y_{i,t}}{Y_{j,t}} = \frac{\left(\frac{X_{i,t}}{X_{C,t}} \right) \times G_{C,t}}{\left(\frac{X_{j,t}}{X_{C,t}} \right) \times G_{C,t}} = \frac{X_{i,t}}{X_{j,t}}$$

Equation 3 Advantage of compilation methodology (applied)

However, the real growth rate for each province over time has been changed due to this calculation, which has been displayed in the equation below so that no single factor can be ruled out.

$$\frac{Y_{i,t}}{Y_{i,(t-1)}} = \frac{\left(\frac{X_{i,t}}{X_{C,t}} \right) \times G_{C,t}}{\left(\frac{X_{i,(t-1)}}{X_{C,(t-1)}} \right) \times G_{C,(t-1)}}$$

Equation 4 Shortcoming of compilation methodology (applied)

Since our main purpose of data compilation is to achieve a reliable comparability between Chinese provinces and other countries in the world over the time series, the real growth rate for each province has been considered less important and therefore the methodology above (*Equation 2*) has been chosen. For sub-national units of India and the U.S., the raw data of GDP per capita is more or less similar to the raw data of sub-China, and therefore, we have applied the same methodology (*Equation 2*) for the compilation.

4. Discussion

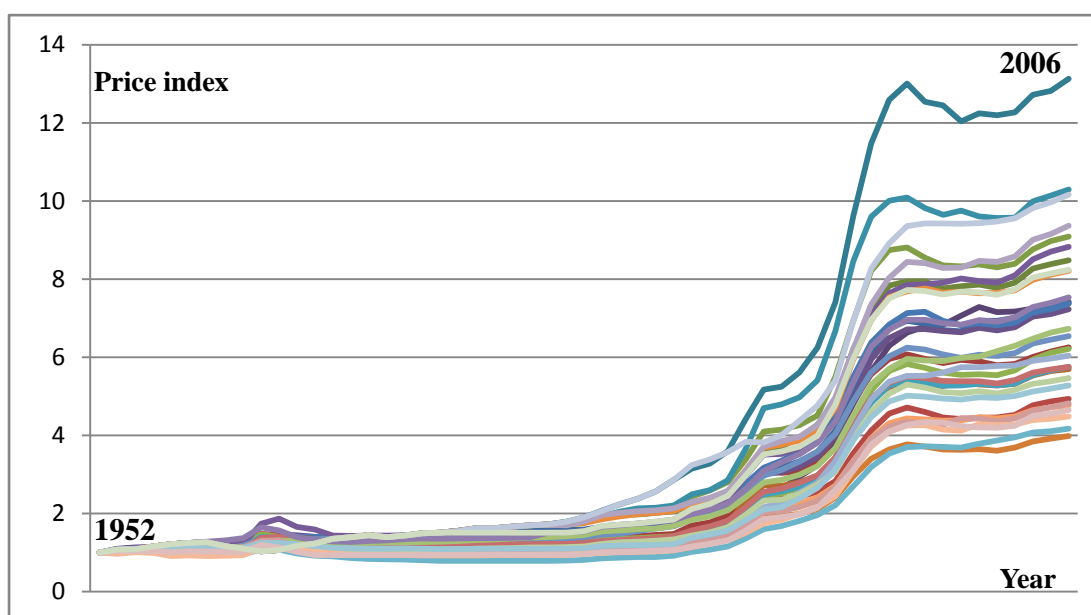
There is another feature of the sub-national units' economy data. Even though each sub-national unit within the same country is using the same currency, there still exists an internal difference of PPP (Purchasing Power Parity) among Chinese provinces, Indian states and the states of the U.S., and this keeps changing all the time. This case is due to a complex disparity in socio-economical development, availability of natural resources and other historical problems. In addition, this feature has brought in some bias to our data compilation.

4.1 GDP per capita of sub-national units of China

The compilation methodology we have chosen implicitly assumes that the price levels in all provinces are the same, i.e. assuming that all the provincial PPP-rates are 1 for all years. However, it is likely that there are substantial differences in the price levels of the provinces. Ideally, we would like to adjust for this, e.g. using some kind of provincial PPP-rates.

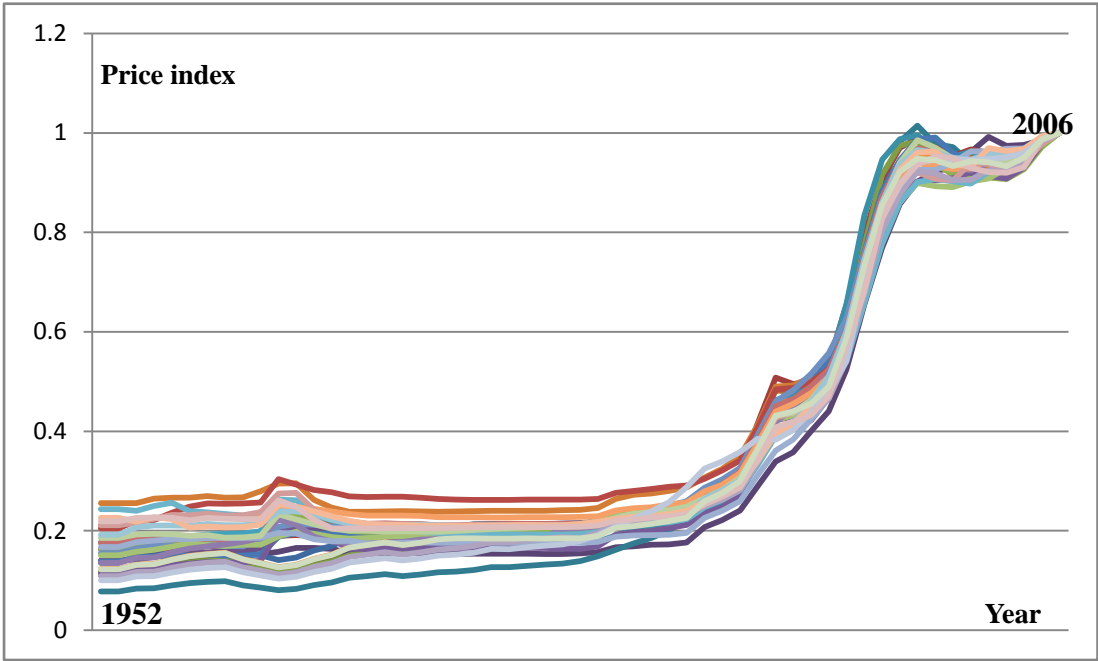
However, so far we are unable to find any data on provincial price levels. The National Bureau of Statistics only provides time series of inflation rates for each province and China total. We could get some sense of the size of the problem by assuming that all the provinces had the same price-levels in 1952, and then applying the inflation rates for all the subsequent years. This is done in *Graph 4* below.

The price level will be greatly unequal among different provinces of China for the latest decade. The biggest difference in 2006 is between Hebei and Chongqing (330%), which is too high to be acceptable.



Graph 6 Price index (1952 =1)

Instead, if we assume that all the price-levels were the same in 2006, the price level will be greatly unequal in 1952, with the biggest difference being 328%. This can be seen in *Graph 5* below.



Graph 7 Price index (2006=1)

Thus, the PPP-rates of the provinces seem to display some significant movements over time. The question is how much a “provincial PPP-adjustment” would affect our results.

We can use these two extreme assumptions to calculate two new sets of provincial GDP per capita. What we have done is to consider the price indexes as being the actual price levels of the provinces and use them to calculate a “provincial PPP-rate”. This provincial PPP-rate is used to adjust the provincial GDP per capita, so that the new indicator becomes:

$$Y_{i,t}^A = P_{C,t}^A / P_{i,t}^A * X_{i,t} * \left(\frac{G_{C,t}}{X_{C,t}} \right)$$

Equation 5 Compilation with price level assumptions

$Y_{i,t}^A$ is provincial *GDP per capita* in *fixed PPP 2005*, for province *i* in year *t*, assuming that the provinces had the same price levels in the year *A*.

$P_{C,t}^A$ is the corresponding price index for China as a whole. We calculate two new GDP per capita indicators with *A=1952* and *A=2006*, respectively.

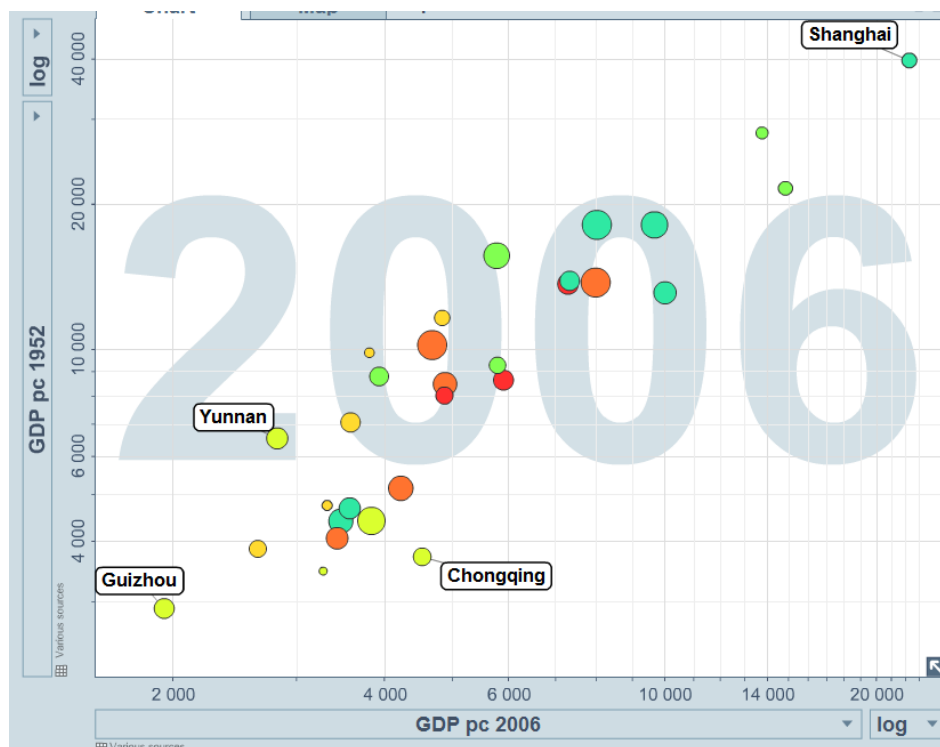
$P_{i,t}^A$ is the price index for province *i*, in year *t*, assuming that the provinces had the same price levels in the year *A*.

$X_{i,t}$ is the provincial *GDP per capita* in *current price Yuan*, for province *i* in year *t*, as given by National Bureau of Statistics.

$G_{C,t}$ is the GDP per capita in fixed PPP 2005, for total China in year t , as given by Gapminder (2008).

$X_{C,t}$ is the GDP per capita in current price Yuan, for total China in year t , as given by National Bureau of Statistics.

We plot these two indicators in Graph 6 below for the year 2006.



Graph 8 Provincial GDP per capita in 2006.
 (“GDP pc 2006” assumes that $A=2006$. “GDP pc 1952” assumes that $A=1952$.)

It is evident that the differences between these two extreme assumptions are substantial for several provinces, i.e. they switch their ranking (such as *Yunnan* and *Chongqing*). However, the broad pattern remains, for this year as well as for the earlier years. Hence, we believe that the lack of provincial prices certainly decrease the precision substantially, if we think of our indicator in terms of purchasing powers. However, the overall pattern is fairly robust with respect to this.

4.2 GDP per capita of sub-national units of India

The raw data from the "National Human Development Report 2001" of India is using *Rupees at 1980-1981 prices* as the counting unit, and so far we don't have any data about the price level of sub-national units of India. Thus, we have assumed in the compilation process that the price levels among the different states of India have always been equal over time. According to the general socio-economical status of India, especially the big differences between urban and rural areas, this assumption surely bears substantial bias for the final results of the compilation. However, we currently don't have other better options due to the availability of data, and the results from this compilation method still reveals the internal disparities among the different Indian states,

which is identical to the overall results we could conclude from the raw database in "National Human Development Report 2001" of India. So we believe the bias due to our assumption is acceptable.

4.3 GDP per capita of sub-national units of the U.S.

The time series data of GDP per capita for the individual states of the U.S. is quite similar to India in that no price level data is available. So we have applied again the assumption for compilation that the price levels among different states of the U.S. have always been equal over time. However, we believe the bias brought by this assumption should be much smaller than the one for India, because the U.S. has much smaller differences in economy among different states due to the highly integrated communication, traffic, and market system within the U.S. Therefore, the data quality of the U.S. sub-national GDP per capita from our compilation is thought to be the most reliable among these 3 data sets.

5. Conclusion

In order to achieve reliable comparability between the sub-national units' economy with the rest of the world, we have chosen the methodology that links the relative economical "positions" of each sub-national unit in China, India and the U.S. to the "position" of the individual nation in the world during that year. This method could maximize the comparability over time, but the growth rate of each sub-national unit after compilation is not consistent with the original database. Further studies are needed to achieve a higher level of data consistency and comparability over time series.

Also, due to the unavailability of the price level data among different sub-national units in these 3 countries, we have to assume that the price level among all sub-national units for a country have kept stable over time. Fortunately this assumption generates a more-or-less similar result compared to the one we could conclude from the raw database, and the bias brought by our assumption is subsequently considered acceptable. Hopefully in the near future we will be able to discover better solutions to this problem.

References

Bureau of Economic Analysis, Regional Economic Accounts. (Downloaded November 2008)

<http://www.bea.gov/regional/gsp/>

Flickr website for India. (Downloaded April 2009) <http://www.flickr.com/photos/sprol/53850243/>

Gapminder (2009). <http://www.gapminder.org/world>

Government of India. (2005) National Human Development Report 2001.

Government of United States. (1970) National Income and Wealth.

Lindgren, M. (2009) Documentation for GDP Per Capita by Purchasing Power Parities for countries and territories.

National Bureau of Statistics of China. (2005) China compendium of statistics 1949-2004.

Official website of National Bureau of Statistics of China. (Downloaded June 2008)

<http://www.stats.gov.cn/index.htm>

Wikipedia website for China. (Downloaded June 2008)

http://en.wikipedia.org/wiki/Template:PRC_provinces_big_imagemap

Wikipedia website for U.S. (Downloaded November 2008)

http://commons.wikimedia.org/wiki/File:Map_of_USA_with_state_names.svg