LUND UNIVERSITY
School of Economics and Management

# Reliability and Validity of GDP Time Series in Asia

Sebastian Maric

May 2018

*Abstract*: There is much literature on the topic of statistical comparability of African countries. However, the focus on Africa has left an information gap for other regions. This paper aims to fill some of that gap by comparing GDP data on Asian countries from some of the most frequently used datasets. Additionally, by comparing the base years and versions of the system of national accounts of Asian countries, an indication of the comparability of their national accounts is acquired. It is concluded that although the general comparability between the datasets is good, there are cases of large differences that can cause data users to come to different conclusions depending on what dataset is being used. To find and reduce the impact of these cases, it is recommended to cross-compare the datasets and, in the case of growth rates, average longer periods when possible. It is also found that the quality of Asian countries' statistics can vary substantially and therefore caution is advised when comparing countries. Finally, an improvement to the World Bank's statistical capacity indicator is suggested since it would have great benefits for data users.

Keywords: Asia, Comparability, GDP, National accounts, Statistics

Supervisor: Morten Jerven
Examiner: Thor Berger

# Acknowledgements

I would like to thank my family and girlfriend for their continued support and for putting up with me constantly talking about national accounts.

# Table of Contents

# List of Tables

# List of Figures

# 1 Introduction

Despite having several problems as an economic indicator for countries, GDP is still one of the most common economic indicator we have (Pilling, 2018; OECD Observer, 2005; Fioramonti, 2014). It is a measurement used to measure the economic activity of countries and is often used in the process of decision-making. For example, international institutions such as the World Bank gives loans to low and middle-income countries based on their GNI (which GDP is a large part of) while governments and central banks use GDP and GNI numbers in the policy-making process (Jerven, 2014; World Bank, 2012).

Seeing as GDP is such an important indicator and so widely used, it is critical to get it right. However, using data on GDP in decision-making or to draw conclusions is problematic since research has shown that the data differs depending on what dataset is used. For example, after creating rankings of 45 sub-Saharan countries according to GDP per capita using three of the most used international datasets, Jerven (2013a) concluded that the average change in rank for a given country between the datasets is 7 ranks. These differences are surprising considering the datasets are supposed to measure GDP for the same period and for the same countries.

Considering that the majority of the literature has focused on African countries, it should be a priority to determine whether problems such as these are constricted to sub-Saharan Africa, or if it is a general problem that most countries suffer from. This thesis aims to fill some of that gap by using similar methods to what Jerven (2013a, 2011a) has applied to sub-Saharan African countries and applying them to Asian countries. To inspect the problem from different perspectives, the analysis starts off with a general approach to get an overview of how comparable Asian countries' GDP per capita data are across different datasets. It then moves on to comparing the annual GDP growth rates of three countries to get more detailed information on how they differ between datasets. Finally, it attempts to determine the comparability between the national accounts of Asian countries.

## 1.1 Outline

Section 2 consists of a literature review and will provide an overview of some of the identified problems relating to comparability across datasets and comparing national accounts of different countries. Section 3 will then go through where the data I use come from and how it was changed to fit the analysis as well as the methods that are used in the thesis. Furthermore, it explains terms such as base year. It will also go through the limitations of the essay and how the selection of countries was performed. Section 4 consists of a data analysis and a discussion for each of the three methods. It ends with a summarizing discussion. Section 5 briefly explains the conclusions that have been drawn from each data analysis and provides recommendations for both data users and data compilers.

## 1.2 Research Questions

The goal of this essay is to investigate to what extent GDP data is comparable for Asian countries across different datasets, and to what extent Asian countries' national accounts are comparable. To find out, I will start off with my three research questions:

- "To what degree are Asian countries' data on GDP per capita comparable across the major international datasets?"
- "To what degree is the data on GDP growth rates from India, Pakistan, and the Philippines comparable across the official national accounts and the major international datasets?"
- "To what degree are the national accounts of Asian countries comparable?"

# 2 Literature Review

The goal of the literature review is to provide an overview of the different problems of GDP comparability both across datasets and across countries. It is important to understand that most of the literature on this specific topic revolves around the problems existing in Africa and as such, the literature review is also focused on Africa**.**

Jerven (2011b) explains that to be able to differentiate between economic growth and price increases when estimating income, prices are taken from a base year. The resulting information is often referred to as a constant series since the prices are held constant until the base year is updated. Jerven continues explaining that there are several problems relating to the base year for income estimates. First, since the prices are held constant until the base year is updated, any changes in prices will not be accounted for in the income estimates. Secondly, the importance of different sectors in the base year also stays the same. This means that structural changes in the economies will not be shown in the income estimates until the base year is updated. Third, sometimes the WDI and national statistical offices use different base years. This makes the comparability across these datasets, as well as across countries, poor.

Jerven (2013b, pp.20-21) explains that when international datasets compile their data on GDP, they take data from the national accounts produced by statistical offices from each country. He then explains that there are different series that the compilers have to choose from. Furthermore, he notes that the different versions sometimes also cover the same years and the process of choosing series and fitting them together are not being done in a transparent manner. This makes it hard for a data user to determine what dataset is better than the other.

Beyond the problems of collecting data, there are also problems with the international datasets themselves. For example, Young (2010) found that the PWT version 6.1 did not have any benchmark study of prices in 24 out of the 45 sub-Saharan African countries included in the dataset. Moreover, when comparing the World Bank's data with estimates from national statistical offices, Jerven (2013a) found that for the year 2009 or 2010, only 18 out of 48 countries had prepared estimates. The World Bank however, reported data for constant as well

as current prices for these countries up until 2009. So it seems that for the World Bank, quantity wins over quality of data.

Dudley Seers (1952) was hesitant of the benefits of using per capita income estimates for international comparison when a developing country is included in the comparison. He writes:

> In the hands of authorities, such international comparisons may yield correlations which throw light on the circumstances of economic progress, and they tell us something about relative inefficiencies and standards of living, but they are very widely abused. Do they not on the whole mislead more than they instruct, causing a net reduction in human knowledge (Seers, 1952, p.160)?

A similar view is shown by Deaton and Heston (2010), they recommend caution when comparing countries that are very different and especially when dealing with data that comes from countries with poor statistical capacity.

As discussed in Jerven (2013b, p.32), more research is needed to establish whether a bias exists in the data in developing countries. He mentions that Paul Collier (2009, p.9) is one of the few who have made a guess about the bias and he has suggested that the problems that exist around gathering economic data in the poorest countries have created an overestimation of their economies. Jerven (2013b, p.32) believes otherwise, stating that problems connected to data-gathering usually results in an underestimation of economies. An exceptional case of this is the 2010 upward revision of Ghana. The revision, which was done using newer methods and information from newer surveys concluded that the Ghanaian economy was undervalued by about 60%, and so Ghana went from being a low-income country to a lower-middle-income country overnight (Jerven, 2013c). Seeing as an extra 60% of GDP in Ghana went unnoticed until 2010, to what extent can we trust the data we have? The revision of Nigeria's national accounts in 2013 which showed an increase of estimated GDP of about 89%, showed that Ghana was not a unique case (The Economist, 2014).

Underestimation of GDP estimates is likely in many other developing countries for several reasons. According to Jerven (2011c), one of the most important reasons is how troublesome and expensive it can be to estimate the informal sector. He later explains that a lot of the times, the informal sector is large in developing countries which means that a sector that contributes a lot to developing countries may go largely unaccounted for.

The problem of GDP comparability between countries is further worsened by the varying quality of data, capabilities, and methods of countries' national accounts (Jerven & Johnston, 2015; Jerven, 2013c; Heston, 1994). Some of these issues are what Devarajan (2013, p. 9) refers to when he writes that Africa is "facing a statistical tragedy". He draws attention to the different methods used by different countries, the outdated population censuses of many countries, and the infrequency of poverty estimates and their inability to be compared over time. The underlying cause for these problems, he suggests, is that statistics are political and therefore many times controlled by both international and national politics. One of his suggested ways of fixing the statistical tragedy in Africa is transparency, which is in line with Jerven's (2012) thoughts**.** Jerven believes that the international dataset providers should be more transparent and provide metadata – information that the data user needs to understand and confidently use the data – such as definitions, sources, and methods used.

What I have described may seem quite dismal but it appears much is being done, at least in Africa, to improve the situation of statistics. Partnerships are being made, capacity is being built, more resources are being used, and overall, more is being done to create what Kiregyera (2013, p. 8) refers to as a "statistical renaissance" in counter to Devarajan's (2013, p. 9) mention of the "statistical tragedy".

Many developing countries have a hard time performing population censuses since most middle- and low-income countries have poor registration systems (Powell, 1981). Despite this, few improvements seem to have been made in civil registration in developing countries (Setel, Macfarlane, Szreter, Mikkelsen, Jha, Stout, AbouZahr, 2007). Additionally, Carr-Hill (2012) concludes that population censuses are increasingly being replaced by household surveys, resulting in population undercounting. To add to this, Carr-Hill (2014) writes about how problematic both population censuses and even more so household surveys can be when it comes to including the poorest of the poor. He mainly writes about the implications this has when assessing the progress towards development goals, however, this may have implications when trying to calculate per capita aggregates as well. Furthermore, Carr-Hill (2012) takes the view that population censuses are becoming less common while household surveys are taking their place. Jerven (2012) on the other hand, states that censuses and larger data collection projects are being preferred at the cost of frequent survey data.

Jerven (2014) writes that it is likely that the provision of statistics is worse in Africa than in other parts of the world. He continues with stating that poorer economies will have poorer

quality statistics as they have less resources to use and are less likely to keep formal records of their economic activities.

While scholars agree on the importance of being critical with the usage of data, Mudde and Schedler (2010, p. 412) note that when some scholars only have low-quality data to choose from, they choose the "bad data are better than no data" approach. Both Srinivasan (1994) and Woods (2014) come to similar conclusions and argue that the measurement errors in data are often underplayed or even ignored by scholars.

Srinivasan (1994, p. 4) criticizes publications of international agencies, stating that they give a "misleading, if not altogether false" representation of the quality of their data. Srinivasan has several recommendations. First, that international agencies should spend more resources to improve the statistical capacity of developing countries. Second, that efforts should be made to document to what extent developing countries are using recent numbers and represent reality. And finally, that cross-checking and validation should be documented when there are several sources available in a developing country.

To give a short summary, there are many problems with data quality and data aggregation that requires improvement, such as varying quality in data collection in the different official statistical offices and the international databases showing very different results for the same year and country (Devarajan, 2013; Jerven, 2013b, pp. 11-32). Furthermore, there is an expected agreement in the literature that one should always be critical towards the sources used, however, underplaying or ignoring poor data quality still appears to be a problem in practice (Deaton & Heston, 2010; Jerven, 2013b, pp. 8-9; Mudde & Schedler, 2010; Srinivasan, 1994; Woods, 2014). But a trend can still be seen among scholars to be consistent in recommending stronger action and transparency (Devarajan, 2013; Jerven, 2012) as well as highlighting errors from international organizations (Jerven, 2013a; Young, 2010).

# 3 Methodology

## 3.1 Data

To answer my first research question, I will use data on GDP per capita based on purchasing power parity in constant 2011 international dollars from three of the most used international databases on economic statistics and their most recent releases – the Penn World Tables (PWT) version 9.0 released in 2016, the World Development Indicators (WDI) from September 2017, and the Maddison dataset released in 2018. GDP based on purchasing power parity in constant 2011 international dollars can be explained as the purchasing power of a country when it has been converted into constant 2011 USD (OECD, 2018).

Due to the general nature of my first research question, I have also taken a general approach to answering the question. I have included all Asian countries of which the datasets had data on GDP per capita for the year 2010. The PWT do not have data on GDP per capita, so instead I divided the "cgdpe" (Expenditure side real GDP) variable with the "pop" (population) variable to go from GDP to GDP per capita.

For my second research question, I use data on growth rates from 1961-2012 from the PWT, the Maddison dataset, and the WDI and I have also included data from the official statistical offices of India, Pakistan, and the Philippines. The data was collected from the respective online websites.

Since the Maddison dataset only reports real GDP per capita numbers, I calculated the real GDP by multiplying the pop (population) variable with the rgdpnapc (real GDP per capita) variable. The official statistical offices and the WDI have data on growth rates available online while simple calculations are needed to get growth rates from both the Maddison dataset and the PWT. This was done using the following formula:

$$\frac{GDP_{final} - GDP_{Initial}}{GDP_{Initial}}$$

$GDP_{initial}$ represents the GDP of the year you want to calculate the growth rate of while $GDP_{final}$ represents the following year's GDP.

The official national growth rates for India and Pakistan come from a linked series (meaning they've been linked across several shorter series to produce a consistent series) found on the websites of their statistical offices. The official growth rates from India and Pakistan are calculated using GDP numbers at factor cost. "A Dictionary of Economics" (Black, Hasimzade & Myles, 2017) explains that GDP at factor cost means using the market prices minus indirect taxes, plus subsidies. The data for the Philippines' real GDP growth rates come from two linked series but they match perfectly where they overlap. However, the Philippine series does not display by which method it is calculated. I have been unable to get in contact with the official statistical office in the Philippines to ask them.

The prices used in the official series are as follows: the Philippines – constant 2000 national prices, India – constant 2004-2005 national prices, Pakistan – constant 1981 national prices from 1961-1999 and for the years 2000-2012 they use constant 2000 national prices.

The three international datasets use data based on constant national prices when calculating GDP. The WDI use constant national prices based on what base year the country is currently using. This means that the WDI calculates growth rates using 2000 national prices for the Philippines, 2011/2012 national prices for India, and 2005/2006 national prices for Pakistan. Both the Maddison dataset and the PWT use constant 2011 national prices. After growth rates have been computed, the international datasets convert their data into U.S. dollars to make their statistics comparable.

The section that examines the comparability between countries uses data from the WDI. The WDI receives their data on base years and system of national accounts (SNA) versions used from the respective statistical offices of each country. When a country reports using two base years, for example 2005/2006, it means that they use a fiscal year instead of the calendar year. The World Bank explains it like this:

> If a country's fiscal year ends before June 30, data are shown for the first year of the fiscal period; if the fiscal year ends on or after June 30, data are shown for the second year of the period (World Bank, 2018a).

For example, if a country reports using 2005/2006 as their base year and their fiscal year ends before June 30, it means that the prices they use come from aggregating data from the 3rd and 4th quarter of 2004 and the 1st and 2nd quarter of 2005 (World Bank, 2018a).

## 3.2 GDP per capita Comparison

You often see headlines like "The 25 poorest countries in the world" in articles and online websites or read that a certain country has a certain GDP per capita. But considering that this information depends on what database the data comes from, sometimes it is not easy to determine which country is the poorest, or what the level of GDP per capita is in a country. Table 1 shows the ranking of 43 Asian countries according to GDP per capita for the year 2010. This makes it easier to see the differences and therefore the comparability between each dataset. To make it easier to see the differences in rank between the datasets in Table 1, I have also included Table 2 which shows the change of rank for each country and each set of datasets. Table 2 also shows the average change of rank for each country. This approach allows me to not only give an overview of the datasets and how they overlap and differ, but also clearly see the differences across the datasets for any country.

## 3.3 Growth Rate Comparison

To give a contrasting view to the general nature and cross-sectional analysis of the first research question, I will use case studies on India, Pakistan, and the Philippines that examine how well the growth rates of real GDP from the years 1961-2012 compare for each country across the datasets. Table 3, 4, and 5 shows a correlation matrix for each country to give an overview of how well the growth rates compare across all four datasets. To highlight the differences between the datasets, Figure 1, 2, and 3 shows the annual disagreement between the highest and lowest reported growth rates among the datasets for each of the three countries and each year between 1961-2012. To lessen the impact of outliers in the data, Table 6, 7, and 8 shows the difference between the datasets by creating a table for each country that displays the average growth rate of different time-periods for each dataset. They also show the highest and lowest average growth rate across all datasets for each period as well as the difference between these two.

## 3.4 National Accounts Comparison

While the other two sections focus on the comparability between the datasets, this section aims to examine the comparability between the national accounts of Asian countries. It will begin with a detailed comparison between the national accounts of Australia and Nepal. By comparing two countries with different statistical capacities, I hope to highlight how large the differences can be between the data in the same dataset. Additionally, to make a general comparison of Asian countries national accounts, Figure 4 and 5 shows the distribution of countries based on what base year and what version of the SNA the countries are using.

National accounts usually provide GDP data in both constant and current prices. Jerven (2013a) explains that the prices that are used in the constant series are determined by the base year. He notes that the advantage of using constant prices is that it enables national accountants to differentiate between economic growth and price increases. Furthermore, he explains that the base year also determines the importance of each sector. In other words, if a sector is small in the base year, it will still contribute a small amount to the GDP even if it has grown large over time. This means that the base year plays a deciding role in how representative the national accounts are of the economic situation in a country.

The United Nations (2018a) explains that the SNA are international standards that were created to achieve harmonization of the production of statistics. It is also explained that as the world economy changed and improved ways of compiling data were introduced, these international standards were also updated. The world economy in the 1960s is very different from the economy we live in today and therefore, it is crucial that the countries have one of the more recent versions of the SNA. Each country has their own SNA to follow but they are built on the SNA from the United Nations.

## 3.5 Limitations

A limitation of the GDP per capita comparison is its lack of scope over time since it only shows the rankings from 2010. Had I had more time, I would have included other years in the analysis as well. However, since there is a risk that 2010 may be a year in which the rankings correlated exceptionally well or bad compared to other years, further testing of other years

beyond this thesis is needed to get a definite answer to what degree the datasets are comparable.

Only three countries were included in the data analysis of the GDP growth rates because of the time constraint. I cannot be sure that the GDP statistics of the Philippines, India, and Pakistan are representative enough for the results to be generalized to other developing countries in Asia.

Another noteworthy point is that the official statistical offices of countries sometimes revise their estimates and because of this the data that come from the official statistical offices may not be the same later on.

## 3.6 Selection of Countries

The selection was made from the World Bank's lending groups that include Asian countries; East Asia and Pacific, Europe and Central Asia, Middle East and North Africa, and South Asia.[1] All non-Asian countries were excluded except for Australia and New Zealand. The countries that had data for GDP per capita for the year 2010 are included in the selection. The countries that were excluded specifically because the PWT were missing data are Syria, Taiwan, and Afghanistan. The countries excluded specifically because the Maddison dataset was missing data are Brunei, Fiji, Macao, Bhutan, and the Maldives. Furthermore, both the Maddison dataset and the PWT did not have any data on Guam, Kiribati, the Marshal Islands, Micronesia, Nauru, Northern Mariana Islands, Palau, Samoa, Solomon Islands, Timor-Leste, Papua New Guinea, Tonga, Tuvalu, Vanuatu, and Gaza. The WDI, the Maddison dataset, and the PWT were all missing data for French Polynesia, North Korea, and New Caledonia. Most of the excluded countries are island nations.

The countries for the growth rate comparisons were chosen on the basis of data availability and population. Data availability was the most constraining factor in the selection of countries. Data availability also played a larger role in the selection than population but I tried choosing countries with high populations to get a representation of as many people as

[1] The categories can be found here: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

possible. My selection was also narrowed down to the Asian countries considered low-income countries or lower-middle income countries as of 2018 by the World Bank because I wanted to examine developing countries more closely. The countries I chose are the Philippines, India, and Pakistan. All three countries have historical series available which has allowed me to compare the growth rates of the countries as far back as 1961.

Australia and Nepal were chosen for the national accounts comparison specifically because of the economic contrast between them. When looking at economic databases, it is easy to forget that the quality of the data may differ substantially between countries since they are all displayed in the same way. By choosing countries that I assumed have different levels of statistical capacity, it was possible to highlight how much the process of creating national accounts can differ between countries.

The countries included in the section that examines the general comparability between Asian countries' national accounts use the same group of countries that the GDP per capita analysis does.

# 4 Analysis and Discussion

## 4.1 GDP per capita Comparison

To determine how well the PWT, the Maddison dataset, and the WDI compare, Table 1 shows the rankings of 43 Asian countries according to GDP per capita based on purchasing power parity for each dataset for the year 2010. The richest countries are at the bottom and the poorest countries at the top. Only countries which have data on GDP per capita available for the year 2010 across all three datasets are included. Many countries show no differences or a difference of one rank between the datasets but there are also countries that jump several ranks. The country that has the highest change of rank across all datasets is Uzbekistan which is ranked #13 in both the PWT and the Maddison dataset but ranked #8 in the WDI, a difference of 5 ranks. Another country that changes a lot between the datasets is Turkmenistan which is ranked #24 in the PWT, #25 in the Maddison dataset, and #21 in the WDI.

**Table 1**
Rankings of Asian Economies According to GDP per capita in Three Datasets (International USD)

| Rank | PWT | | Maddison | | WDI | |
|---|---|---|---|---|---|---|
| 1 | Nepal | 1813 | Nepal | 2022 | Nepal | 1986 |
| 2 | Bangladesh | 2311 | Cambodia | 2322 | Tajikistan | 2106 |
| 3 | Cambodia | 2434 | Bangladesh | 2421 | Bangladesh | 2443 |
| 4 | Tajikistan | 2438 | Tajikistan | 2893 | Cambodia | 2523 |
| 5 | Kyrgyzstan | 2851 | Kyrgyzstan | 3381 | Kyrgyzstan | 2790 |
| 6 | Myanmar | 3493 | Myanmar | 3384 | Myanmar | 3721 |
| 7 | Laos | 3770 | Laos | 3791 | Laos | 3984 |
| 8 | Pakistan | 4052 | Pakistan | 4133 | Uzbekistan | 4240 |
| 9 | Yemen | 4088 | India | 4353 | Pakistan | 4284 |
| 10 | India | 4227 | Yemen | 4363 | India | 4405 |
| 11 | Vietnam | 4231 | Vietnam | 4465 | Yemen | 4479 |
| 12 | Philippines | 5454 | Philippines | 5449 | Vietnam | 4486 |
| 13 | Uzbekistan | 6618 | Uzbekistan | 6812 | Philippines | 5597 |
| 14 | Armenia | 7063 | Indonesia | 7378 | Armenia | 6703 |
| 15 | Georgia | 7263 | Mongolia | 7487 | Georgia | 6734 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16 | Mongolia | 7298 | Armenia | 7642 | Mongolia | 7709 |
| 17 | Indonesia | 7457 | Sri Lanka | 7844 | Indonesia | 8433 |
| 18 | Sri Lanka | 7615 | Georgia | 8305 | Sri Lanka | 8563 |
| 19 | Iraq | 9404 | Iraq | 9426 | Jordan | 9473 |
| 20 | China | 9476 | China | 9475 | China | 9526 |
| 21 | Jordan | 9728 | Jordan | 10 594 | Turkmenistan | 9942 |
| 22 | Thailand | 13 078 | Thailand | 12 586 | Iraq | 12 718 |
| 23 | Azerbaijan | 14 041 | Azerbaijan | 12 866 | Thailand | 13 487 |
| 24 | Turkmenistan | 14 808 | Turkey | 16 211 | Azerbaijan | 15 950 |
| 25 | Turkey | 16 135 | Turkmenistan | 16 391 | Lebanon | 16 281 |
| 26 | Lebanon | 16 921 | Lebanon | 16 988 | Iran | 17 444 |
| 27 | Iran | 17 254 | Malaysia | 17 653 | Turkey | 17 959 |
| 28 | Kazakhstan | 18 011 | Iran | 17 716 | Kazakhstan | 20 097 |
| 29 | Malaysia | 19 780 | Kazakhstan | 17 738 | Malaysia | 21 107 |
| 30 | Russia | 20 115 | Russia | 20 255 | Russia | 23 108 |
| 31 | Israel | 30 869 | Israel | 29 319 | Israel | 29 600 |
| 32 | New Zealand | 31 154 | New Zealand | 30 904 | South Korea | 30 352 |
| 33 | South Korea | 32 304 | South Korea | 32 325 | New Zealand | 32 253 |
| 34 | Japan | 34 721 | Japan | 35 477 | Japan | 35 750 |
| 35 | Bahrain | 39 572 | Bahrain | 37 062 | Bahrain | 41 107 |
| 36 | Saudi Arabia | 40 431 | Saudi Arabia | 39 370 | Australia | 41 385 |
| 37 | Australia | 43 068 | Oman | 43 260 | Oman | 45 335 |
| 38 | Oman | 43 778 | Hong Kong | 44 857 | Saudi Arabia | 45 421 |
| 39 | Hong Kong | 48 067 | Australia | 45 398 | Hong Kong | 48 108 |
| 40 | United Arab Emirates | 56 538 | United Arab Emirates | 58 565 | United Arab Emirates | 57 580 |
| 41 | Kuwait | 62 529 | Singapore | 61 827 | Singapore | 72 105 |
| 42 | Singapore | 69 228 | Kuwait | 63 719 | Kuwait | 75 204 |
| 43 | Qatar | 121 944 | Qatar | 123 057 | Qatar | 125 141 |

Sources: PWT (2015); The Maddison Dataset (2018); WDI (2017)

Table 2 is based on Table 1 and shows the average change of rank across all datasets for each country, as well as the average change of rank across each set of datasets. The first three columns in the last row shows the average change of rank for a given country for each pair of datasets. Out of these 43 countries, 28 has an average change below 1 rank, while 11 of those 28 are a perfect match across all datasets. Out of the remaining 15 countries, 7 have an average change of 2 ranks or more. The two countries with the largest average difference are Uzbekistan with an average change of about 3.3 ranks and Turkmenistan with an average change of about 2.7 ranks. The countries with the second highest average change of rank are Georgia, Indonesia, Iraq, Turkey, and Australia with an average change of 2 ranks. There is an average change of 0.6 ranks between the PWT and the Maddison dataset, between the PWT and the WDI an average change of 0.8 ranks, and between the Maddison dataset and the WDI an average change of 1.2 ranks. The average change of rank of a given country across all

datasets is 0.8 ranks. This can be put in contrast to Jerven's (2010) findings that the average change in rank for sub-Saharan countries between the WDI and the Maddison dataset (although he used different versions than I do) is 7 ranks.

**Table 2**

Difference in Ranking According to GDP per capita Between Three Datasets

| Country | PWT-Mad | PWT-WDI | Mad-WDI | Average |
|---|---|---|---|---|
| Nepal | 0 | 0 | 0 | 0 |
| Bangladesh | 1 | 1 | 0 | 0,7 |
| Cambodia | 1 | 1 | 2 | 1,3 |
| Tajikistan | 0 | 2 | 2 | 1,3 |
| Kyrgyzstan | 0 | 0 | 0 | 0 |
| Myanmar | 0 | 0 | 0 | 0 |
| Laos | 0 | 0 | 0 | 0 |
| Pakistan | 0 | 1 | 1 | 0,7 |
| Yemen | 1 | 2 | 1 | 1,3 |
| India | 1 | 0 | 1 | 0,7 |
| Vietnam | 0 | 1 | 1 | 0,7 |
| Philippines | 0 | 1 | 1 | 0,7 |
| Uzbekistan | 0 | 5 | 5 | 3,3 |
| Armenia | 2 | 0 | 2 | 1,3 |
| Georgia | 3 | 0 | 3 | 2 |
| Mongolia | 1 | 0 | 1 | 0,7 |
| Indonesia | 3 | 0 | 3 | 2 |
| Sri Lanka | 1 | 0 | 1 | 0,7 |
| Iraq | 0 | 3 | 3 | 2 |
| China | 0 | 0 | 0 | 0 |
| Jordan | 0 | 2 | 2 | 1,3 |
| Thailand | 0 | 1 | 1 | 0,7 |
| Azerbaijan | 0 | 1 | 1 | 0,7 |
| Turkmenistan | 1 | 3 | 4 | 2,7 |
| Turkey | 1 | 2 | 3 | 2 |
| Lebanon | 0 | 1 | 1 | 0,7 |
| Iran | 1 | 1 | 2 | 1,3 |
| Kazakhstan | 1 | 0 | 1 | 0,7 |
| Malaysia | 2 | 0 | 2 | 1,3 |
| Russia | 0 | 0 | 0 | 0 |
| Israel | 0 | 0 | 0 | 0 |
| New Zealand | 0 | 1 | 1 | 0,7 |
| South Korea | 0 | 1 | 1 | 0,7 |
| Japan | 0 | 0 | 0 | 0 |
| Bahrain | 0 | 0 | 0 | 0 |
| Saudi Arabia | 0 | 2 | 2 | 1,3 |

| | | | | |
|---|---|---|---|---|
| Australia | 2 | 1 | 3 | 2 |
| Oman | 1 | 1 | 0 | 0,7 |
| Hong Kong | 1 | 0 | 1 | 0,7 |
| United Arab Emirates | 0 | 0 | 0 | 0 |
| Kuwait | 1 | 1 | 0 | 0,7 |
| Singapore | 1 | 1 | 0 | 0,7 |
| Qatar | 0 | 0 | 0 | 0 |

Sources: PWT (2015); The Maddison Dataset (2018); WDI (2017)

### 4.1.1 Discussion

Except for the cases of Uzbekistan and Turkmenistan, the datasets match well when it comes to ranking countries' GDP per capita. It can see that there is a larger difference between the WDI and the other datasets compared to the difference between the other two datasets. Furthermore, there is no major difference between the richer half of the countries compared to the poorer half. Still, the differences are in most cases nothing major and shouldn't make data users draw different conclusions based on what dataset they use. The difference between the GDP per capita numbers of the datasets tells a different story however. Since all of the three datasets report their data for GDP per capita in 2011 US prices, one would expect any differences between the reported GDP per capita data to be small. This is the case in many of the listed countries but there are also several countries that show some differences and a few countries that show very large differences between the datasets. There are some countries that show differences that may not seem large to begin with but are substantial in proportion to their reported GDP per capita. Both Turkmenistan and Uzbekistan, which we've seen have large differences between the rankings of the datasets, show relatively large differences compared to their reported GDP per capita. The PWT reports a GDP per capita of 14 808 USD for Turkmenistan, the Maddison dataset 16 391 USD, and the WDI 9942 USD. The difference between the Maddison dataset and the WDI is 6449 USD. The largest difference for Uzbekistan is between the Maddison dataset which reports 6812 USD, and the WDI reports 4240 USD. The largest difference between the datasets in terms of numbers comes from Kuwait, which the PWT reports a GDP per capita of 62 529 USD, the Maddison dataset reports 63 719 USD, and the WDI reports 75 204 USD. Put in numbers, that's a difference of 12 675 USD. Another example of a large difference is Singapore for which the PWT reported 69 228 USD, Maddison reported 61 827 USD, and the WDI reported 72 105 USD. That's a difference of 10 278 USD between the WDI and the Maddison dataset. The examples shown above are some of the most extreme cases and Table 1 also shows countries with differences

that aren't as extreme but still have a difference of thousands of USD between the datasets. However, many countries match well across the datasets. Since the comparability varies so much across countries, GDP data should always be cross-checked with other sources to determine whether the data match across the datasets or not. Sadly, if they do not match well, it is hard for a data user to determine which dataset is best for a specific purpose.

A possible explanation to why these differences exist is that different datasets use different methods, and these different methods can vary in their ability to calculate economies of a certain structure (Deaton & Heston, 2010). For example, one method may be good at estimating economic activity in countries that rely heavily on oil exports while another method may be better at estimating the economic activity in countries that mostly rely on financial services. There is no method that can perfectly calculate the GDP of every country, and if there were we wouldn't need more than one international dataset. Using different methods for different types of countries wouldn't work well either since that would mean these types of countries wouldn't be comparable. Another possible factor to the differences is that different datasets sometimes have different population estimates for the same country and the same year. Furthermore, if a dataset released a version recently while it has been years since another dataset released a version, any revisions from countries' national accounts that happened between those years will only be included in the most recent database. For example, any revisions to any country's national accounts that has happened since the latest version of the PWT was released in 2015 will be included in the latest version of the Maddison that was released in 2018 but not in the PWT.

According to Jerven (2011a), a country may have national accounts series that overlap but use different base years, and the dataset compilers sometimes have to put these different versions together to get a complete time series. I could not find how this was done in the user guides of the international datasets but the process is another potential source of the differences between the datasets. As Jerven (2011a, p. 383) puts it: "It is unsatisfactory to work with data where no proper sources are given and where there is no immediate indication as to why different sources disagree". He later writes that the only way to avoid the problem is to go to the source of the data, the national accounts. However, he continues with noting that it is not practical since they are not available online. Jerven (2016) suggests that the ways that the primary data is changed by each dataset may be the main reason for the differences between the datasets.

The GDP per capita comparisons have enabled me to, at least on a general level, answer my first research question about to what degree the data on GDP per capita from Asian countries are comparable across the PWT, the Maddison dataset, and the WDI. Table 1 and 2 have shown that the rankings according to GDP per capita, with a couple of exceptions, compare quite well across the three datasets. When we look at the actual numbers, however, Table 1 shows that the data on GDP per capita have at least minor differences between the datasets for all countries. But for some countries they vary considerably. Therefore, the comparability between the countries can be considered good overall, but data users should cross-check GDP data with other datasets to find the cases where there are large differences.

The problem here is that it is hard for a data user to determine what dataset should be chosen when there are large differences. Jerven (2016) tells us that in most cases, the datasets do not provide an immediate indication of why different datasets report different numbers for the same measurement. A suggestion to improve the problem is for the dataset compilers to create online websites or separate databases that provide the source data, go through the process of compiling the data, and gives step by step instructions on how to use the data. A partial solution to the problem comes from Jerven (2012) who suggests that the dataset compilers should provide metadata along with the dataset to enable data users to use the datasets effectively. The WDI has done a good job of this in their beta version of their database. It provides metadata for countries, for example, what base years they use, when the latest population census was performed, what SNA the countries use, and so on. It also provides information about their series, for example, definitions on GDP per capita and purchasing power parity, and what prices were used to calculate them. Additionally, the World Bank's knowledge base contains searchable information about the World Bank's methods and definitions. If a data user can't find what they're looking for in the metadata or the knowledge base, they can also e-mail the WDIs' support which in my experience has answered within a day or two. The only problem is that it is on another website and not directly accessible from the WDI online database. Both the Maddison dataset and the PWT have user manuals which both have introductions that do a good job at explaining the datasets and variables, but as they go on they become harder to follow for anyone without experience in statistics. Finally, information about what a term means or how something is calculated can be found in the World Bank's knowledge base.

## 4.2 Growth Rate Comparison

The discussion section of the GDP per capita comparisons explains some reasons to why one would expect differences across the datasets and they apply to the growth comparisons as well. However, a reason specific to this section is that the variables I have used from the datasets were not all computed using the same national prices. As explained in the data section, the prices that the WDI use depend on the countries' reported base year. The PWT and the Maddison dataset however, use constant 2011 prices. There are other methodological differences between the datasets but discussing all of them is outside of the scope of this thesis. To compare the methodology of the datasets, the data manuals of the PWT (Feenstra, Inklaar, and Timmer, 2015), the Maddison dataset (Bolt, et al. 2018), as well as the World Bank's (2018b) knowledge base can be compared.

Table 3, 4, and 5 shows the correlation between the reported growth rates of the Philippines, India, and Pakistan; for the years 1961-2012; as reported by official numbers, the PWT, the Maddison dataset, and the WDI. This test should give an indication of how much the different methods used and the factors described above affect the datasets. When rounding up to the closest hundredth, the Philippines show a perfect correlation across all data sources. Overall, India's correlation matrix shows quite good correlations as well. Pakistan however, shows a worse correlation between the datasets and especially bad is Pakistan's correlation between the Maddison data and the national data. The average difference between the highest and lowest reporting dataset for the years 1961-2010 is 0.03% for the Philippines, 1.00% for India, and 1.47% for Pakistan.

**Table 3** – Correlation Matrix of Annual Growth Rates for the Philippines, 1961-2012

| | Philippines | PWT | Maddison | WDI |
|---|---|---|---|---|
| Philippines | 1,00 | | | |
| PWT | 1,00 | 1,00 | | |
| MADDISON | 1,00 | 1,00 | 1,00 | |
| WDI | 1,00 | 1,00 | 1,00 | 1,00 |

Sources: National accounts of the Philippines (2012, 2018); PWT (2015); The Maddison Dataset (2018); WDI (2017)

**Table 4** – Correlation Matrix of Annual Growth Rates for India, 1961-2012

| | India | PWT | MADDISON | WDI |
|---|---|---|---|---|
| India | 1,00 | | | |
| PWT | 0,93 | 1,00 | | |
| MADDISON | 0,99 | 0,93 | 1,00 | |
| WDI | 0,98 | 0,96 | 0,97 | 1,00 |

Sources: National accounts of India; PWT (2015); The Maddison Dataset (2018); WDI (2017)

**Table 5** – Correlation Matrix of Annual Growth Rates for Pakistan, 1961-2012

| | Pakistan | PWT | MADDISON | WDI |
|---|---|---|---|---|
| Pakistan | 1,00 | | | |
| PWT | 0,92 | 1,00 | | |
| MADDISON | 0,77 | 0,86 | 1,00 | |
| WDI | 0,86 | 0,95 | 0,84 | 1,00 |

Sources: National accounts of Pakistan; PWT (2015); The Maddison Dataset (2018); WDI (2017)

To get a closer look at how large the differences in growth rates are between the datasets, Figure 1, 2, and 3 shows the highest and lowest reported growth rate for each year from 1961-2012 for each country across all four datasets. In Figure 1 you can barely see the other line in the figure, which means that the growth rates of the Philippines match almost perfectly. The biggest difference across all years was in 1963 when the PWT reported a growth rate of 6.81% while the Maddison dataset reported 7.07%; a difference of 0.26%. Every difference in the Philippines series larger than 0.02% is caused by the PWT. Moving on to Figure 2, we can see that there are several years in India which have large differences between the datasets. 1968, 1969, 1970, 1999, and 2008 all have a difference larger than 2% between the highest and lowest reporting dataset. The year in which the largest difference between the datasets in India can be found is 1969 when the PWT reported a growth rate around 4.5-4.7% higher than the other dataset depending on what dataset the PWT is compared to. In Figure 3 we can see that the differences between the highest and lowest reporting dataset for Pakistan are larger than India. In 13 out of the 51 years, there's a difference larger than 2% between the datasets. That's more than 1 fifth of the years included in the comparison. For Pakistan, the year with the largest difference across the datasets is 1965 where the largest difference between the datasets is 6%, the Maddison dataset reported a growth rate of 4.42% while the others reported growth rates of around 9% and 10%.

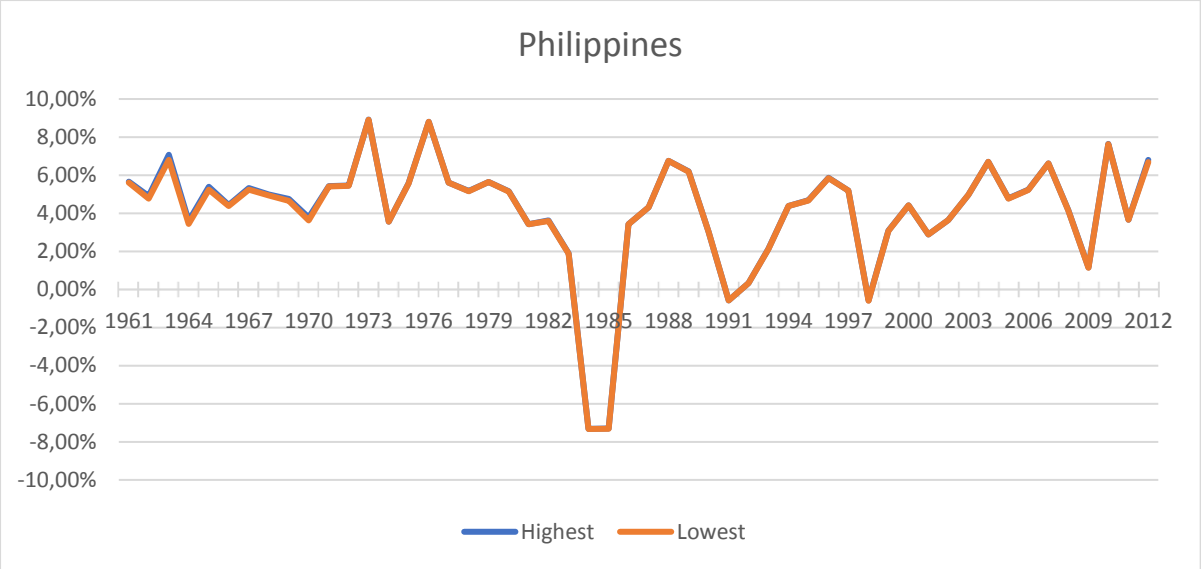


**Figure 1**: Annual Disagreement Between the Highest and Lowest Reported Growth Rates Across Four Datasets, the Philippines.

Sources: National accounts of the Philippines (2012, 2018); PWT (2015); The Maddison Dataset (2018); WDI (2017)

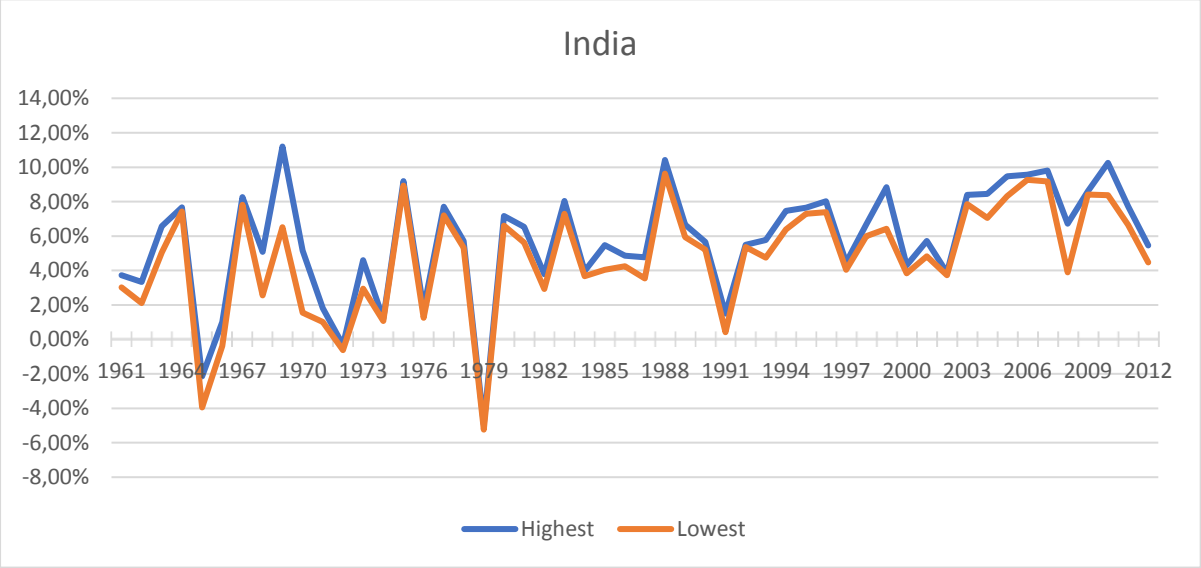


**Figure 2**: Annual Disagreement Between the Highest and Lowest Reported Growth Rates Across Four Datasets, India.

Sources: National accounts of India; PWT (2015); The Maddison Dataset (2018); WDI (2017)

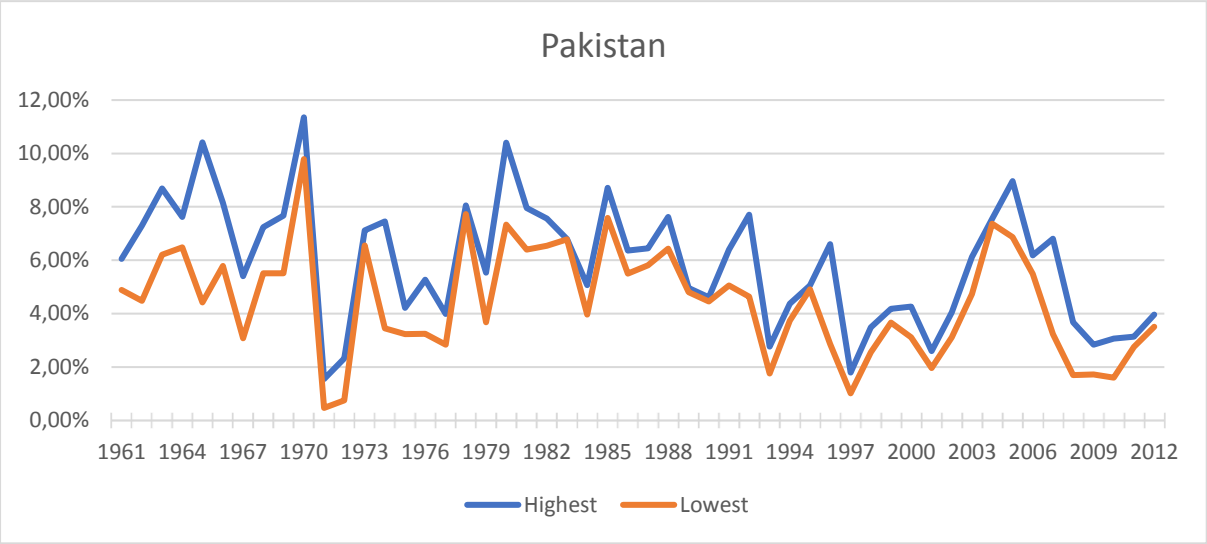


**Figure 3**: Annual Disagreement Between the Highest and Lowest Reported Growth Rates Across Four Datasets, Pakistan

Sources: National accounts of Pakistan; PWT (2015); The Maddison Dataset (2018); WDI (2017)

When going through the growth rates I found that in 1969 – the year with the largest difference between the datasets for India – the PWT reported a growth rate of 11.2% for India while the other datasets reported growth rates around 6.6%. So, the PWT reported a growth rate about 4.6% higher than the others. The following year however, the PWT reported a growth rate about 3.5% lower than the others. Either this is a coincidence or it was a correction. If it is a correction, it is likely that a mistake was made when the dataset compilers were calculating India's growth rates for 1969 which then made way for a "natural" correction for the following year when there was no mistake and everything was done correctly. I'm going to use an example to illustrate what I mean by "natural" correction. Let's assume that the PWT, the Maddison dataset, and the WDI report a GDP of 100 USD for country X in the year 1999. The PWT makes a mistake and reports a GDP of 110 in the year 2000 when it should've reported a GDP of 102. The PWT's reported growth rate for 1999-2000 will then be 10% instead of 2%. Meanwhile, let's say the other datasets all reported a growth rate of 2%. The following year, the economy of country X is booming and when the GDP for the year 2001 is calculated and everything is done correctly, the PWT reports a GDP of 120 and the other datasets report a GDP of 120 as well. The growth rate of the PWT will be 9% while the other datasets will report a growth rate of 17.6%. Because of the mistake from the previous year, the PWT will naturally report a lower growth rate the year following the mistake. I found a few other possible corrections among the growth rates of the three countries but the corrections were spread out over several years, which means it is unlikely

that it was a “natural” correction and instead the large differences are probably the results of the different methods that the datasets use.

I could only find this happening for 1969 and 1970 in India so it may just be that the methods of the PWT produced different growth rates than the other datasets for 1969 and 1970 and that they happened to look like they are somewhat balancing each other out. However, there is a possibility that something went wrong when the PWT compilers were calculating the growth rates for 1969. Jerven (2016) discusses something similar happening in Tanzania when there was a gap between two different versions of the official national accounts of Tanzania which somehow ended up with the PWT reporting a growth rate of -33% for 1988. He continues with describing how the PWT had reported a growth rate of 20%, a number that is not backed by any other evidence and is not even close to what other datasets are reporting. Finally, he concludes that it is likely that the PWT simply overestimated the growth rate for 1988 in Tanzania. The high growth rate that the PWT reported for Tanzania in 1987 could in turn have led to the low growth rate in 1988. Situations such as these enforce the view that cross-comparisons should be used when dealing with datasets such as these. The problem is not that things like this happen, everyone is bound to make mistakes after all, but rather that it is not brought up and explained. If problems such as these aren’t brought to light, someone might use the data to make incorrect assumptions. Jerven (2011a) shows an example by highlighting that the -33% growth rate that the PWT reported for Tanzania in 1988 was used as an example by Durlauf, Johnson, and Temple (2005) to argue that negative output shocks are typical of low-income countries.

Table 6, 7, and 8 show the average growth for different time periods across all four datasets for each country. The differences in the Philippines are not large enough to consider significant. India has small differences, but they are all below 1% and the largest difference is 0.95% in the period 1961-1965. As in the other exercises, Pakistan’s comparability is the worst out of the three countries, but there are no major differences and the largest difference in Pakistan is 1.15% in the period 1971-1975. It seems that averaging longer periods does well in reducing the impact that years with large differences have. For example, the largest difference between the datasets across all countries was found in Pakistan in 1965, in which there was a 6% difference between the Maddison dataset and the WDI. If we compare this with the difference between the Maddison average and the WDI average from 1961-1965, which is 1.11%, we can see the difference averaging can make. In the case of 1965, the

growth rate differed from a growth rate of 4.42%, which one could normally find in a developed country which is doing well, to a very high growth rate of 10.42% which is closer to what a rapidly industrializing country may experience.

**Table 6**: Growth Rates of Different Time Periods, the Philippines 1961-2012

| Philippines | Official | PWT | Maddison | WDI | Max | Min | Difference |
|---|---|---|---|---|---|---|---|
| 1961-1965 | 5,23% | 5,27% | 5,23% | 5,23% | 5,27% | 5,23% | 0,04% |
| 1966-1970 | 4,62% | 4,60% | 4,63% | 4,62% | 4,63% | 4,60% | 0,03% |
| 1971-1975 | 5,78% | 5,78% | 5,78% | 5,78% | 5,78% | 5,78% | 0,00% |
| 1976-1980 | 6,07% | 6,07% | 6,07% | 6,07% | 6,07% | 6,07% | 0,00% |
| 1981-1985 | -1,14% | -1,14% | -1,14% | -1,14% | -1,14% | -1,14% | 0,00% |
| 1986-1990 | 4,74% | 4,74% | 4,74% | 4,74% | 4,74% | 4,74% | 0,00% |
| 1991-1995 | 2,19% | 2,19% | 2,19% | 2,19% | 2,19% | 2,19% | 0,00% |
| 1996-2000 | 3,59% | 3,59% | 3,59% | 3,59% | 3,59% | 3,59% | 0,00% |
| 2001-2005 | 4,60% | 4,60% | 4,60% | 4,60% | 4,60% | 4,60% | 0,00% |
| 2006-2010 | 4,96% | 4,96% | 4,96% | 4,96% | 4,96% | 4,96% | 0,00% |

Sources: National accounts of the Philippines (2012, 2018); PWT (2015); The Maddison Dataset (2018); WDI (2017)

**Table 7**: Growth Rates of Different Time Periods, India 1961-2012

| India | Official | PWT | Maddison | WDI | Max | Min | Difference |
|---|---|---|---|---|---|---|---|
| 1961-1965 | 2,84% | 3,76% | 2,81% | 3,49% | 3,76% | 2,81% | 0,95% |
| 1966-1970 | 4,66% | 5,08% | 4,69% | 4,57% | 5,08% | 4,57% | 0,51% |
| 1971-1975 | 3,08% | 2,91% | 3,06% | 2,94% | 3,08% | 2,91% | 0,17% |
| 1976-1980 | 3,24% | 3,22% | 3,30% | 3,23% | 3,30% | 3,22% | 0,08% |
| 1981-1985 | 4,91% | 5,38% | 5,05% | 5,17% | 5,38% | 4,91% | 0,47% |
| 1986-1990 | 5,89% | 6,36% | 6,18% | 5,97% | 6,36% | 5,89% | 0,47% |
| 1991-1995 | 5,23% | 5,18% | 5,32% | 5,10% | 5,32% | 5,10% | 0,21% |
| 1996-2000 | 6,22% | 5,80% | 5,94% | 6,09% | 6,22% | 5,80% | 0,42% |
| 2001-2005 | 6,75% | 6,99% | 6,92% | 6,74% | 6,99% | 6,74% | 0,25% |
| 2006-2010 | 8,62% | 8,34% | 8,39% | 8,34% | 8,62% | 8,34% | 0,29% |

Sources: National accounts of India; PWT (2015); The Maddison Dataset (2018); WDI (2017)

**Table 8**: Growth Rates of Different Time Periods, India 1961-2012

| Pakistan | Official | PWT | Maddison | WDI | Max | Min | Difference |
|---|---|---|---|---|---|---|---|
| 1961-1965 | 6,79% | 7,12% | 6,32% | 7,43% | 7,43% | 6,32% | 1,11% |
| 1966-1970 | 6,74% | 7,24% | 7,15% | 7,06% | 7,24% | 6,74% | 0,50% |
| 1971-1975 | 4,34% | 3,62% | 3,19% | 3,22% | 4,34% | 3,19% | 1,15% |
| 1976-1980 | 5,34% | 5,98% | 6,28% | 6,23% | 6,28% | 5,34% | 0,95% |
| 1981-1985 | 6,69% | 6,58% | 7,00% | 6,78% | 7,00% | 6,58% | 0,42% |
| 1986-1990 | 5,60% | 5,80% | 5,60% | 5,80% | 5,80% | 5,60% | 0,20% |
| 1991-1995 | 4,90% | 4,65% | 4,61% | 4,65% | 4,90% | 4,61% | 0,29% |
| 1996-2000 | 3,98% | 3,27% | 2,97% | 3,27% | 3,98% | 2,97% | 1,01% |
| 2001-2005 | 5,25% | 5,02% | 5,44% | 5,02% | 5,44% | 5,02% | 0,42% |
| 2006-2010 | 4,22% | 3,43% | 3,08% | 3,43% | 4,22% | 3,08% | 1,14% |

Sources: National accounts of Pakistan; PWT (2015); The Maddison Dataset (2018); WDI (2017)

### 4.2.1 Discussion

The observations in the analysis raises the question of what should be considered “real” growth rates and what should be considered “artificial” growth rates after a dataset reports a substantially higher or lower growth rate for the same year compared to the other datasets, and then corrects it the following year? If datasets produce data that seem out of place without explaining the reasons or giving a way to examine the data used in the calculations, the only ways to determine whether outliers are caused by mistakes or if it is a coincidence would be to cross-check datasets, examine the primary data, and then compare methods. However, the datasets do not give access to the primary data, and they do not show exactly how they compile their own data. That leaves data users with cross comparing the datasets. A data user could make the assumption that if there’s one dataset that differ from the other 3 datasets, the lone dataset is more likely to be wrong. However, this method does not work if there’s a scenario where two pairs of datasets report different growth rates and even less so if all datasets reported growth rates that differ from each other. Of course, data users should also keep in mind that even if several datasets come to the same conclusion, it does not mean that their estimates are always accurate.

The large differences between the datasets make the comparability between the datasets low but averaging does seem to do well in reducing the impact of the years with a large difference among the datasets. Even when averaging however, cross-comparisons between the datasets should be actively used to determine how safe it is to use data from only one dataset. If there

is a consensus on the growth rates, it is more likely that the datasets are reporting numbers closer to the "true" growth rates.

Considering the Philippines is the richest in GDP per capita out of the three countries and has the best comparability while Pakistan is the poorest and has the worst comparability, it is possible that there is a correlation between a higher GDP per capita and a better comparability between the datasets. Jerven (2014) suggests that poorer countries will generally provide worse statistics since they will have less resources to spend on improving statistics. Then again, just because the resources are available, it does not mean that a government will prioritize statistics in their budget. If we examine how the comparability between the datasets have developed from 1961-2012, we can see in Table 6 that the differences have been shrunk over time in the Philippines. The development is harder to see in India but Table 7 shows that there is a general reduction in the differences. Table 8 suggests the opposite however. If we compare the period 1961-1986 to 1987-2012, the differences have even increased. Considering that two out of the three countries have experienced lower differences as they got richer, there is a possibility that there is a correlation between a higher GDP per capita and a better comparability. This is only what the results of the comparison between these three countries hint at however, and more countries have to be tested to determine whether this is the case or not across Asia.

Overall, the results of the growth rate comparisons for the three countries differ too much to be able to make a generalization to other Asian countries. What's interesting is that despite the fact that there should be differences between the datasets because they use different methods, the Philippines' growth rates match almost perfectly across the datasets. The differences that did show up are minor and mostly caused by the PWT. I would expect there to be differences between the other datasets as well, especially considering that the WDI use national prices from 2000 in their calculations while the PWT and the Maddison dataset use national prices from 2011 for their calculations. A possible reason behind why they match so well is that they all used the same series of GDP when computing growth rates and made minor adjustments to the data or filled any gaps in the data with their own estimates which resulted in the small differences that do exist. If that is the case, however, then Jerven (2016) was most likely correct when he suggested that the fact that datasets use their sources in different ways is the factor that contributes the most to the differences across the datasets.

Sadly, it can't be examined further since the datasets do not provide their source data and the ways in which they changed the primary data.

Furthermore, it is understandable that mistakes are made by the compilers of datasets but they should acknowledge and explain mistakes. The PWT does not seem to have acknowledged the -33% growth rate it reported for Tanzania in 1988 despite the fact that Jerven (2011a) discovered it more than a decade ago and despite the attention it has been given. Even if a dataset reports a very high or very low growth rate compared to other datasets and it is not caused by a mistake, and it has been given attention from data users, the dataset compilers should still try to explain the reasons behind why such a large difference exists.

The datasets are supposed to measure GDP for the same countries and the same periods, yet they differ several percentages in growth rates in some cases. The differences are minor for most years but the large differences that do exist have consequences for the comparability of the datasets. However, cross-comparing between the datasets seem to be a relatively fast way of finding potential mistakes. Additionally, by averaging longer time periods a data user should in most cases be able to find better agreement between the datasets. Furthermore, as suggested in the section that compares the GDP per capita data between the datasets, the compilers should provide easily accessible and easily understood metadata that makes it easier for data users to understand why differences exist and how the data is compiled.

## 4.3 Country Comparability

### 4.3.1 Comparison of the National Accounts of Australia and Nepal

To show how the process of statistics may differ between countries I will show some examples from the SNA of Nepal and Australia. It should highlight the differences in quality that data users cannot see when they compare countries in databases. Keep in mind that Nepal is one of the poorest countries in Asia according to GDP per capita while Australia is one of the richest, so the contrast will not be this large between all countries. It should also be noted that Australia's documents that explain their SNA are much more detailed than Nepal's and as such, Nepal's descriptions of their methods sometimes leave things unanswered. I will begin with going through the way the two countries approach timeliness, coherence, and accuracy. Afterwards, I will give a few examples of some problems that the countries face.

The Nepalese bureau of statistics (Nepal CBS, 2017) comments on their timeliness – which essentially means the delay between gathering the data and releasing the data – stating that the timeliness of their data is a problem because the analysis of both censuses and surveys are not completed on time. They explain that the causes for this are a lack of manpower as well as motivation to perform the analysis. This is in contrast to Australia bureau of statistics (Australia BS, 2016) where they write more about the balance between timeliness and accuracy, and that they rely on feedback from the users of the data to find the best combination of the two. They continue with giving detailed information on how long different kinds of estimates are usually made available after the reference period as well as noting that Australia national accounts rate well compared to other countries' timeliness.

There is also a large difference in coherence between the two countries. A high coherence means that comparable methods and definitions are used by data that measure the same phenomena. Nepal's bureau of statistics (Nepal CBS, 2017) writes that coherence is the main challenge they face. They continue, noting that the problem comes out of lack of coordination between the statistical office and connected agencies. They also mention that their lack of coherence has led to outputs that are not easy to interpret which in turn has resulted in confusion and mistrust in their statistics. Australia's (Australia BS, 2016) SNA does not seem to have any larger problems with coherence because they have followed international standards and definitions for statistics.

For most data users however, accuracy is the most important part of national accounts. Accuracy can be seen as how close to the true value an estimation is. For example, how close a GDP estimate is to the true GDP. Australia's bureau of statistics (Australia BS, 2016) writes that it is impossible to find an objective overall measure of the accuracy of national accounts since they draw data from so many sources which all have their own problems. To counter this, subjective assessments of accuracy are done on individual aspects of the national accounts. In practice, this means that the statistical offices in Australia grades categories such as GDP, household data, and government on a scale from A-D where A is the best and D is the worst. Overall, it seems that the Australian statistical offices take great care in improving the accuracy of their national accounts and explain problems they have. For example, they acknowledge that there are difficulties with measuring legal underground activity made to avoid taxation, but also explain that they estimate its value by regularly analyzing the individual components of GDP and other evidence from their taxation office. Whereas

Australia seems to put a lot of focus on accuracy in their estimates, Nepal's (Nepal CBS, 2017) bureau of statistics does not write much about their accuracy. They do however mention that it is measured by the data's validity and reliability. They write:

> The validity depends on many stages of the implementation of the census and surveys. They cannot be measured numerically. However; the reliability of the estimates depends on the standard errors of estimates. They have been measured in the poverty measures derived from NLSS [Nepal living standards survey] 2003-04 and 2010-11. The reliability of the data has been identified in the agriculture sample surveys in details. Other than these, it's hard to find the test of reliability for the survey estimates (Nepal CBS, 2017, p.282).

Another big problem for the statistics in Nepal (Nepal CBS, 2017) seems to be the quality of their administrative data. Their registration system does not have the required coverage to provide quality statistics and most of their registration and vital statistics system is run without modern technology. In contrast to this, Australia's national accounts (Australia BS, 2016) use administrative data for collecting information on the household sector, taxes, and finances. They also describe that they take extra efforts to make sure that the administrative data follows the quality requirements of the SNA.

The lack of demand for statistics is another topic that can be frequently seen in Nepal's SNA (Nepal CBS, 2017). Perhaps this is what has led to the lack of motivation from statistical staff members that is described as a big challenge in their SNA. For example, they write that the reason that the analysis of censuses and samples are often late is because of the lack of employees and motivation to perform the analysis. They later write that without an increased demand for statistics, it will be hard to improve statistics. To achieve that demand, they aim to increase statistical literacy among data users and people in general.

There is an apparent contrast between the two systems of national accounts. The problems seem to be large and many for the statistical system in Nepal and fewer and smaller for Australia's statistical system. Both the problems and the way the problems are dealt with in Australia's SNA are explained in detail but Nepal's descriptions often leave things up for interpretation. Moreover, there is a noticeable difference between the problems that the two countries experience. The problems that Australia writes about are either very specific or problems that exist in all countries, such as difficulties with measuring the informal economy.

Nepal (Nepal CBS, 2017) however, describes fundamental problems like a lack of motivation from their staff, lack of communication between different statistical agencies, and the fact that the traditional system of keeping hard copies of publications is widespread. With this in mind, can we really consider data from these two countries comparable? Although it is easy to forget, it is important to remember that data from economic databases are rarely of the same quality if they come from different countries. The international community should put more emphasis on enabling developing countries to catch up in statistical capacity since a better understanding of countries' economies will be beneficial to understanding economic growth. As we gain a better understanding of economic growth, research on global questions such as poverty and inequality will likely benefit as well.

### 4.3.2 General Comparison of National Accounts in Asia

Figure 4 and 5 shows the distribution of what base year and SNA the 43 countries included in the GDP per capita comparison section report using. The WDI does not have any data on what SNA Kyrgyzstan uses and Singapore reports using a mix of the 1993 and 2008 SNA. The base years and version of SNA that each country uses can be seen in Appendix A. When Statistical offices calculate national accounts they usually use a base year from which they gather prices. An alternative to using a base year is the chained constant price method which essentially means that instead of using the prices from a base year they use prices from the previous year. For example, using the chained constant price method to calculate national accounts for 2018 would be done using 2017 as the base year. Arguably, the chained constant price method provides a better representation of the economic reality in an economy compared to using a base year. Hence, most high-income countries use this method. However, using this method does not ensure high-quality statistics. Judging the comparability between the countries' statistics using only what base year and SNA they use is far from ideal but it should at least give an indication of how much resources, efforts, and time is being put into the statistics for the countries.

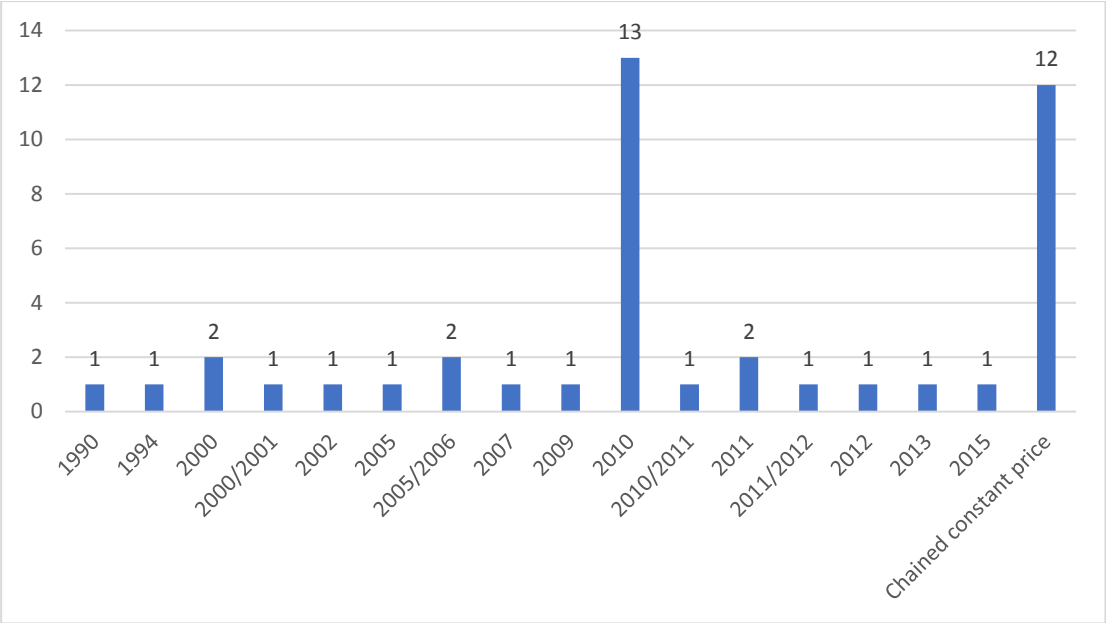


**Figure 4**: Asian Countries by Base Year

Source: WDI (2017)

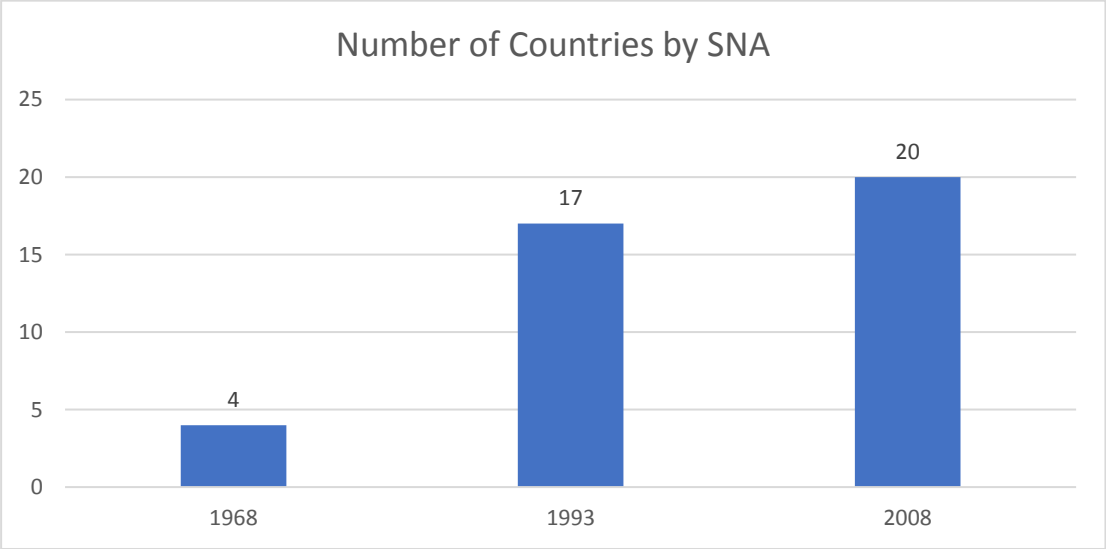


**Figure 5**: Asian Countries by Version of the System of National Accounts

Source: WDI (2017)

The way in which two countries perform their national accounts may differ substantially and one of the major factors in the quality of national accounting is what version of the SNA are being used. The United Nations (2018a) explains that the systems of national accounts are international standards of recommendations on how to perform national accounts. They note that new versions came up as the world economy changed and new methods were introduced:

the first version was released in 1953 and new versions have come out in 1968, 1993 and 2008. Best case scenario when it comes to comparability between countries would be if every country followed the same rules and standards but as Figure 4 and 5 shows, that is not the case.

Going through the difference between the versions of the SNA is outside of the scope for this paper but I will briefly explain the 1968 SNA, 1997 SNA, and 2008 SNA. None of the countries included in Figure 5 use the 1953 SNA and so, it will not be included. To get an indication of how much the coverage changed with each new version, you can look at how many pages each of the versions consists of. The 1953 SNA is 46 pages long, the 1968 SNA is 250 pages, the 1993 SNA is 700 pages, and the 2008 SNA is about 600 pages long but unlike the other versions, it has two instead of one column of texts per page (Lequiller & Blades, 2014, pp. 440-448). It is not hard to imagine that as the SNA develop and contain more and more information, they become harder for national accountants to follow. To combat the difficulties with the increasingly complicated SNA, the United Nations (2018b) started releasing handbooks of national accounting many years ago. If followed correctly, however, the national accounts of countries across the world should become more detailed and harmonized, resulting in better comparability between countries.

Lequiller and Blades (2014, pp.440-448) discuss the different version of the SNA. They note that the 1968 SNA was a major step forward in measuring economies. They note that the 1968 SNA focused more on international comparisons through for example constant prices and the limitations of using constant prices. They then move on to the 1993 SNA which was published by the several major international agencies whereas the 1968 SNA was published by the United Nations alone. To showcase how much more work went into the 1993 SNA compared to the 1968 SNA, they write that about 15 people wrote the 1968 SNA while more than 50 people took part in writing the 1993 SNA. They explain that the 1993 SNA took greater care to deal with issues that many countries faced when using the previous version. Zieschang (2012) from the IMF writes in a PowerPoint presentation made for the UN economic commission for Europe that the 2008 SNA did not include any fundamental changes and was mostly made to stay consistent with other statistical manuals, to keep up with the changing economy, to introduce improved methods, and to clarify prominent issues.

Out of the 43 listed countries, 12 use the chained constant price method and several of the remaining 31 countries that use base years are in grave need of an update. Both Yemen (1990)

and Jordan (1994) haven't updated their base year in more than two decades. Nepal (2000/01), Bangladesh (2005/06), Cambodia (2000), Pakistan (2005/06), the Philippines (2000), Iraq (2007), Thailand (2002), and Turkmenistan (2005) all have base years that are over a decade old. This means that out of the 31 countries that use a base year, 10 countries have base years that are more than a decade old and 21 countries have base years that are less than a decade old.

A country that has the most recent SNA from 2008 does not seem to mean that it also has a recently updated base year. Similarly, having an older SNA does not mean that a country has an outdated base year. For example, the Philippines uses the 2008 SNA but uses the year 2000 as a base year which means it has been almost two decades since an update. Another example is Myanmar which uses the 1968 SNA but its base year is from 2010/2011. Considering there are no fundamental differences between the 1993 SNA and the 2008 SNA (Zeischang, 2012) the comparability between them should be acceptable. Since 17 out of the 43 countries use the 1993 SNA while 20 out of the 43 countries use the 2008 SNA, the general comparability between the countries based on what version of the SNA they use is quite good. Kyrgyzstan does not report using a SNA and Iraq, Kuwait and Myanmar use the 1968 SNA which was designed to work for a world economy different from the economy we live in today. If we assume that the 1968 SNA is not comparable with the 1993 and 2008 SNA and remove the four countries that uses the 1968 SNA, we have 37 countries left that can be considered comparable.

One of the factors the World Bank's statistical capacity indicator uses to determine the score of a country's methodology is how recently their base year was updated. However, the indicator does not show the actual base year used. Instead, it gives countries a score of +1 if they use the annual chain linking method or if the base year was updated within the last 10 years, and if it has been more than 10 years the country receives a score of 0. If we apply that 10-year limit to Figure 4 and assume that countries whose base year are within 10 years of each other are comparable, we can determine how many countries that are comparable to the two large groups of countries that use the same base year in Figure 4. We can see that the group of 11 countries that use the chained constant price method is comparable with countries that use 2007 or later as their base year. If we exclude the three countries that use the 1968 SNA and Kyrgyzstan which does not report using an SNA, there are 39 countries left. This means that 26 out of the 39 countries are comparable to the group that uses the chain constant

price method. The group of 12 countries that use the base year 2010 is comparable with 37 out of the 39 countries. So, accounting for both the base year and the SNA, the comparability between the countries national accounts in Asia can be considered good.

If we are stricter and instead follow the IMF recommendation that base years should be updated every 5th year (Jerven, 2013a), we instead see that the countries that use 2010 as their base year and the countries that use the chained constant price method are no longer comparable. It should also be noted that, out of the 43 countries that were included from the beginning, only the 12 countries that use the chained constant price method plus Qatar (2013) and China (2015) comply to the IMF's recommendation. The group of countries with 2010 as their base year is then comparable with countries that have a base year anywhere from 2005-2015 which means 22 out of 39 countries. The group that uses the chained constant price method are comparable with countries with a base year from 2012-2022, which means 15 out of 39 countries. So, depending on whether we set the limit of comparability to 10 or 5 years difference between their base years, the general comparability of national accounts in Asia can be considered either good or bad. In my opinion, the 5-year limit is more appropriate since prices and economies can change a lot in a decade. Additionally, using a base year that hasn't been updated in 10 years will normally result in inaccuracies. If a country that is in the process of restructuring their economy or largely depend on the export of volatile exports uses a base year that is 10 years old, the inaccuracies will be even larger.

### 4.3.3 Discussion

There are several reasons as to why the general comparability of national accounts in countries can be low. First, conflicts such as wars can have consequences for a country's statistics. For example, if a country is suffering from a conflict – such as the still ongoing civil war in Myanmar or the 2003-2011 war in Iraq – it may have consequences for its data collection and one should be careful when using data from countries that have suffered or is suffering from conflicts that may have disturbed the data collection process. Second, some countries and especially developing countries, have large sectors of their economy that is not recorded. Examples are illegal economic activities or subsistence agriculture. For countries that do not have a lot of such activities, it usually is not a big problem as it wouldn't have had a large effect on the national accounts anyway. It becomes a problem in statistics since this kind of activity is larger in some countries and is hard to record and include in national

accounts. In practice, this means that countries that do not have the ability and resources to make somewhat accurate estimates of its unrecorded activity are usually undervalued and therefore their national accounts lose comparability with other national accounts. Third, politics seem to have an effect on statistics. For example, Magee and Doces (2014) have concluded that authoritarian regimes are prone to exaggerating their growth rates. Additionally, Geda and Yimer (2014) analyzed the political and economic climate in Ethiopia and compared estimates from international data sources with Ethiopia's reported growth rates and came to the conclusion that the official numbers are probably heavily overstated. Fourth, the general comparability of countries' national accounts is also a question of how the data is collected. For example, Eurostat (2017) mentions that high-income countries are more likely to keep administrative data compared to low-income countries. Eurostat also mentions that administrative data is an important part in national accounts and therefore, a lack of such data is likely to cripple national accountants' ability to produce quality statistics. Finally, Carr-Hill (2014) explains how the poorest of the poor are normally excluded from household surveys, adding to the problems of accuracy in developing countries. Therefore, these are things that data users have to be informed about if they hope to accurately judge a country's national accounts. However, it is unreasonable to expect data users to be country experts for them to be able to effectively use the data. One solution is to use the World Bank's statistical capacity indicator, which creates estimates of countries statistical capacity by rating its methodology, source data, and periodicity. However, its way of measuring the different factors is unsatisfactory due to its simplicity and it is rigid in taking its measurements rather than adaptive. Furthermore, it does not include high-income countries.

A solution that could change the way we view statistics is if a measurement that covers all countries and is more extensive than the World Bank's statistical capacity indicator was created. For example, instead of giving countries a higher score based on whether or not they have a base year that has been updated in the last 10 years, it could gradually decrease how much points are given as more time passes since the base year was updated. Furthermore, as a modifier to the base year score, the indicator could also analyze how much countries are affected by price changes and change the score accordingly. For instance, if a country relies on exports with volatile prices, it should update its base year more often than a country that relies on exports with stable prices. Another change that could be made is that it could show a score for different kinds of statistics. Instead of showing scores for methodology, source data, and periodicity like it does now, it could also show one score for economic data, one for

environmental data, one for health data, and so on. Another idea is to use the method that Magee and Doces (2014) used to estimate how much authoritarian regimes overstated their growth rates. They used nighttime satellite images to measure how many new lights have shown up, and then compared the annual light growth to the annual GDP growth. An indicator that could accurately measure these factors could be great for the world of statistics. To add to this, international organizations should also provide more assistance to help countries with poor statistical capacity improve their data.

The aim of this comparison was to get an overview of how comparable Asian countries' national accounts are. Figure 4 and 5 have shown the distribution of what base year and what version of the SNA the countries are using but it is not enough to give a definite answer of the level of comparability of the countries national accounts. However, considering that both the base year and the SNA plays a crucial role in the quality of a country's national accounts, it should work as an initial indicator. Furthermore, neither the World Bank's statistical capacity indicator or Figure 4 and 5 shows the quality of the data collected for each country. To get a full assessment of the comparability between countries' national accounts, data users would have to examine the countries' SNA. Only then can a data user look at each country's methods, what problems the statistical office encountered, how problems were solved and to what degree they were solved, and so on. So, when data users are using the national accounts of a country or data that have been aggregated by using national accounts – such as the WDI, the PWT, and the Maddison dataset – I recommend inspecting the SNA of each country that the data user is examining to get an idea of the quality of the data. If that is not possible due to a high number of countries or lack of time, data users should at least utilize the World Bank's statistical capacity indicator. If countries that are not included in the statistical capacity indicator are being examined, data users can instead perform the same comparison as the ones in figure 4 and 5 by downloading the data from the WDI. Another solution that could help data users is if the World Bank's statistical capacity indicator's coverage and measurements were improved.

## 4.4 Summarizing Discussion

The thesis has found that with the exceptions of Uzbekistan and Turkmenistan, the rankings of the three major international datasets according to GDP per capita are comparable. When

the actual GDP per capita numbers were compared however, several countries with relatively large differences between the datasets were found. Therefore, GDP per capita data should always be cross-checked between several sources. The majority of countries do not have major differences however, and therefore the comparability between the datasets can be considered quite good. It should be noted that only data from 2010 were tested and that other years have to be tested to see if the results are similar for different years.

It was also found that the comparability of growth rates across the datasets for the Philippines, India, and Pakistan was very different depending on what country was examined. The Philippines' growth rates matched almost perfectly across the datasets. India's growth rates did not match as well as the Philippines, there were a few large differences but most other differences were below 1.00%. Pakistan's growth rates matched worse than the others and the overall difference was larger compared to the other countries. For Pakistan, there were several years that had a difference large enough for data users to come to different conclusions depending on what dataset is used. The largest difference between the datasets across all countries is found in 1965 for Pakistan, a difference of about 6% between the Maddison dataset and the other three datasets. Similar to the GDP per capita section, the overall comparability is quite good but the cases where large differences can be found is problematic for the comparability between the datasets. Just as with GDP per capita comparisons, data users should cross-compare the datasets when it comes to GDP growth rates as well. Moreover, as the analysis found, a useful tool for data users can be to average the growth rates of longer periods instead of examining individual years since it seems to do well in reducing the impact that years with large differences have. Finally, since the comparability between the datasets differs so much between the countries, no generalizations to countries in Asia should be made from these results without examining more countries. With that said, the results do hint that a higher comparability of growth rates between the datasets goes hand in hand with a higher GDP per capita but again, more countries have to be tested.

For GDP comparisons, it would be easier for data users to determine the quality of the estimations from different datasets if the datasets showed more transparency. For example, if there is an outlier in the datasets, and it has been given attention by data users, it should still be acknowledged and explained by the dataset. Furthermore, not being able to look at the source data (Jerven, 2013a) and not having quality metadata (Jerven, 2012) is limiting when it comes to understanding the data. A possible solution could be for the international datasets to

provide their source data as well as online instructions and examples to enable data users to gain a better understanding of the data.

The analysis of comparability between countries' national accounts found that 4 out of the 43 countries are not comparable because of what SNA they use and the WDI does not have any data on what SNA Kyrgyzstan uses. Additionally, the analysis found that out of the remaining 38 countries, there are two large groups of countries that use the same base year. The first group consists of the 12 countries that use 2010 as their base year and the second group consists of the 11 countries that use the constant chained price method. Using 5 years difference between countries' base years as a limit to when countries' national accounts are comparable, it was found that the two groups of countries are not comparable with each other. With the IMFs' recommendation of updating base years every 5$^{th}$ year, the overall statistical situation in Asia is not looking good in terms of what base years are used. Similarly, the overall comparability between the national accounts is poor, except for the group that uses 2010 as their base year which can be considered comparable with 25 out of the 43 countries. Data users should therefore at the very least compare countries' base years and SNA versions when comparing data from two countries. Additionally, data users should use the World Bank's statistical capacity indicator to get another indicator of countries comparability. Considering the difference that could be seen when comparing the SNA of Australia and Nepal, best practice when comparing GDP data that use national accounts as their primary source is to examine the SNA of the countries the data originally come from. To take the responsibility of quality checks away from the data user and put it in the hands of international institutions, an extensive upgrade to the World Bank's statistical capacity indicator could be developed. The improved statistical indicator could then be used together with datasets to determine what countries' statistics are comparable. This would also help international organizations pinpoint which countries are in most need of improving their statistical capacity.

I've compared the results from my analysis to analyses done on African statistics with similar methods by Jerven (2011a, 2013a, 2013b), and it is apparent that Asian countries are indeed suffering from the same statistical problems as African countries are, but to a smaller extent. For example, the average change of rank between the rankings of GDP per capita between the Maddison dataset and the WDI was 7 ranks. For Asian countries, it was 1.2 ranks. Another example is that whereas Jerven (2011a) showed that Tanzania's growth rates had an average

difference of 6% between the highest and lowest reported growth rate for the years 1961-2001, the Philippines, India, and Pakistan had average differences of 0.03%, 1.00%, and 1.47% respectively for the years 1961-2012. More research should be done on the topic of statistics, and especially on specific countries, to determine to what degree the statistics from each country can be trusted.

# 5 Conclusion

While African countries have been the focus of most literature on the topic of statistical problems, this thesis has examined how the statistical situation looks in Asia. By comparing different datasets and their reported GDP per capita and GDP growth rates for different countries, the thesis has found that the comparability between the datasets is good overall but some cases show substantial differences between the datasets. Therefore, cross-comparisons should always be used to locate these cases. Furthermore, averaging longer periods of growth is recommended to reduce the impact that these cases have. Data users then have to make their own assessment of whether the differences are at acceptable levels or not. To help data users fully understand the data, dataset compilers should be more transparent with their methods and provide the source material they use. It should also be noted that to be sure of the results, similar methods should be applied to more countries and different years.

The analysis of the countries' SNA showed that the national accounts of Asian countries should not be considered comparable on a general level. There are however tools that data users can use to get an indication of what countries are at comparable levels of statistical capacity. First, a data user can use the World Bank's statistical capacity indicator. Second, by using data from the WDI, a data user can quickly compare countries' base years and versions of the SNA. Third, the data user can examine and compare the different methods that countries use in their SNA. While the first two options should only be used as initial indicators, the third option can be time consuming and is impractical when comparing many countries. Therefore, I suggest a transfer of the responsibility of quality checks from the data users to international institutions. This could be done by upgrading the methods and coverage of the World Bank's statistical capacity indicator.

With the data analysis in mind, the overall comparability of the GDP data between the datasets can be seen as comparable only on a general level. This is because large differences exist between the datasets as well as between the quality of countries' estimates. Furthermore, data users should be aware that although data are presented in the same way for different countries, they may not have the same level of quality.

Finally, three things should be emphasized. First, more research is required on the subject of comparability of statistics. Second, international organizations should focus more on helping all countries reach the same level of statistical capacity. Third, although well-known organizations give credibility to data, all data users must stay critical of sources regardless of their origin.

# References


Australian Bureau of Statistics, (2016) 'Australian System of National Accounts. Concepts, Sources and Methods' [pdf] Available online: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/C5ACA29422243B56CA257F7D00177D09/$File/52160_2015_.pdf [Accessed 15 May 2018]

Black, J. Hashimzade, N. Myles, G. (2017) 'A Dictionary of Economics' 5th Edition. Oxford University Press. Available online: http://www.oxfordreference.com.ludwig.lub.lu.se/view/10.1093/acref/9780198759430.001.0001/acref-9780198759430-e-1138?rskey=sbvCZ3&result=1212 [Accessed 28 May 2018]

Carr-Hill, R. (2012), 'Missing Millions and Measuring Progress towards the Millenium Development Goals with a focus on Central Asian States' *Central Asian Journal of Global Health*, Vol. 1, No. 1. Available online: https://cajgh.pitt.edu/ojs/index.php/cajgh/article/viewFile/24/16 [Accessed 28 May 2018]

Carr-Hill, R. (2014) 'Measuring Development Progress in Africa: The Denominator Problem' *Canadian Journal of Development Studies*, 35:1, 136-154. Available online: https://www-tandfonline-com.ludwig.lub.lu.se/doi/abs/10.1080/02255189.2014.884969 [Accessed 2 May 2018]

Central Bureau of Statistics, (2017). 'A compendium of National Statistical System of Nepal. Kathmandu, Nepal: Central Bureau of Statistics.' [pdf] Available online: http://cbs.gov.np/image/data/2017/National%20Statistical%20System%20of%20Nepal_FINAL.pdf [Accessed 15 May 2018]

Collier, P. (2007) 'The Bottom Billion: Why the Poorest Countries Are Failing and What Can Be Done About It.' Oxford: Oxford University Press

Deaton, A. & Heston, A. (2010) 'Understanding PPPs and PPP-based National Accounts' *American Economic Journal: Macroeconomics*, Vol. 2, No. 4, pp. 1-35. Available online: https://scholar.princeton.edu/deaton/publications/understanding-ppps-and-ppp-based-national-accounts [Accessed 11 May 2018]

Devarajan, S. (2013) 'Africa's statistical tragedy', *Review of Income and Wealth*, Vol. 59, Special Issue, Series 59, pp.9-15. Available online: https://onlinelibrary-wiley-com.ludwig.lub.lu.se/doi/abs/10.1111/roiw.12013 [Accessed 28 May 2018]

Durlauf, S., P. Johnson and J. Temple (2005) 'Growth econometrics' *Handbook of Economic Growth*, vol. 1, part A, pp. 555–667. Available online: https://econpapers.repec.org/bookchap/eeegrochp/1-08.htm [Accessed 28 May 2018]

Eurostat (2017) 'Building the System of National Accounts – Administrative Sources' [pdf] Available online: http://ec.europa.eu/eurostat/statistics-explained/pdfscache/13318.pdf [Accessed 25 May 2018]

Feenstra R. Inklaar R. Timmer M. (2015) 'The Next Generation of the Penn World Table' *American Economic Review*, Vol. 105, No. 10, p. 3150-3182, Available online: www.ggdc.net/pwt [Accessed 10 May 2018]

Fioramonti, L. (2014) 'The World's Most Powerful Number: An Assessment of 80 Years of GDP Ideology' [pdf] *Anthropology Today* Vol. 30, No. 2, pp. 12-15. Available online: https://onlinelibrary-wiley-com.ludwig.lub.lu.se/doi/epdf/10.1111/1467-8322.12098 [Accessed 27 May 2018]

Geda, A. and Yimer, A. (2014) 'The Political Economy of Growth, Poverty and Inequality in Ethiopia, 2000-2013' in *Reflections on Development in Ethiopia: New Trends, Sustainability and Challenges.* Forum for Social Studies, pp. 31-66. Available online: https://www.researchgate.net/publication/264004552_The_Political_Economy_of_Growth_Poverty_and_Inequality_in_Ethiopia_2000-2013 [Accessed 27 May 2018]

Heston, A. (1994) 'A Brief Review of Some Problems in Using National Accounts Data in Level of Output Comparisons and Growth Studies', *Journal of Development Economics*, Vol. 44, Issue. 1, pp. 29-52. Available online: https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/pii/0304387894000042 [Accessed 15 May 2018]

India Ministry of Statistics and Programme Implementation (n.d) 'Summary of Macro Economic Aggregates at Constant (2004-05) Prices, 1950-51 To 2013-14' Available online: http://www.mospi.gov.in/data [Accessed 11 May 2018]

Jerven, M. & Johnston, D. (2015) 'Statistical Tragedy in Africa? Evaluating the Data Base for African Economic Development', *The Journal of Development Studies*, Vol. 51, No. 2, pp. 111-115. Available online: https://www-tandfonline-com.ludwig.lub.lu.se/doi/abs/10.1080/00220388.2014.968141 [Accessed 15 May 2018]

Jerven, M. (2010) 'The Relativity of Poverty and Income: How Reliable are African Economic Statistics?' *African Affairs*, Vol. 109, No. 434, pp. 77-96. Available online: https://academic-oup-com.ludwig.lub.lu.se/afraf/article/109/434/77/72195 [Accessed 28 May 2018]

Jerven, M. (2011a) 'Growth, Stagnation or Retrogression? On the Accuracy of Economic Observations, Tanzania, 1961-2001' *Journal of African Economies*, Vol. 20, No. 3, pp.377-394. Available online: https://academic-oup-com.ludwig.lub.lu.se/jae/article/20/3/377/694998 [Accessed 2 May 2018]

Jerven, M. (2011b) 'Counting the Bottom Billion: Measuring the Wealth and Progress of African Economies', *World Economics, Vol*. 12, Issue. 4, pp. 35-52. Available online:

https://www.worldeconomics.com/Journal/Papers/Counting%20the%20Bottom%20Billion.details?ID=494 [Accessed 7 May 2018]

Jerven, M. (2011c) 'Users and Producers of African Income: Measuring African Progress', *African Affairs* Vol. 110, Issue. 439, pp. 169–190. Available online: https://academic-oup-com.ludwig.lub.lu.se/afraf/article/110/439/169/163787 [Accessed 20 May 2018]

Jerven, M. (2012) 'Poor Numbers! What We Know About Income and Growth in Sub-Saharan Africa' [pdf] Ithaca: Cornell University Press. Available online: https://www.cgdev.org/doc/17-NOV-CGD-Poor%20Numbers-Jerven.pdf [Accessed 14 March 2018]

Jerven, M. (2013a) 'Comparability of GDP Estimates in Sub-Saharan Africa: The Effect of Revisions in Sources and Methods Since Structural Adjustment' The Review of Income and Wealth, Series 59, Special Issue, pp. 16-36. Available online: https://onlinelibrary-wiley-com.ludwig.lub.lu.se/doi/abs/10.1111/roiw.12006 [Accessed 10 May 2018]

Jerven, M. (2013b) 'Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It.' Cornell University Press.

Jerven, M. (2013c) 'Briefing: for Richer, For Poorer: GDP Revisions and Africa's Statistical Tragedy' *African Affairs*, Vol. 112, Issue. 446, 2013, pp. 138-147. Available online: https://academic-oup-com.ludwig.lub.lu.se/afraf/article/112/446/138/10183 [Accessed 7 May 2018]

Jerven, M. (2014) 'Measuring African Development: Past and Present. Introduction to the Special Issue', *Canadian Journal of Development Studies*, Vol 35, Issue. 1, pp. 1-8. Available online: https://www-tandfonline-com.ludwig.lub.lu.se/doi/abs/10.1080/02255189.2014.876617 [Accessed 7 May 2018]

Jerven, M. (2016) 'Discrepancies: Why Do GDP Growth Rates Differ?" *Review of Agrarian Studies*, vol. 6, no. 1. Available online: http://www.ras.org.in/discrepancies_why_do_GDP_growth_rates_differ [Accessed 4 May 2018]

Kiregyera, B. (2013), 'The Dawning of a Statistical Renaissance in Africa' Conference on African Economic Development: Measuring Success and Failure.' [pdf]Vancouver. Available online: http://mortenjerven.com/wp-content/uploads/2013/04/AED_Panel_8-Kiregyera.pdf [Accessed 14 March 2018]

Lequiller, F. & D. Blades (2014), 'Understanding National Accounts: Second Edition', OECD Publishing. Available online: http://dx.doi.org/10.1787/9789264214637-en [Accessed 11 May 2018]

Maddison Project Database version 2018. Bolt, Jutta, Robert Inklaar, Herman de Jong and Jan Luiten van Zanden (2018), ”Rebasing ’Maddison’: New Income Comparisons and the Shape of Long-Run Economic Development”, Maddison Project Working paper 10. Available online: https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2018 [Accessed 10 May 2018]

Bolt, Jutta, Robert Inklaar, Herman de Jong and Jan Luiten van Zanden (2018) ’Rebasing ’Maddison’: New Income Comparisons and the Shape of Long-run Economic Development’ [pdf] Available online: https://www.rug.nl/ggdc/html_publications/memorandum/gd174.pdf [Accessed 25 May 2018]

Magee Christopher S. P. and John A. Doces (2014) ‘Reconsidering Regime Type and Growth: Lies, Dictatorships, and Statistics. *International Studies Quartely*. Vol. 59, Issue 2, pp. 223-237. Available online: https://onlinelibrary-wiley-com.ludwig.lub.lu.se/doi/abs/10.1111/isqu.12143 [Accessed 25 May 2018]

Mudde, C & Schedler, A. (2010) ‘Introduction: Rational Data Choice.’ *Political Research Quarterly*, Vol.63, Issue 2, pp. 410–416. Available online: https://works.bepress.com/cas_mudde/37/ [Accessed 25 May 2018]

OECD (2018) ‘Purchasing Power Parities – Frequently Asked Questions (FAQs)’ Available online: http://www.oecd.org/sdd/prices-ppp/purchasingpowerparities-frequentlyaskedquestionsfaqs.htm [Accessed 26 May 2018]

OECD Observer (2005) ‘Is GDP a Satisfactory Measure of Growth?’ Available online: http://oecdobserver.org/news/archivestory.php/aid/1518/Is_GDP_a_satisfactory_measure_of_growth_.html [Accessed 21 May 2018]

Pakistan Bureau of Statistics (n.d) ‘National Accounts Tables’ [pdf] Available online: http://www.pbs.gov.pk/sites/default/files/national_accounts/tables/table1.pdf [Accessed 1 May 2018]

Philippine Statistics Authority, (2012), ‘ANNUAL National Accounts Linked Series (1946 - 2010)’ Available online: http://psa.gov.ph/nap-press-release/data-series [Accessed 3 May 2018]

Philippine Statistics Authority, (2018), ‘ANNUAL National Accounts (1998 - 2017)’ Available online: http://psa.gov.ph/nap-press-release/data-series [Accessed 8 May 2018]

Pilling, D. (2018) ‘The Growth Delusion: How the World is Wrongly Obsessed by GDP’ [Youtube] The Foreign Correspondents’ Club, Hong Kong. Available online: https://www.youtube.com/watch?v=_QHnO1k-tgo [Accessed 21 May 2018]

Powell, NP. (1981) ‘Major Obstacles to Achieving Satisfactory Registration of Vital Events and the Compilation of Reliable Vital Statistics.’ [pdf] Technical papers of the International Institute for Vital Registration and Statistics, No 15. Bethesda (MD): International Institute

for Vital Registration and Statistics. Available online: https://www.cdc.gov/nchs/data/isp/15_major-obstacles-to-achieving-satisfactory-regist-of-vital-events---compilations-of-reliable-vital-stat_02.pdf [Accessed 19 May 2018]

Seers, D. (1952) 'The Role of National Income Estimates in the Statistical Policy of an Underdeveloped Area', *The Review of Economic Studies*, Vol. 20, No. 3, pp. 159-168. Available online: https://academic-oup-com.ludwig.lub.lu.se/restud/article-abstract/20/3/159/1597496 [Accessed 28 May 2018]

Setel P. Macfarlane S. Szreter S. Mikkelsen L. Jha P. Stout S. AbouZahr C. (2007) 'A Scandal of Invisibility: Making Everyone Count by Counting Everyone' The Lancet. Vol. 386, Issue 9598, pp. 1569-1577. Available online: https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/pii/S0140673607613075?via%3Dihub [Accessed 17 May 2018]

Srinivasan T.N. (1994) 'Data Base for Development Analysis: An overview' *Journal of Development Economics*, Vol. 44, Issue 1, pp.3-27. Available online: https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/pii/0304387894000034 [Accessed 15 May 2018]

The Economist (2014) 'Revised figures show that Nigeria is Africa's largest economy' Available online: https://www.economist.com/news/finance-and-economics/21600734-revised-figures-show-nigeria-africas-largest-economy-step-change [Accessed 20 May 2018]

United Nations (2018a) 'Updating the System of National Accounts (SNA)' Available online: https://unstats.un.org/unsd/nationalaccount/snaUpdate.asp [Accessed 11 May 2018]

United Nations (2018b) 'National Accounts Publications' Available online: https://unstats.un.org/unsd/nationalaccount/pubs.asp [Accessed 11 May 2018]

Woods, D. (2014) 'The Use, Abuse, and Omertà on the "Noise" in the Data: African Development, Democratization, and Urban Bias.' *Canadian Journal of Development Studies* 35: 120-135. Available online: https://www-tandfonline-com.ludwig.lub.lu.se/doi/abs/10.1080/02255189.2014.873021 [Accessed 25 May 2018]

World Bank (2012) 'Getting to Know the World Bank' Available online: http://www.worldbank.org/en/news/feature/2012/07/26/getting_to_know_theworldbank [Accessed 14 March 2018]

World Bank (2017) 'World Development Indicators'. Available online: http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators [Accessed 14 March 2018]

World Bank (2018a) 'Is all the WDI data based on calendar year or fiscal year reporting period' Available online: https://datahelpdesk.worldbank.org/knowledgebase/articles/201203-is-all-the-wdi-data-based-on-calendar-year-or-fisc [Accessed 23 May 2018]

World Bank (2018b) 'Knowledge Base' Available online: https://datahelpdesk.worldbank.org/#knowledgebase [Accessed 28 May 2018]

Young, A. (2010) 'The African Growth Miracle' [pdf] *Journal of Political Economy*, Vol. 120, No. 4, pp. 696-739. Available online: http://personal.lse.ac.uk/YoungA/TheAfricanGrowthMiracle.pdf [Accessed 25 April 2018]

Zieschang (2012) 'SNA 2008: An Essential Tool for Economic Policy and Monitoring' Published by United Nations Economic Commission for Europe. Powerpoint available for download: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2012/Introduction_of_the_2008_SNA_for_Users.pptx [Accessed 11 May 2018]

# Appendix A

Table 9
Base year and system of national accounts currently in use as reported by the WDI

| Country | Base year | SNA |
|---|---|---|
| Yemen | 1990 | 1993 |
| Jordan | 1994 | 1968 |
| Cambodia | 2000 | 1993 |
| Philippines | 2000 | 2008 |
| Nepal | 2000/2001 | 1993 |
| Thailand | 2002 | 2008 |
| Turkmenistan | 2005 | 1993 |
| Bangladesh | 2005/2006 | 1993 |
| Pakistan | 2005/2006 | 2008 |
| Iraq | 2007 | 1968 |
| Turkey | 2009 | 2008 |
| Vietnam | 2010 | 1993 |
| Mongolia | 2010 | 2008 |
| Indonesia | 2010 | 2008 |
| Sri Lanka | 2010 | 2008 |
| Lebanon | 2010 | 2008 |
| Malaysia | 2010 | 2008 |
| South Korea | 2010 | 2008 |
| Bahrain | 2010 | 1993 |
| Saudi Arabia | 2010 | 2008 |
| Oman | 2010 | 1993 |
| United Arab Emirates | 2010 | 1993 |
| Kuwait | 2010 | 1968 |
| Singapore | 2010 | 1993/2008 |
| Iran | 2011 | 1993 |
| Myanmar | 2010/2011 | 1968 |
| Russian | 2011 | 1993 |
| India | 2011/2012 | 2008 |
| Lao People's DR | 2012 | 2008 |
| Qatar | 2013 | 1993 |
| China | 2015 | 2008 |
| Tajikistan | Chained constant price | 1993 |
| Kyrgyzstan | Chained constant price | - |
| Uzbekistan | Chained constant price | 1993 |
| Armenia | Chained constant price | 2008 |
| Georgia | Chained constant price | 1993 |
| Azerbaijan | Chained constant price | 1993 |
| Kazakhstan | Chained constant price | 1993 |
| Israel | Chained constant price | 2008 |

| | | |
|---|---|---|
| New Zealand | Chained constant price | 2008 |
| Japan | Chained constant price | 2008 |
| Australia | Chained constant price | 2008 |
| China, Hong Kong | Chained constant price | 2008 |

Source: WDI (2017)