The University of Queensland

Faculty of Business, Economics and Law

School of Economics

MEASUREMENTS OF TRUE INCOME: DO PWT, UQICD AND WDI BEHAVE SIMILARLY?

Hoang Diep PHAN

16 November, 2016

Word Count: 20,000

Abstract

Internationally comparable income is obtained using purchasing power parity (PPP) adjustments over space and time across countries. The method used to adjust can lead to differences in the PPP adjusted income. National accounts GDP in local currency can be converted to PPP adjusted income, and some authors in the poverty measurement literature have suggested using income obtained from household surveys. In their recent QJE paper, Pinkovskiy and Sala-i-Martin (2016a) (P&Sa) proposed a combination method whereby optimum weights are given to GDP and surveys' based income. They used World Development Indicators (WDI)'s GDP income for their study. This thesis extends their work in three directions: First, the robustness of the findings is tested by using four alternative PPP adjusted GDP income to WDI, they are PWT 7.1, PWT 8.1, PWT 9.0 and UQICDv2.1.1. This also extends the recent NBER paper (P&Sb) which evaluated WDI income against PWT 7.1, PWT 8.0 and PWT 8.1. Second, the robustness of the findings is tested by running the models by income and geographical country groups. Third, the robustness of the findings is tested by using an alternative econometric estimation approach to obtain the optimum weights.

Table of Contents

List of figures and tables

List of figures

List of tables

List of Abbreviations

PWT	Penn World Table
WDI	World Development Indicator
UQICD	The University of Queensland International Comparison Database
ICP	International Comparison Program
WB	World Bank
GDP	Gross Domestic Product
PPP	Purchasing Power Parity
P&S	Pinkovskiy and Sala-i-Martin
P&Sa	Pinkovskiy and Sala-i-Martin (2016a)
P&Sb	Pinkovskiy and Sala-i-Martin (2016b)
P&Sa, b	Pinkovskiy and Sala-i-Martin (2016a, b)
NTL	Nighttime Lights
IPI	Integrated Poverty Index
ALI	Average Light Index

Chapter 1 Introduction

1.1 Motivations

Measuring income per capita accurately is a challenging task. We now have various sources of data on comparable income across countries. The Penn World Table (PWT) and World Development Indicator (WDI) are considered the most popular income datasets. These datasets are the most frequently used in cross country empirical studies by many scholars, independent organizations or government agencies all around the world. Recently, the University of Queensland has developed a new source of internationally comparable income per capita which is called The University of Queensland International Comparison Database (UQICD). Although different datasets or even different versions of the same datasets (for example, different versions of PWT) use different methodologies to construct income data, they are all based on primary data collected under the International Comparison Program (ICP) of the United Nations statistical office, and managed by the World Bank (WB). However, the problem is that these sources of data show a high degree of inconsistency. For example, according to PWT 7.1, Gross Domestic Product (GDP) per capita in international Purchasing Power Parity (PPP) dollar, at 2005 constant prices of Ghana in 2010 is $2094. Yet, from PWT 8.1, PWT 9.0, WDI (2012), and UQICDv2.1.1 datasets, the same numbers are $2952, $2691, $1478 and $1560 respectively[1]. The last two ICP rounds have been conducted in 2005 and 2011. The PWT 7.1 and PWT 8.1 do not include ICP 2011. All others include 2011 ICP in the construction of PPP adjusted income. Table 1.1 presents a list of some countries which have large differences in real GDP per capita in the year 2005 from four sources of GDP. We can immediately see the differences for the same country and same year. These differences have considerable impact on research results. For example, Ciccone and Jarociski (2010) found that using different versions of PWT results considerably changed the validity of many growth determinants such as the role of government or international trade. Hence, if you are a researcher or policy maker, you must answer the three following questions

1. Which source of data do you use to conduct a study or make a decision?

2. Which source of data is the best in capturing true income per capita?

[1] PWT data are downloaded from http://www.rug.nl/research/ggdc/data/pwt/. UQICD data are downloaded from http://uqicd.economics.uq.edu.au/data.php. WDI data are taken from the paper of Pinkovskiy and Sala-i-Martin (2016a), however, these are available from the World Bank.

3. How do different datasets behave in a particular research?

Table 1.1: Countries with large differences in PPP real GDP per capita 2005 reported in WDI (2012), PWT 7.1, PWT 8.1 and UQICD

Country	Large difference			
	WDI (2012)	PWT 7.1	PWT 8.1	UQICD
Turkmenistan	4,762	10,348	8,507	-
Tonga	4,139	7,640	-	4,020
Samoa	3,831	6,541	-	3990
Kiribati	2,342	3,830	-	1960
Barbados	17,965	28,575	23,991	17,700
Grenada	10,031	15,890	7,853	10,500
Suriname	6,128	9,659	5,702	-
Vanuatu	3,500	5,331	-	3,290
Guyana	2,536	3,727	-	2,930
Belize	6,254	8,981	7,034	7,050
Guatemala	4,062	5,589	3,683	5,070
Ghana	1,208	1,658	2,174	1,280
Thailand	6,675	6,966	8,335	6,980
Vietnam	2,161	2,088	2,525	2,330
Malaysia	12,011	10,482	14,387	12,400

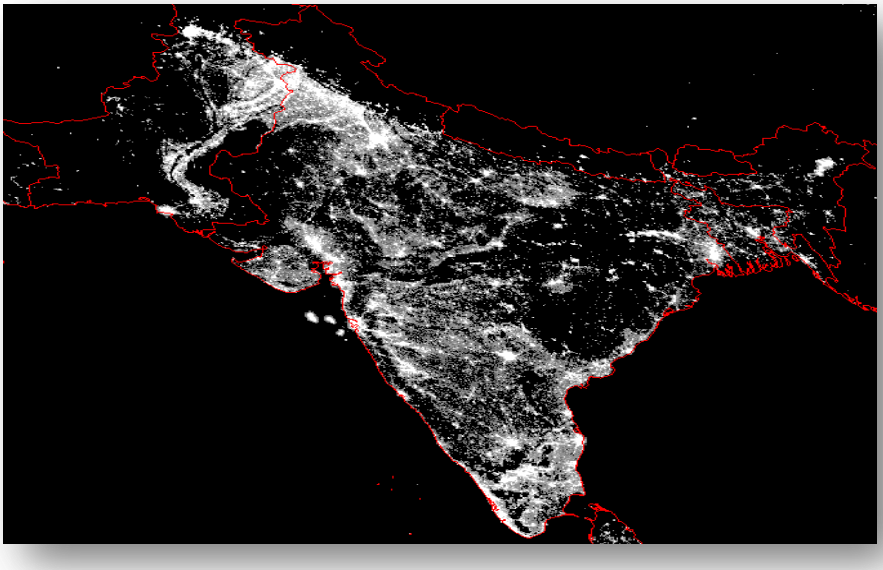
(Source: adapted from Ram and Ural 2014 , page 641)

To find a key to answer the first question, we should first try to answer the second and the third questions. In an attempt to answer the second question, relevant findings were achieved by Pinkovskiy and Sala-i-Martin (2016b) (P&Sb). They used satellite-recorded nighttime light (NTL) data to evaluate the performance of different series of GDP estimates. They assumed that lights as a measure of income have measurement errors uncorrelated with measurement errors of different international income data. They found that WDI is often better in explaining true income than PWT, while PWT 7.1 series appear to perform better than PWT 8.0 and PWT 8.1. The comparison made in P&Sb is between WDI based on ICP 2011, NTL and PWT 7.1, PWT 8 and PWT 8.1 are based on ICP 2005. The thesis brings PWT 9 (based on ICP 2005 and 2011) and UQICDv2.1.1 based on all available ICP (1970-2011) to estimate robustness of P&Sa' s modeling.

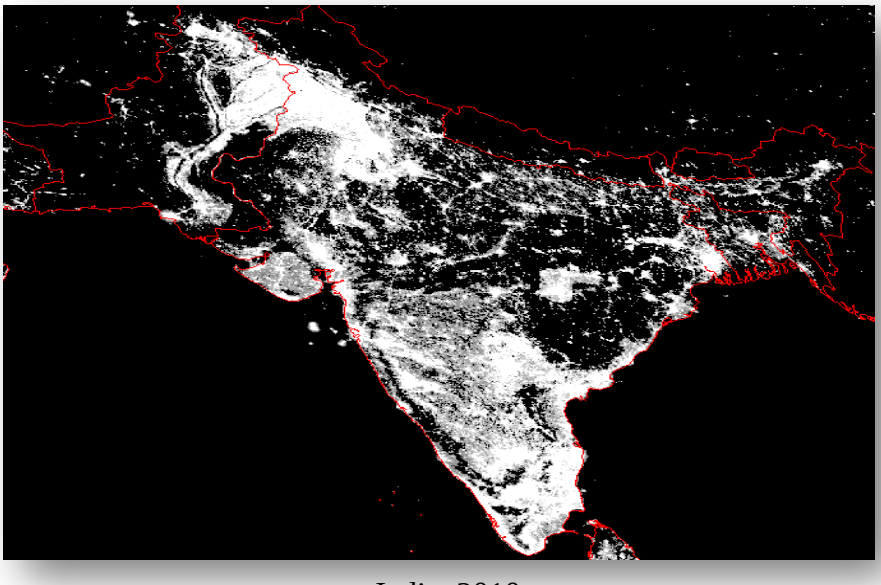
Recently, several papers have studied various applications of NTL in economic activities, including Sutton and Costanza (2002), Ebener et al. (2005), Paul et al. (2009), Elvidge et al. (2009), Henderson et al. (2012), Elvidge et al. (2012) , Keola et al. (2015), Mellander et al. (2015). Although there are a variety of findings, these studies generally agree that NTL are strongly correlated with economic activity. For example, Figure 1.1 illustrates lights in India between two separate years, 1994 and 2010. We can see that there is a substantial increase in light intensity. To be more precise, the growth rate of lights per capita in India is 112% between 1994 and 2010. The growth rate of national accounts consumption per capita is 100% for India for the same period (Pinkovskiy and Sala-i-Martin, 2016a, page 3).[2] Hence, there is a strong link between NTL and economic activities.

In this thesis, I will examine how different international income estimates behave in a particular set of contexts and thus provide some evidences to answer the third question. In order to fulfill this task, I will build on the work of Pinkovskiy and Sala-i-Martin (2016a, b) (P&Sa, b). In particular, in their research, they use NTL to predict true income. The approach in P&Sa is to optimally weight GDP from WDI and survey means from household surveys. This thesis will extend P&Sa's work by using four different GDP series from both older sources (PWT 7.1, PWT 8.1) and newer sources (UQICDv2.1.1 and PWT 9.0) in three directions. I will present these in detail in section 1.2.

[2] National accounts consumption per capita data are taken from WDI data.



India, 1994



India, 2010

Figure 1.1 (source Pinkovskiy and Sala-i-Martin 2016a, page 43)

1.2 Objectives

This thesis aims to test the behavior of five different international income datasets through extending P&Sa's modelling work in the paper "*Lights, Camera,....Income! Illuminating the National Accounts-Household Surveys Debate*". In particular, the study of Pinkovskiy and Sala-i-Martin (2016a) is significant since it provides substaintial insight into the behaviour of national

accounts GDP per capita and surveys means from household surveys in predicting the best linear unbiased estimator of true income[3]. However, the approach P&Sa used to check the robustness of their findings is incomplete and can be extended in three directions.

Firstly, the thesis evaluates the robustness of P&Sa's results when GDP is measured by WDI and four other alternative measures of income (PWT 7.1, 8.1, 9.0 and UQICDv2.1.1) are used to evaluate whether aggregate GDP per capita or GDP measured through household surveys are correlated with NTL.

Secondly, P&Sa's econometric approach to the estimation of the relative optimal weights on GDP per capita and survey means for calculating true income is indirect and standard errors can only be obtained through bootstrapping. We propose to use a combination forecast approach as an econometric alternative. Specifically, we assume that log GDP per capita and log survey means are unbiased proxies for log true GDP per capita.[4] By doing so, we can use combination forecast method to estimate directly the combination weights with standard errors. Again, we will use five different series of GDP estimates to check the sensitivity of the combination forecast approach.

Thirdly, P&Sa have drawn their inferences based on aggregate data for the world on average. This does not means these inferences will be still robust with disaggregated data. Therefore, we will build on the work of P&Sa, by re-estimating their model using World Bank income groups and geographical groups. Once again, five different GDP measures will be used to check the sensitivity of the results.

The empirical findings of previous studies suggest that different versions of PWT perform differently(Ciccone and Jarociski, 2010), or the performance of World Development Indicators (WDI) data is different from PWT (Pinkovskiy and Sala-i-Martin 2016b). Will these facts be maintained in the context of this study? This thesis will take into account the previous findings again.

To sum up, the main objective of this thesis is to check the robustness of P&Sa's findings that we should give 100% weight on GDP per capita, or zero weight on survey means in estimating true income (Pinkovskiy and Sala-i-Martin 2016a). Due to large discrepancies between national

[3] In their papers, P&S used national accounts or GDP per capita from WDI datasets

[4] This assumption is different with that of P&S. In their model, P&S assume that log GDP per capita and log survey means are biased proxies for log true GDP per capita

accounts GDP per capita and household survey means, policy makers or researchers face dilemma about whether it is right to use GDP per capita from national accounts, or to use survey means from household surveys. This is particularly important to decide policies related to poverty alleviation in developing countries (Pinkovskiy and Sala-i-Martin 2016a). Therefore, the results of this thesis are important since it provides further evidence in this regard. In order to fulfill this aim, this study will be conducted in three ways. Firstly, we will replicate P&Sa's work replacing GDP per capita from WDI by four difference alternative sources of GDP per capita. Secondly, we use combination forecast approach to re-estimate the optimal weights on GDP per capita and survey means. Finally, we will check whether P&S's findings are robust in disaggregated model.

1.3 Contributions

This thesis possesses two key contributions to the existing literature. The first is the extension of P&Sa and P&Sb to include UQICD and PWT 9.0 which are newer datasets not yet as widely considered in the literature. The UQICD presents a new approach which incorporates all the ICP benchmarks to construct the PPP and income data, while the traditional GDP measures from WDI or PWT use only one or two ICP benchmarks to construct data (Rao et al., 2014). UQICDv2.1.1 was released in November 2015 and it was the first to include ICP 2011. PWT 9.0 was released on June 9, 2016 and included the 2011 ICP. Since each source of GDP per capita plays the same role in the research, we can therefore compare the behavior of UQICD or PWT 9.0 with other sources of GDP per capita.

Secondly, the study contributes by proposing the use of a combination forecast approach to estimate P&Sa's theoretical weights. The method provides a direct estimation, unlike P&S's where the quantities of interests are found through plugging in estimates from auxiliary regressions (the method is presented in Chapter 4).

This thesis will divided into seven chapters: Chapter 2 will review the general applications of NTL data, and discusses in more detail the particular applications of NTL in accessing the existing GDP measures in the literature. Chapter 3 describes how NTL data are generated and a brief overview of other GDP data sources. Chapter 4 presents the methodology used in this thesis including the method of P&Sa and combination forecast approach. It also provides the statistical test used to check the results of P&Sa across income groups and regions. Chapter 5 and Chapter 6 are devoted to showing the empirical results of this study. Finally, Chapter 7 concludes and discusses the findings of this study.

Chapter 2 Literature Review

2.1 A brief look at nighttime lights studies

This review will show the diverse applications of recent studies in using satellite-recorded nighttime lights to estimate different aspects of human welfare of nations. While there is some disagreement about the accuracy of nighttime lights data in measuring economic growth in developing countries, there is considerable literature which supports the view that nighttime lights data are a good proxy measure for economic activities.

While there have been diverse findings, there is general agreement in the literature that nighttime lights have a strong correlation with economic activity. For example, early studies such as Elvidge et al. (1997) (cited in Chen and Nordhaus, 2011) using data for 21 countries showed a strong correlation (R^2=0.97) between illuminated area and GDP (both in logarithms form). Furthermore, a high correlation between luminosity and GDP per square kilometer at the national level was also found by Sutton and Costanza (2002). Later studies have focused on using nighttime lights for the prediction of income per capita at the subnational level. For example, Ebener et al. (2005) found that GDP per capita at the national and subnational levels can be predicted by using lit area and percentage of frequency of lighting. To be more specific, the model produces better results in predicting GDP when climate and agriculture are considered. Later, developing Ebener's model, Sutton et al. (2007) added estimated urban population at the state level to deal with the saturation problem in traditional luminosity image and used the new model in four countries: India, the United States, China and Turkey. In this model, a log-log linear relationship between the size of urban regions and population was used to estimate "urban population'' of each state of a country. Then they concluded that the estimation of aggregate national measures of GDP was improved substantially by spatial disaggregation of estimates.

The nighttime luminosity data have also been used to investigate poverty at the gridded, sub-national and national levels (Tilottama et al., 2013). For example, in their research, Elvidge et al. (2009) have used the nighttime lights images and the LandScan population data to construct a globally consistent and reliable measure of poverty at the one kilometer square grid level. The result showed that about 2.2 billion people of the global population were estimated to be living in poverty compared to the 2.6 billion estimated by the World Development Indicators. Further work by Wang et al. (2012) used three-year DMSP/OLS night-time light satellite imagery to estimate poverty degree for 31 provincial areas in China. Principal component analysis (PCA)

has also been used to establish an integrated poverty index (IPI) adopting 17 socio-economic indexes, which was demonstrated to offer a good representative of poverty level in China. The result found a high correlation between Average Light Index (ALI) and IPI (R^2=0.85), which suggested that the nighttime luminosity data can be useful in analysing poverty at the regional level. In a similar vein, Elvidge et al. (2012) established The Night Light Development Index to measure the distribution of income and wealth at national and subnational scale.

Another advantage of using night lights data is that they are used as a potential source to estimate informal economic activity and remittances, which was significantly achieved by Paul et al. (2009). The research developed a model to measure informal economy and remittances for Mexico using the spatial patterns of nighttime lights and the more reliable data of Gross State Product (GSP) of the United States (U.S.). They found that using their method to estimate Mexico's informal economy and remittances increased the official estimate of GNI by 150 per cent. Other applications of nighttime lights would also be significant in evaluating inequality in human development (Elvidge et al., 2012).

In spite of numerous useful applications, using nighttime lights data possesses certain limitations. Keola et al. (2015) argued that the value-added by agriculture and forestry, which account for a large share of national income in developing countries, might not be explained by nighttime lights alone. These two sectors release less or even zero observable nighttime lights. Therefore, economic activities can be underestimated by using the nighttime luminosity data. Another shortcoming of using nighttime light data is that satellites cannot record light intensity which is higher than 63 digital number, causing an underestimate of the development level of a region, especially in the booming metropolis of most developed countries (Bluhm and Krause, 2016). Furthermore, over glow and blooming also exert influence on the quality of light data, and different generations of satellites may produce inconsistent data (Pinkovskiy and Sala-i-Martin 2016a, page 7-8).

The above discussion has highlighted the benefits of using nighttime lights data as a proxy for measuring economic activities. It reviews a range of various applications as well as listing some flaws of NTL data. Nevertheless, there is still a level of uncertainty about the robustness of NTL in measuring economic activity.

2.2 GDP measurement comparison

In the preceding section, the review has indicated nighttime lights might be a reasonable proxy for measuring economic activity in a number of settings. In recent works by Pinkovskiy and Sala-i-Martin (2016a, b) nighttime lights have been compared to alternative measures of internationally comparable real GDP per capita. In this section I will examine the measurement of internationally comparable real GDP per capita.

To date, Penn World Table, World Development Indicators and International Comparison Program are the three most frequently used sources of international data on GDP per capita in almost every cross-country study. All these sources are used by numerous researchers and other users worldwide. However, there is a large misunderstanding of how these sources of PPP adjusted GDP are constructed and therefore related. The ICP has been collecting basic data on prices of a basket of commodities through surveys conducted at participating countries since the early 1970s (see The History of ICP at http://go.worldbank.org/WLPETUYSO0). These surveys are not conducted regularly, only sporadically. The last two rounds have been truly global (2005 and 2011). At each round of ICP the data are used to produce PPPs that can be used to construct real GDP per capita for that year by dividing the national accounts GDP per capita in domestic currency by the corresponding set of PPPs to obtain constant price GDP per capita. PWT and WDI provide real GDP per capita for a large number of countries. These measures are obtained by extrapolations (across countries that did not participate) and over time of the ICP PPPs. The observed differences across PWT and WDI versions result from the methodology used to extrapolate as well as the ICP round used to construct the extrapolations.

A few studies have undertaken a comparison between them. Rati (2009) used a descriptive comparison and the computation of cross-country income inequality to compare between real (PPP) GDP from ICP and WDI and PWT. The study found that there are enormous differences for many countries between these three datasets. Further work by Ram and Ural (2014) using similar methods and more updated versions of ICP and WDI and PWT data revealed the same finding. Using the ICP 2005 version as a benchmark to compare between the WDI and PWT data, huge differences between WDI and ICP data are shown in a small number of countries, whereas there are several cases for which PWT 7.1 and the ICP shows huge differences for many countries**.** Deaton and Aten (2014) provide an overview of the inconsistencies between ICP 2011 and 2005, which have led to the differences between the latest version of WDI, based on 2011,

and its previous version. For example, the estimates of Chinese GDP was 40% lower in the 2005 PPP revision compared to the previous version, while the subsequent (2011) PPP revision did not incorporate this change yet. Although Rati (2009) and Ram and Ural (2014) provide a comparison between WDI and PWT, these studies are hampered by lack of statistical framework in their methodology, as well as lack of an independent benchmark in measuring GDP per capita with WDI and PWT.

Another section of literature review has focused on assessing different versions of the Penn World Table (PWT) data and WDI. The PWT income data has been used in most cross-country empirical work by numerous scholars and other users throughout the world. At the latest count Google Scholar recorded 4997 citations in September 2016 for the original paper by Summers and Heston (1991). Since then PWT has had substantial revisions and updates to the data (the latest version is 9th vintage), and users of the PWT typically choose the most updated versions in their research (Pinkovskiy and Sala-i-Martin, 2016b). However, the different versions of the PWT show substantial variation in the estimates, and this leads to a significant effect on research outcomes. For example, in their research Ciccone and Jarociski (2010) found that when using the PWT 6.2 data instead of the previous version PWT 6.1 data, the robust determinants of economic growth (the role of government, international trade, geography and demography) have changed substantially. Another research also concluded that "*measurement error is a serious problem in the data of many countries in the Penn World Tables*" (Dawson et al. (2001), page 1008). Another significant work is by Johnson et al. (2012). Specifically, they suggested a comparison between the two versions PWT 6.1 and PWT 6.2 released in 2002 and 2006 respectively. The standard deviation of the changes was approximately 1.1% in the average annual growth rates of countries over a period of 29 years (1970-1999) – a great difference compared to countries' average growth rate of 1.56%. For an intuitive example, the study noted that the five worst growth performers in Africa using the former version no longer remain so when using the latter version. PWT 8 has seen substantial changes in the underlying construction of the PWT (Feenstra et al., 2015). Pinkovskiy and Sala-i-Martin (2016b) have recently compared PWT 7 and PWT 8. While earlier versions of the PWT were anchored on the latest ICP benchmark data, PWT 9 is anchored on the last two (2005 and 2011).

In the above part of this section, I have explained that a great deal of literature focused on the discussion of using WDI, PWT and ICP as the main sources of real GDP per capita measurement. The remainder of this section will be devoted to a consideration of the new source of real income

data: “The University of Queensland International Comparison Database”. The latest version is the UQICDv2.1.1 that was released in November 2015. The UQIDC was first introduced by Rao et al. (2010). The study developed an economic framework for the construction of Purchasing Power Parities (PPPs) which allow us to combine all the PPP benchmark data from various phases. The UQICD dataset was constructed by improving the econometric method used in the production of the PWT and WDI. The key difference between the UQICD and PWT & WDI is that instead of anchoring on one or two particular selected ICP benchmark data, the UQICD approach uses all ICP benchmarks as well as additional OECD PPP measurements obtained in some intermediate years between ICP benchmarks. For example, while the WDI (2014) PPP series were created by taking 2011 ICP benchmark data (released in April, 2014), the PWT 8.0 have used 2005 and 2011 consecutive benchmarks to interpolate PPPs for intervening years for nations that were available in both benchmarks and for those countries that were available in only one benchmark, PPPs from that year were extrapolated backwards and forwards (Rao et al., 2014). Furthermore, UQICD combines information from national level data in its econometric approach, where a model for the national price levels, assuming heteroskedastic and spatially correlated errors across countries provides cross-sectional predictions of PPPs for non-participating countries and years where no ICP data are available. The cross-sectional model along with data on country specific price movements are combined using a state–space formulation and optimum predictions of PPPs are acquired. This approach is applied to construct panels of PPPs for 181 countries over the period 1971 to 2012. Therefore, the new constant price real incomes series (UQICD) can be used as an alternative source of measuring income and wealth.

The review of literature in this section has largely emphasized various sources of PPP-adjusted GDP estimates including both the old (PWT5, 6 and 7, WDI) and the new (PWT8 and UQICD). However, no studies to date draw a robust conclusion about which source is the best and most reliable in measuring true income.

2.3 Nighttime lights and GDP per capita measurement

In this section I will review recent literature where NTL has been used as an independent benchmark to access existing GDP measures and estimate true growth rates.

A growing literature in economics has paid more attention to artificial night light data and significant works have been achieved to deal with estimation errors by linking these observations

with economic growth. The work of Henderson et al. (2012) changed the way in which nighttime lights data are applied. Using an error-measurement approach, they pioneered the development of a formal statistical analysis to compare nighttime lights data with traditional output measures. The main assumption is that the measurement error in luminosity data is uncorrelated with the measurement error in national accounts GDP; they drew a conclusion that the information on lights growth was roughly equal to the estimation of true growth rate of income for countries with good national income accounts data. By contrast, for countries with poor national income accounts, the optimal estimate of true income growth is a combination with approximately equal weights. Furthermore, for low quality data countries, the study has shown that new estimated average annual growth has as much as 3 percentage point difference from the official data. However, there is disagreement in the literature as to exactly how NTL predicts economic growth. Bickenbach et al. (2016), using regional data in Brazil and India, argue that there are still considerable divergences between night light growth and observed GDP growth not only in statistical terms but also in economic terms across regions. This disagreement suggests that there is a need for further exploration of the robustness of NTL data in measuring economic growth and it is the purpose of this study to examine it in more detail.

The most recent influential work in the literature on NTL is Pinkovskiy and Sala-i-Martin (2016a). They used a data-driven method to deal with measurement error in order to construct the optimal combination of what is labeled “national accounts measured GDP per capita” and “survey means”. The national accounts based real GDP per capita used is that constructed by the World Bank using ICP 2005, and therefore they are the earlier version of the World Bank’s WDI. The second measure of GDP per capita used is also constructed by the World Bank using ICP 2005 or 2011; however, national income is obtained from data from more than one thousand household surveys across 131 developing countries, and 21 high income countries made available by the poverty research group at the World Bank. The study uses the data based on ICP 2005. This measure will be referred to in this thesis as Survey Income per capita. The study compared these two “development indicators” to the evolution of satellite-recorded nighttime lights. As a result, they found that the growth of the economies in poor countries has been higher than estimates proposed in the survey. Secondly, a faster rise in living standards in developing countries and a more equal world income distribution than suggested by the GDP based survey are key findings of the study. Finally, the study concluded that for developing countries that are richer and growing faster, the performance of the GDP based survey appears to be worse,

whereas national accounts better captures desirable outcomes for the poor (such as longer life expectancy, better education, and access to safe water). In their study, although Pinkovskiy and Sala-i-Martin (2016a) provided various controls to check the robustness of their findings, it is limited in that only WDI GDP per capita are used without sensitivity check with other sources of real GDP per capita.

Subsequently, Pinkovskiy and Sala-i-Martin (2016b) have released a new discussion paper where nighttime lights have been used as an independent measurement of GDP per capita to evaluate PWT and WDI datasets in estimating true income. Generally, they find that older vintages of PWT are not necessarily prevailed by the newer ones. Specifically, the study noted that the chain series index in PWT 7.1 performs considerably better than the equivalent variables to it in PWT 8.0 and PWT 8.1. Additionally, another key finding is that WDI is better in capturing unobserved true income compared to all the vintages of PWT. Furthermore, the authors found that each successive PPP revision has resulted in an improvement of WDI in predicting true income. This was a particularly significant finding because this is the first study that used a statistical framework to evaluate different sources of income per capita. The current thesis will use the same data sets (with the addition of UQICD and PWT 9.0) in the empirical sections.

To sum up, the review has shown a new application of NTL, using error measurement approach not only to predict true economic growth but also to be an independent benchmark in assessing various sources of GDP per capita. There is a noted gap in the existing literature of checking the robustness of NTL and GDP per capita data. The above studies have used both developed and developing countries in the same model. As the review has already indicated, there might be significant differences in the reliability of NTL across countries. This thesis will add to the literature by using UQICD as well as Survey Income, WDI and PWT measures of GDP per capita in examining the robustness of NTL in aggregate as well as disaggregated models.

Chapter 3 Data description

3.1 PWT

This thesis uses three versions of the Penn World Table (version 7.1, 8.1 and version 9.0), which were released in July 2012, April 2015 and June 2016 respectively. The differences between version 8.1 and 7.1 are described clearly by Feenstra et al. (2015). UQICD v.2 incorporates all ICP data including the ICP 2011 benchmark price survey results. PWT versions 7 and 8 do not incorporate the ICP 2011. The very recently released PWT9 incorporates ICP 2011.

For the PWT 7.1 version, there are two major changes in comparison to the PWT 7.0: first, the updating and a number of revisions of national accounts; and second the initial PPP for investment has changed (Heston et al., 2012). The vintage 7.1 includes 189 countries over a period 1950-2010 and has 2005 as the reference year. Up to version 7.1, the PWT bases different countries' national accounts data in local currency, and the ICP benchmark price survey used a regression to estimate missing survey data in the benchmark year to allocate PPP to countries.[5] In the next step, national growth rates of consumption, investment and government spending are weighted together in order to compute PPP-adjusted GDP away from the benchmark year (Pinkovskiy and Sala-i-Martin, 2016b).

The PWT version 8.1 firstly introduces real GDP on the expenditure side and real GDP on the output side. Another key contribution is that the PWT 8.1 uses multiple ICP benchmarks to calculate GDP in constant prices, thus making growth rates to be less sensitive as new benchmarks appear. As Feenstra et al. (2015) note, these measures are never changing to future price data, as well as invariant over time unless there exists a revision of underlying national accounts data.

With regard to the PWT 9.0, there were three major changes made to the data compared to PWT 8.1. Firstly, it uses new PPP data of ICP 2005 and 2011 and other sources. Secondly, it uses updated and extended National Accounts data till 2014. Thirdly, the data set revises data for factor input and labour cost shares (Feenstra et al., 2016).

In this paper, I will use variable "rgdpch" (real GDP per capita in constant international dollars for 2005) from PWT version 7.1, and variable "rgdpe" (real GDP on the expenditure side) from

[5] For PWT 8.0, this imputation no longer exists.

version PWT 8.1 and version PWT 9.0. I use different variables from three vintages of PWT because: 1) the variable "rgdpch" does not appear in PWT 8.1 and PWT 9.0, 2) the "rgdpch" and "rgdpe" can be comparable since they are both generated from modified growth rates using price surveys, 3) both variables are used to compare the standards of living across countries over time (Pinkovskiy and Sala-i-Martin, 2016b).

The PWT 7.1 vintage can be downloaded at

http://www.rug.nl/research/ggdc/data/pwt/pwt-7.1

The PWT 8.1 vintage can be downloaded at

http://www.rug.nl/research/ggdc/data/pwt/pwt-8**.1**

The PWT 9.0 vintage can be downloaded at

http://www.rug.nl/research/ggdc/data/pwt/pwt-9.0

3.2 UQICD

The University of Queensland International Comparison Dataset (UQICD) has been constructed by D.S.P. Rao, A.N. Rambaldi and H.E. Doran and their team at the University of Queensland. The construction of UQICD is described in detail in "UQICD Version 2.0 User Guide" (Rao et al., 2014). This thesis will use variable "CRGDP" since this variable is the closest corresponding variable to the variable "RGDPCH" in PWT 7.1 and "RGDPE" in PWT 8.1. As shown in (Rao et al., 2014), "CRGDP" series are computed by the following formula:

$$CRGDP_{it}^{2005} = \frac{GDP_{it}}{PPP_{it}^{2005}}$$

Where

PPP_{it}^{2005} represents the PPP of currency of country expressed in 2005 prices

GDP_{it} represents the gross domestic product of country i in period t expressed in local currency units in current prices.

$CRGDP_{it}^{2005}$ is the constant price real GDP

Data on real GDP per capita at constant 2005 year prices is taken from UQICD version 2 (2.1.1) which was released in November 2015. The series are available for download at http://uqicd.economics.uq.edu.au/data.php

As noted by Rao et al. (2014), a new econometric approach has been used in UQICD in constructing extrapolated PPPs, thus resulting in some advantages in comparison to PWT and WDI. Most importantly, the panels from UQICD are the results of using all the sources to produce optimal extrapolations of PPP. Second, for each of the PPPs available from UQICD, there exists the possibility of attaching measures of reliability. Third, the new economic framework allows constructing extrapolations which can meet various needs. Fourth, the choice of reference country does not affect the predicted panels of PPPs from UQICD. Finally, every predicted PPP available from UQICD can be explained as weighted sums of all available information.

3.3 Nighttime Lights

This thesis will use the same nighttime lights data as the paper "*Lights, Camera,....Income! Illuminating the National Accounts-Household Surveys Debate*" which was written by Pinkovskiy and Sala-i-Martin (2016a)[6]. The rest of this section will be used to describe how nighttime lights data are produced.

The Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) is responsible for collecting data on lights at night. Images are sent by a series of orbiting satellites, for every 30 arc-second output pixel (approximately 1 square kilometre at the equator) between 65 degrees south and 75 degrees north latitude. Every location on the Earth is observed by each satellite at 20:30 to 22:00 local time. In the next step, these raw data are processed and preserved at the National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center (NGDC) and they then deliver the final data to the public. In processing, cloud cover, snow, auroral activity (the northern and southern lights) and forest fires are removed from these images. The final output is available for download at

http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

The luminosity data is available for the period during 1992-2013. A six-bit digital number (DN) in grid format represents the intensity of lights. The range of digital numbers is from 0 (no light)

[6] The author is indebted to Maxim Pinkovskiy for providing their data for this study.

to 63, see Table 2.1 (Keola et al. 2015, page 326). The sum of all the digital numbers across pixels is used to produce light proxy for aggregate income

$$\text{Lights}_{j,t} = \sum_{i=1}^{63} i * (\#\ of\ pixels\ in\ country\ j\ and\ year\ t\ with\ DN = i)$$

Almost all studies on nighttime lights in economics used the above formula (Henderson et al. 2012; Pinkovskiy and Sala-i-Martin 2016a and 2016b; Chen and Nordhaus 2011). According to Henderson et al. (2012), in the case that there are multiple satellites in the same year, the logarithms of aggregate luminosity measure will be averaged.

Table 3.1: Nighttime light data for selected countries 1992–2009 average

	Cambodia	Lao PDR[a]	Vietnam	Burundi	Myanmar	Mongolia
DN0	98.92%	98.53%	74.60%	98.63%	98.39%	99.69%
DN1-2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DN3-5	0.37%	0.61%	11.35%	0.38%	0.59%	0.18%
DN6-10	0.40%	0.56%	9.10%	0.57%	0.66%	0.08%
DN11-20	0.16%	0.18%	2.85%	0.19%	0.23%	0.03%
DN21-62	0.15%	0.13%	2.02%	0.22%	0.13%	0.03%
DN63	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%
Gini(DN)	0.994	0.991	0.848	0.992	0.990	0.998
Pop. Density	78.97	26.60	274.53	335.84	78.33	1.69
Percent urban	19.56	30.83	29.15	10.14	31.00	65.54
GDP per capita, PPP 2005 ($)	1882.09	2002.11	2610.56	473.92	-	3608.60
GDP per capita, 2000 ($)	587.99	561.52	775.76	150.28	-	1249.14

Source: Keola et al. (2015).

3.4 WDI

This thesis will use national accounts data variable “GDP per capita, PPP, constant 2005 international dollars’’ from the WDI of World Bank (WB). This source of dataset provides relatively sufficient data for almost all countries without missing any data. GDP per capita from

PWT and WDI has been used in almost every cross-country study by numerous scholars and other users (Pinkovskiy and Sala-i-Martin 2016a; Ram and Ural 2014). In comparison to PWT, as Henderson et al. (2012) noted, WDI provided more reliable rank groups of countries based only on the quality of their national accounts data. Furthermore, national growth rates are not modified by WDI, while national growth rates are modified by the PWT extrapolation methodology (Johnson et al. 2012 cited in Pinkovskiy and Sala-i-Martin 2016a).

3.5 Other Data

Other Data such as Survey Income per capita, log electricity production, log area, log shares of GDP in agriculture were provided by the author of the paper "*Lights, Camera,....Income! Illuminating the National Accounts-Household Surveys Debate*". (Pinkovskiy and Sala-i-Martin 2016a, page 8).

Chapter 4 Methodology

The methodology used in this study includes a re-estimation of P&Sa's model where the measure of GDP per capita and consumption per capita from WDI in the original paper is replaced by GDP per capita and consumption per capita from UQICD, PWT 7.1, PWT 8.1 and PWT 9.0 to check the sensitivity of their findings. Secondly, I will use a combination forecast approach to assess the results of P&S's method using statistical measures of uncertainty. Finally, I study the robustness of their findings across regional country groupings and income level country groupings.

4.1 Replication of Pinkovskiy and Sala-i-Martin (2016a) – Extending to include UQICD, PWT 7.1, PWT 8.1 and PWT 9.0

Mathematical Framework

In this section, I will present a mathematical framework that Pinkovskiy and Sala-i-Martin (2016a) used to estimate the best unbiased linear predictor of log true income per capita $(y_{i,t}^{*})$ for country i in year t. This framework assumes that true income per capita is the result of some nonstationary exogenous process. It is impossible to observe $(y_{i,t}^{*})$ directly, whereas log light per capita $(y_{i,t}^{L})$, log measured GDP per capita $(y_{i,t}^{G})$, and log survey mean income $(y_{i,t}^{S})$ for country i and year t can be observable directly. The relationship between log true income per capita and these data will be shown by the following system of equations:

$$y_{i,t}^{L} = \beta^{L} y_{i,t}^{*} + \varepsilon_{i,t}^{L} \tag{1}$$

$$y_{i,t}^{G} = \beta^{G} y_{i,t}^{*} + \varepsilon_{i,t}^{G} \tag{2}$$

$$y_{i,t}^{S} = \beta^{S} y_{i,t}^{*} + \varepsilon_{i,t}^{S} \tag{3}$$

Where $\varepsilon_{i,t}^{L}$, $\varepsilon_{i,t}^{G}$, $\varepsilon_{i,t}^{S}$ are error terms. Henderson et al. (2012) and Chen and Nordhaus (2011) have used another framework which is very close to this approach. A slight difference is that while P&Sa assume that log GDP per capita is a biased proxy for log true income ($\beta^{G} \neq 1$), these two papers' assumption is that log GDP per capita does not deviate from log true income ($\beta^{G} = 1$). In other words, this key assumption can be used to explain the fact that, if $\beta^{G} > \beta^{S}$, on average, national accounts captures true national income better than household surveys, which result in a persistent divergence of mean of GDP per capita and surveys over time.

There are two assumptions. First of all, the error terms $\varepsilon_{i,t}^{L}$, $\varepsilon_{i,t}^{G}$, $\varepsilon_{i,t}^{S}$ and true income are mean independent, which means,

$$E\left(\varepsilon_{i,t}^{L} \mid y_{i,t}^{*}\right) = E\left(\varepsilon_{i,t}^{G} \mid y_{i,t}^{*}\right) = E\left(\varepsilon_{i,t}^{S} \mid y_{i,t}^{*}\right) = 0 \tag{A1}$$

The second and most important assumption, is that measurement error in using observed light ($\varepsilon_{i,t}^{L}$) is uncorrelated with measurement error in household surveys ($\varepsilon_{i,t}^{S}$) and national accounts ($\varepsilon_{i,t}^{G}$) is conditional on true income. That is,

$$E\left(\varepsilon_{i,t}^{G}\varepsilon_{i,t}^{L} \mid y_{i,t}^{*}\right) = E\left(\varepsilon_{i,t}^{G}\varepsilon_{i,t}^{S} \mid y_{i,t}^{*}\right) = 0 \tag{A2}$$

Henderson et al. (2012) and Chen and Nordhaus (2011) have also used the second assumption to explain why luminosity data are useful. This is an appropriate assumption since nighttime lights data are collected by satellites from space without border limits or institutional structures, while national income accounts and survey income data are obtained within national institutional frameworks.

P&Sa' approach aims to find the best linear unbiased estimator of log true income per capita $y_{i,t}^{*}$ by using $y_{i,t}^{G}$ and $y_{i,t}^{S}$,

$$z_{i,t} = \gamma_G y_{i,t}^{G} + \gamma_S y_{i,t}^{S} \tag{4}$$

In order to minimize the mean squared error, the objective function should be:

$$(\gamma_G^{*}, \gamma_S^{*}) = \arg\min_{\gamma_G, \gamma_S} E\left(\left(y_{i,t}^{*} - \gamma_G y_{i,t}^{G} - \gamma_S y_{i,t}^{S}\right)^2\right) \tag{5}$$

Subject to

$$E(\gamma_G^{*} y_{i,t}^{G} + \gamma_S^{*} y_{i,t}^{S} \mid y_{i,t}^{*}) = y_{i,t}^{*} \tag{6}$$

Or

$$\gamma_G^{*}\beta^{G} + \gamma_S^{*}\beta^{S} = 1 \tag{7}$$

Now, substituting equation (2) and (3) into the value function equation (5), we get

$$E\left(\left(y_{i,t}^{*} - \gamma_G y_{i,t}^{G} - \gamma_S y_{i,t}^{S}\right)^2\right) =$$
$$(1 - \gamma_G^{*}\beta^{G} - \gamma_S^{*}\beta^{S})\, E\left(\left(y_{i,t}^{*}\right)^2\right) + \gamma_G^2 var\left(\varepsilon_{i,t}^{G}\right) + 2\gamma_G\gamma_S cov\left(\varepsilon_{i,t}^{G}, \varepsilon_{i,t}^{S}\right) + \gamma_S^2 var\left(\varepsilon_{i,t}^{S}\right) \tag{8}$$

And using equation (7), equation (5) will become:

$$(\gamma_G^*, \gamma_S^*) = \arg min_{(\gamma_G, \gamma_S)}\{{\gamma_G}^2 var(\varepsilon_{i,t}^G) + 2\gamma_G\gamma_S cov(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S) + \gamma_S^2 var(\varepsilon_{i,t}^S)\} \quad (9)$$

Subject to $$\gamma_G^*\beta^G + \gamma_S^*\beta^S = 1 \quad (10)$$

Solving this problem, we obtain:

$$\frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*} = \frac{\beta^G var(\varepsilon_{i,t}^S) - \beta^S var(\varepsilon_{i,t}^G)}{\beta^G var(\varepsilon_{i,t}^S) + \beta^S var(\varepsilon_{i,t}^G) - (\beta^G + \beta^S) cov(\varepsilon_{i,t}^S, \varepsilon_{i,t}^G)} \quad (11)$$

The fraction $\frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*}$ is the weight on log national accounts as a proportion of the total weight that should be received on log national accounts and log survey means in equation (4). However, the right hand side of equation (11) cannot be identified. To solve this problem, P&S have used nighttime lights data to recover the unknown side of equation (11). In particular, they estimate the population regression of log lights per capita on log national accounts and log survey means, that is:

$$y_{i,t}^L = b^0 + b^G y_{i,t}^G + b^S y_{i,t}^S \quad (12)$$

Then, the coefficients of equation (12) must be satisfied:

$$\frac{b^G}{b^G + b^S} = \frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*} \quad (13)$$

Equation (13) can be proved as follow

As we know, the coefficient b^G can be computed by the fomula below:

$$b^G = \frac{var(y_{i,t}^S) cov(y_{i,t}^G, y_{i,t}^L) - cov(y_{i,t}^G, y_{i,t}^S) cov(y_{i,t}^S, y_{i,t}^L)}{var(y_{i,t}^G) var(y_{i,t}^S) - \left(cov(y_{i,t}^G, y_{i,t}^S)\right)^2} \quad (14)$$

Similarity, the coefficient b^S can be identified by using equation (14), replacing $y_{i,t}^S$ by $y_{i,t}^G$. These formulas rely on the covariances and variances of $y_{i,t}^L$, $y_{i,t}^G$ and $y_{i,t}^S$. Under assumptions A1 and A2, and equations (1)–(3), the expression of these covariances and variances are as follow:

$$var\left(y_{i,t}^G\right) = (\beta^G)^2 var\left(y_{i,t}^*\right) + var(\varepsilon_{i,t}^G) \tag{15}$$

$$var\left(y_{i,t}^S\right) = (\beta^S)^2 var\left(y_{i,t}^*\right) + var(\varepsilon_{i,t}^S) \tag{16}$$

$$cov\left(y_{i,t}^G, y_{i,t}^L\right) = \beta^G \beta^L var\left(y_{i,t}^*\right) \tag{17}$$

$$cov\left(y_{i,t}^S, y_{i,t}^L\right) = \beta^S \beta^L var\left(y_{i,t}^*\right) \tag{18}$$

$$cov\left(y_{i,t}^G, y_{i,t}^S\right) = \beta^G \beta^S var\left(y_{i,t}^*\right) + cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right) \tag{19}$$

Substituting equations (15)-(19) into equation (14), we obtain:

$$b^G = \frac{\beta^L var\left(y_{i,t}^*\right)\left[\beta^G var\left(\varepsilon_{i,t}^S\right) - \beta^S cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)\right]}{var\left(y_{i,t}^*\right)\left[var\left(\beta^G \varepsilon_{i,t}^S - \beta^S \varepsilon_{i,t}^G\right)\right] + var\left(\varepsilon_{i,t}^S\right) var\left(\varepsilon_{i,t}^G\right) - cov(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S)} \tag{20}$$

Similarly, we can obtain b^S by replacing β^G with β^S in the numerator and $var\left(\varepsilon_{i,t}^S\right)$ with $var\left(\varepsilon_{i,t}^G\right)$ in the denominator.

Dividing equation (20) by $b^S + b^G$, we get

$$\frac{b^G}{b^G + b^S} = \frac{\beta^G var\left(\varepsilon_{i,t}^S\right) - \beta^S cov\left(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S\right)}{\beta^G var\left(\varepsilon_{i,t}^S\right) + \beta^S var\left(\varepsilon_{i,t}^G\right) - (\beta^G + \beta^S) cov(\varepsilon_{i,t}^G, \varepsilon_{i,t}^S)} = \frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*} \tag{21}$$

Consequently, although we cannot estimate γ_G^* and γ_S^* , as well as the sum $\gamma_G^* + \gamma_S^*$ directly, we can estimate the weight of log income per capita relative to total weight of these two proxies for log true income per capita.

Estimation

In this paper, since data sets have both cross-sectional and time series dimensions, a panel data model with fixed effects will be used to process these data. This approach has been used widely

in the literature on nighttime lights in economics, including Henderson et al. (2012) and Pinkovskiy and Sala-i-Martin (2016a, b). P&Sa also added year fixed effects, country fixed effects, and both year and country fixed effects, in their regression (Wooldridge, 2012)(page 484-492).

To test the hypothesis that $\beta^S/\beta^G = 1$, an instrumental variable estimation is used. From equation (17) and (18), we have

$$\frac{cov(y_{i,t}^S, y_{i,t}^L)}{cov(y_{i,t}^G, y_{i,t}^L)} = \frac{\beta^S}{\beta^G}$$

Therefore, testing the hypothesis $\beta^S/\beta^G = 1$ is equivalent to testing

$$\frac{cov(y_{i,t}^S, y_{i,t}^L)}{cov(y_{i,t}^G, y_{i,t}^L)} = 1$$

The left hand side of the above equation can be estimated by regressing log survey mean on log national accounts, using log nighttime lights as an instrumental variable to correct for measurement error in income.

Bootstrapping

Due to the non-linearities, standard errors are obtained by bootstrapping. In almost all literature on nighttime lights in economics, including Henderson et al. (2012), Pinkovskiy and Sala-i-Martin, (2016a) and Chen and Nordhaus (2011), confidence intervals are computed using bootstrapping. As Wooldridge (2002) noted "*The method of bootstrapping, which is a popular resampling method, can be used as an alternative to asymptotic approximations for obtaining standard errors, confidence intervals, and p-values for test statistics*" (page 378-379) . The process could be done as follows (Wooldridge 2012, page 225-226).

Assume that $\hat{\theta}$ is an estimate of a population parameter, θ. Our desire is to get a standard error for $\hat{\theta}$. This standard error then can be applied to calculate t statistics and confidence intervals. It is considerable to note that a valid standard error can be obtained by constructing the estimate from various random samples taken from the original data.

The implementation of this procedure is not difficult. Firstly, observations will be listed from 1 to n, then n numbers are chosen randomly and entirely by chance with replacement from the list.

This process provides a new data set (of size n) that contains the original data, with many observations appearing several times. For each time we produce a new sample data set from the original data, we can generate an estimate of θ using the same method that was applied on the original data. Define $\hat{\theta}^{(b)}$ as the estimate of bootstrapping sample b. In the next step, m new estimates can be produced by repeating the resampling and estimation m times, $\{\hat{\theta}^{(b)}: b = 1,2, \dots . m\}$. The **bootstrap standard error** of $\hat{\theta}$ is the result of the sample standard deviation of the $\hat{\theta}^{(b)}$.

$$bse(\hat{\theta}) = \left[(m-1)^{-1} \sum_{b=1}^{m} \left(\hat{\theta}^{(b)} - \bar{\hat{\theta}}\right)^2\right]^{1/2}, \tag{22}$$

where $\bar{\hat{\theta}}$ is the average of the bootstrap estimates.

P&Sa use bootstrap to construct approximate confidence intervals for the target parameters in their model, the optimal weights defined as

$$\widehat{w}_G = \frac{b^G}{b^G + b^S}$$

$$\widehat{w}_s = 1 - \widehat{w}_G$$

Where b^G and b^S are the OLS estimates obtained from the regression in (12).

4.2 Combination forecast approach

In this section, I will present a combination forecast approach to estimate the optimal weights of log GDP per capita and log survey means of w_G and w_S. This method was presented in detail in Timmermann (2006) (page 145-146). We propose combination forecast as an alternative to obtaining $\widehat{w}_G$ and $\widehat{w}_S$ and symmetric confidence intervals.

In this approach, we assume $y_{i,t}^L$, $y_{i,t}^G$ and $y_{i,t}^S$ are unbiased predictors of $y_{i,t}^*$

$$y_{i,t}^L = y_{i,t}^* + \varepsilon_{i,t}^L \tag{23}$$

$$y_{i,t}^G = y_{i,t}^* + \varepsilon_{i,t}^G \tag{24}$$

$$y_{i,t}^S = y_{i,t}^* + \varepsilon_{i,t}^S \tag{25}$$

We are expected to predict the log of true income per capita, $y_{i,t}^{*}$, by weighting $y_{i,t}^{G}$ and $y_{i,t}^{S}$ using combination forecasts:

$$y_{i,t}^{*} = w_G\, y_{i,t}^{G} + w_S\, y_{i,t}^{S} + \varepsilon_{i,t}^{*} \tag{26}$$

This approach is different to the one used by P&Sa with the assumptions that $\beta^{L} = 1$, $\beta^{G} = 1$ and $\beta^{S} = 1$ (see equation 1, 2 and 3 in section 4.1). This framework has been used by Henderson et al. (2012), Chen and Nordhaus (2011) ($\beta^{G} = 1$) and Chen and Ravallion (2004), Milanović (2005) ($\beta^{S} = 1$) in literature (cited in Pinkovskiy and Sala-i-Martin 2016a, page 9 and page 17). These assumptions means that log GDP per capita and log survey means are unbiased proxies for the measure of log true income per capita. Furthermore, this approach also uses the same assumptions A1 and A2 as P&Sa's method (see assumptions A1 and A2 in section 4.1). The forecast errors will be

$$e_{i,t}^{G} = y_{i,t}^{*} - \widehat{y_{i,t}^{G}} \text{ and } e_{i,t}^{S} = y_{i,t}^{*} - \widehat{y_{i,t}^{S}} \tag{27}$$

According to the assumptions, we have

$$E\left[e_{i,t}^{G}\right] = E\left[e_{i,t}^{S}\right] = 0 \tag{28}$$

Variances of forecast errors: σ_G^2, σ_S^2. Covariance is σ_{GS}

If the combination weights add up to one, with weights ($w_G,\ w_s = 1 - w_G$), then the combined forecast of log true income per capita will be unbiased

$$\widehat{y_{i,t}^{*}} = w_G\, \widehat{y_{i,t}^{G}} + (1 - w_G)\, \widehat{y_{i,t}^{S}}$$

The forecast error from the combination will be:

$$e_{i,t}^{*} = y_{i,t}^{*} - \widehat{y_{i,t}^{*}} = y_{i,t}^{*} - w_G\, \widehat{y_{i,t}^{G}} - (1 - w_G)\, \widehat{y_{i,t}^{S}} = w_G\, e_{i,t}^{G} + (1 - w_G)\, e_{i,t}^{S}$$

Therefore, combined forecast error is a weighted average of the individual forecast errors:

$$E\left(e_{i,t}^{*}\right) = 0$$

$$Var\left(e_{i,t}^{*}\right) = w_G^2\, \sigma_G^2 + (1 - w_G)^2\, \sigma_S^2 + 2\, w_G\, (1 - w_G)\, \sigma_{GS}$$

Differentiating with respect to w_G and calculating the first order condition, will give

$$w_G^* = \frac{\sigma_S^2 - \sigma_{GS}}{\sigma_G^2 + \sigma_S^2 - 2\sigma_{GS}}$$

$$w_S^* = 1 - w_G^* = \frac{\sigma_G^2 - \sigma_{GS}}{\sigma_G^2 + \sigma_S^2 - 2\sigma_{GS}}$$

Combination weight can be negative if $\sigma_{GS} > \sigma_S^2$ or $\sigma_{GS} > \sigma_G^2$

Note that having a negative weight on a forecast does not imply that it is meaningless - it simply means that the forecast will be able to offset the prediction errors of other models. Furthermore, a higher weight is given to more accurate models (small σ_i^2).

Estimation

We cannot estimate the optimal weights on GDP per capita and survey means from the following equation

$$y_{i,t}^* = w_G\, y_{i,t}^G + (1 - w_G)\, y_{i,t}^S + \varepsilon_{i,t}^* \tag{29}$$

As log true income per capita is unobservable. Instead, under assumption A1 and A2, we can use log light per capita to forecast the optimal weights:

$$y_{i,t}^L = w_G\, y_{i,t}^G + (1 - w_G)\, y_{i,t}^S + \varepsilon_{i,t}^* \tag{30}$$

Weights cannot be negative if $\sigma_{GS} = 0$. Such an assumption would rule out a correlation between $e_{i,t}^G$ and $e_{i,t}^S$ (see (27)) for all i and t. That is the forecast errors are uncorrelated. This can be imposed in this framework as follows.

Subtracting both side of equation (30) for $y_{i,t}^G$, we obtain:

$$y_{i,t}^L - y_{i,t}^G = (1 - w_G)\,(y_{i,t}^S - y_{i,t}^G) + \varepsilon_{i,t}^* \tag{31}$$

4.3 Extending P&S's method to Groups: Income and Region

The modeling conducted with the methodology explained in the section 4.1 has been implemented for a model of all countries across income groups and regions. To study the robustness of these results we estimate the models by using the World Bank income groups and geographical regions**.**

According to the World Bank, economies are classified into four income groupings: low, lower-middle, upper-middle, and high. The World Bank uses gross national income (GNI) per capita to

measure income. Income is calculated in US dollars using the World Bank Atlas method. Low-income countries are nations with a per-capita GNI in the current 2017 fiscal year between $1,025 or less in 2015; those with a GNI per capita between $1,026 and $4,035 are defined as lower middle-income countries; those with a GNI per capita between $4,036 and $12,475 are defined as upper middle-income countries; those with a GNI per capita of $12,476 or more are high income countries (The World Bank Group). In this paper, I will examine the robustness of P&S's findings using three income groups: low, lower-middle and upper-middle. There are a total of 590 observations, with 26 countries in low income group, 37 countries in lower-middle income group and 39 countries in upper-middle income group.

In terms of geographical regions, I investigate three different regions: Asia and Pacific, Africa and Middle East, Latin America and Caribbean. There are 445 observations, with 17 countries in Asia and Pacific, 48 countries in Africa and Middle East, and 22 countries in Latin America and Caribbean.

Now, I need to test whether the null hypothesis that the optimal weight on GDP per capita is 1 ($w_G^* = 1$), or survey means get zero weight ($w_S^* = 0$) are the same between different income groups or different region groups. To test this hypothesis, I will add interaction terms to baseline regression equation (12). The regression equation will be estimated as follow:

$$y_{i,t}^L = D_1 + D_2 + D_3 + b^{G1}D_1y_{i,t}^{G1} + b^{S1}D_1y_{i,t}^{S1} + b^{G2}D_2y_{i,t}^{G2} + b^{S2}D_2y_{i,t}^{S2} + b^{G3}D_3y_{i,t}^{G3} + b^{S3}D_3y_{i,t}^{S3} + \varepsilon_{i,t} \quad (32)$$

Where

$y_{i,t}^L$ Log lights per capita in country i and year t

D_1 dummy variable, $D_1 = 1$ if country is in group 1, otherwise $D_1 = 0$

D_2 dummy variable, $D_2 = 1$ if country is in group 2, otherwise $D_2 = 0$

D_3 dummy variable, $D_3 = 1$ if country is in group 3, otherwise $D_3 = 0$

$y_{i,t}^{G1}$ Log GDP per capita in country i and year t, where country i belongs to group 1

$y_{i,t}^{S1}$ Log survey income per capita in country i and year t, where country i belongs to group 1

$y_{i,t}^{G2}$ Log GDP per capita in country i and year t, where country i belongs to group 2

$y_{i,t}^{S2}$ Log survey income per capita in country i and year t, where country i belongs to group 2

$y_{i,t}^{G3}$ Log GDP per capita in country i and year t, where country i belongs to group 3

$y_{i,t}^{S3}$ Log survey income per capita in country i and year t, where country i belongs to group 3

b^{G1}, b^{S1}, b^{G2}, b^{S2}, b^{G3}, b^{S3} are coefficients

$\varepsilon_{i,t}$ is error term

Now, the null hypothesis tested is

$$w_G^*(group\ i) = 1 \text{ with } i = 1, 2, 3$$

To test the hypothesis we compute an asymptotic χ^2 statistic for non-linear combinations of estimates. The Wald type statistic is calculated using a delta method which is an approximation appropriate in large samples.

The estimates of relative weight on GDP per capita for group 1, group 2 and group 3 will be

$$\widehat{w}_G^{*group1} = \frac{b^{G1}}{b^{G1} + b^{S1}}$$

$$\widehat{w}_G^{*group2} = \frac{b^{G2}}{b^{G2} + b^{S2}}$$

$$\widehat{w}_G^{*group3} = \frac{b^{G3}}{b^{G3} + b^{S3}}$$

Note that group 1, group 2, and group 3 represent low income, lower-middle income, and upper-middle income respectively. When we test the null hypothesis that the optimal weights on GDP per capita are 1 are the same between different income groups. Similarly, group 1, group 2, and group 3 will represent Asia and Pacific, Africa and Middle East, and Latin America and Caribbean respectively. When we test the null hypothesis that the optimal weights on GDP per capita are 1 are the same between different region groups.

Chapter 5 Aggregate Models Results

In this chapter, I will use five different series of GDP per capita from WDI, UQICD, PWT 7.1, PWT 8.1 and PWT 9.0 datasets, as well as household survey means, and satellite-recorded NTL data in order to check the sensitivity of P&Sa's findings in estimating the optimal weight on GDP per capita and survey means for true income.

5.1 Regressions of Nighttime Lights on GDP per Capita and Survey Means

In this section, I replicate Table I columns (1), (3) and (4) in the paper "*Lights, Camera,...Income! Illuminating the National Accounts-Household Surveys Debate*" (Pinkovskiy and Sala-i-Martin 2016a, page 38)[7].

Table 5.1 shows the OLS coefficients of the regression of log nighttime lights per capita on log GDP per capita using five alternative measures of GDP with 680 country-year observations in the developing nations with household survey information. For consistency purposes, I used 680 observations compared to 701 observations of the original paper because of the different availabilities of different datasets. Cluster standard errors by country are used in all cases. As shown in columns 1-4 of Table 5.1, it is clear that there is no significant difference in the coefficient estimates when we use different GDP series. In particular, all coefficients are statistically significant irrespective of fixed effects included. However, if two types of country fixed effects are included, the coefficients on GDP per capita from WDI, PWT 7.1 and UQICD are considerably lower than the case without fixed effects or year fixed effects only. This result reflects the fact that light-generating processes may be different in different countries (Pinkovskiy and Sala-i-Martin 2016a, page 19). The estimates in the first five columns of Table 5.1 are therefore consistent with the assumption that there is strong correlation between the growth rates of nighttime lights per capita with growth rate of GDP per capita regardless of the datasets we consider.

[7] Note that we do not replicate columns (2) and (5) in the original paper because the result will not deviate from the original paper; we only consider the sensitivity of P&S's findings with different sources of income estimate.

Table 5.1: Univariate regression - Replicate columns (1) and (3) of Table I (Pinkovskiy and Sala-i-Martin 2016a, page 38)

Baseline Regressions										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Log Lights OLS	Log Lights OLS	Log Lights OLS	Log Lights OLS	Log Lights OLS	Log Surveys IV, Lights	Log Surveys IV, Lights	Log Surveys IV, Lights	Log Surveys IV, Lights	Log Surveys IV, Lights
No Fixed Effects										
Log GDP_PC WDI	1.16***					0.78***				
	(0.063)					(0.027)				
Log GDP_PC UQICD		1.17***					0.79***			
		(0.066)					(0.029)			
Log GDP_PC PWT 7.1			1.16***					0.767***		
			(0.059)					(0.285)		
Log GDP_PC PWT 8.1				1.01***					0.81***	
				(0.114)					(0.038)	
Log GDP_PC PWT 9.0					1.05***					.80***
					(.112)					(.036)
R2	0.73	0.72	0.74	0.61	0.65	0.82	0.82	0.81	0.64	0.70
Year Fixed Effects										
Log GDP_PC WDI	1.17***					0.78***				
	(0.066)					(0.028)				
Log GDP_PC UQICD		1.18***					0.79***			
		(0.069)					(0.03)			
Log GDP_PC PWT 7.1			1.16***					0.768***		
			(0.062)					(0.029)		
Log GDP_PC PWT 8.1				1.01***					0.81***	
				(0.119)					(0.038)	
Log GDP_PC PWT 9.0					1.06***					.80***
					(.119)					(.036)
R2	0.75	0.74	0.76	0.73	67	0.82	0.82	0.82	0.65	0.71
Country Fixed Effects										
Log GDP_PC WDI	0.55***					0.62***				
	(0.096)					(0.141)				
Log GDP_PC UQICD		0.50***					0.618***			
		(0.094)					(0.139)			
Log GDP_PC PWT 7.1			0.53***					0.644***		
			(0.098)					(0.151)		
Log GDP_PC PWT 8.1				0.44***					0.59***	
				(0.113)					(0.144)	
Log GDP_PC PWT 9.0					.42***					.50***
					(.09)					(.12)
R2	0.97	0.97	0.97	0.97	0.97	0.95	0.95	0.95	0.95	0.95
	Year and Country Fixed Effects									
Log GDP_PC WDI	0.64***					0.33				
	(0.155)					(0.336)				
Log GDP_PC UQICD		0.51***					0.36			
		(0.153)					(0.381)			
Log GDP_PC PWT 7.1			0.57***					0.363		
			(0.152)					(0.367)		
Log GDP_PC PWT 8.1				0.32*					0.383	
				(0.184)					(0.429)	
Log GDP_PC PWT 9.0					.35**					.317
					(.174)					(.347)
R2	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.96	0.95	0.95
Number of Obs	680	680	680	680	680	680	680	680	680	680
Number of Cluster	115	115	115	115	115	115	115	115	115	115

Table 5.1 presents the OLS Estimates of log nighttime lights per capita on log GDP per capita from WDI, PWT 7.1, PWT 8.1, PWT 9.0 and UQICD (columns 1-5). Dependent Variable: log lights per capita. Columns 6-10 provides results for OLS estimates of log survey means on log GDP per capita from WDI, PWT 7.1, PWT 8.1, PWT 9.0 and UQICD using log lights per capita as an instrument variable. Standard errors are in parentheses and clustered by country. Data on nighttime lights, GDP per capita (WDI) and survey means are from the paper "*Lights, Camera,...Income! Illuminating the National Accounts- Household Surveys Debate*" (Pinkovskiy and Sala-i-Martin, 2016a). Data on GDP per capita (UQICD) is from the University of Queensland. PWT 7.1, PWT 8.1 and PWT 9.0 downloaded from http://www.rug.nl/research/ggdc/data/pwt/.

The scatterplots of data in Figures 5.1-5.5 in Results Appendix confirm the strong linear relationship between log lights per capita and log GDP per capita in five datasets. We can see that the slopes of GDP per capita from all sources of data are larger than the slopes of survey means ($\beta^G > \beta^S$), impling that the variance increases at higher rates of log GDP per capita and log surveys means. However, we also see that the distances between the linear lines based on log GDP per capita from WDI, PWT 7.1, UQICD (Figures 5.1-5.3) and on log survey means are wider than the distance between the trendline of log GDP per capita from PWT 8.1, PWT 9.0 and the trendline of survey means (Figures 5.4-5.5). We can test the hypothesis $\beta^S/\beta^G < 1$ by using log light per capita as an instrumental variable for the regressor log GDP per capita from five alternative datasets in the scalar regression model

$$y_{i,t}^S = b^0 + b^G y_{i,t}^G + \varepsilon_{i,t} \tag{33}$$

Columns 6-10 in Table 5.1 provide the estimates of equation (33) with regard to four different types of fixed effects. It is important to note that using five different GDP series does not impact on P&S's findings. Specifically, whatever GDP series we use, we always reject the null hypothesis that $\beta^S/\beta^G = 1$.

Table 5.2 shows the estimates of coefficient b^G and b^S of the bivariate regression of log lights per capita on log GDP per capita from all five datasets and log survey means. They are the point estimates of b^G and b^S in equation (12) in Chapter 4

$$y_{i,t}^L = b^0 + b^G y_{i,t}^G + b^S y_{i,t}^S$$

As mentioned in Chapter 4, the coefficients b^G and b^S satisfy equation 13 in Chapter 4

$$\frac{b^G}{b^G + b^S} = \frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*}$$

In other words, the left hand side of equation (13) is the relative weight of log GDP per capita and the best linear unbiased estimator of log true income per capita. As indicated in columns 1-3 of Table 5.2, it is clear that when using WDI, PWT 7.1 or UQICD datasets, the coefficients of GDP per capita show a similar pattern. In particular, without any fixed effects included, the estimates of b^G are slightly lower than the simple regressions (see columns 1-3 of Table 5.1) and significant at 1%, whereas the estimates of b^S plunge to just one fifth of b^G and not significant at 1% (see row 1, columns 1-3 of Table 5.2). When regressions with country fixed effect are

included (see rows 3-4, columns 1-3 of Table 5.2), the coefficients on log GDP per capita b^G show decreases in magnitude to a certain degree but are still significant at 1%, the coefficients on log survey means are negative and insignificant. By contrast, the use of PWT 8.1 and PWT 9.0 changes the results considerably. Specifically, in the no fixed effects and year fixed effects specifications, the coefficients on log survey means are larger than the coefficients on log GDP per capita and they are both significant at 1% (see rows 1-2 and columns 4-5 of Table 5.2) . As for country fixed effects regression, while the estimate of b^G slightly declines it is still significant, the coefficient on survey means drops sharply and is insignificant (see row 3 and columns 4-5 of Table 5.2). When all fixed effects are included, both estimates of b^G and b^S are insignificant (see row 4 and columns 4-5 of Table 5.2).

To sum up, in this section GDP series from WDI dataset in P&Sa's paper is replaced by GDP series from UQICD, PWT 7.1, PWT 8.1 and PWT 9.0 in replicating columns 1, 3 and 4 of Table I in Pinkovskiy and Sala-i-Martin (2016a, page 38). The results show that while using UQICD, PWT 7.1 datasets does not change the findings of P&S, these findings can change considerably when using PWT 8.1 and PWT 9.1 datasets.

Table 5.2: Bivariate regression - Replicate column (4) of Table I (Pinkovskiy and Sala-i-Martin 2016a, page 38)

Baseline Regressions (cont)					
	(1)	(2)	(3)	(4)	(5)
	Log Lights OLS	Log Lights OLS	Log Lights OLS	Log Lights OLS	Log Lights OLS
No Fixed Effects					
Log GDP_PC (WDI)	1.01***(0.133)				
Log Survey Mean	0.204(0.139)				
Log GDP_PC (UQICD)		0.976***(0.135)			
Log Survey Mean		0.253*(0.135)			
Log GDP_PC (PWT7.1)			1.036***(0.1256)		
Log Survey Mean			0.166(0.137)		
Log GDP_PC (PWT8.1)				0.507***(0.141)	
Log Survey Mean				0.774***(0.148)	
Log GDP_PC (PWT9)					.601***(.171)
Log Survey Mean					.665***(.170)
R2	0.73	0.73	0.74	0.68	0.69
Year Fixed Effects					
Log GDP_PC (WDI)	0.988***(0.135)				
Log Survey Mean	0.240*(0.14)				
Log GDP_PC (UQICD)		0.964***(0.14)			
Log Survey Mean		0.281**(0.139)			
Log GDP_PC (PWT7.1)			1.008***(0.126)		
Log Survey Mean			0.210(0.137)		
Log GDP_PC (PWT8.1)				0.491***(0.142)	
Log Survey Mean				0.804***(0.146)	
Log GDP_PC (PWT9)					.592***(.180)
Log Survey Mean					.692***(.175)
R2	0.75	0.75	0.76	0.71	0.72
Country Fixed Effects					
Log GDP_PC (WDI)	0.597***(0.135)				
Log Survey Mean	-0.058(0.097)				
Log GDP_PC (UQICD)		0.511***(0.134)			
Log Survey Mean		-0.012(0.098)			
Log GDP_PC (PWT7.1)			0.55***(0.141)		
Log Survey Mean			-0.027(0.104)		
Log GDP_PC (PWT8.1)				0.42***(0.161)	
Log Survey Mean				0.039(0.122)	
Log GDP_PC (PWT9)					.41***(.137)
Log Survey Mean					.01(.116)
R2	0.97	0.97	0.97	0.97	0.97
Year and Country Fixed Effects					
Log GDP_PC (WDI)	0.664***(0.183)				
Log Survey Mean	-0.039(0.106)				
Log GDP_PC (UQICD)		0.512***(0.183)			
Log Survey Mean Income		-0.0021(0.110)			
Log GDP_PC (PWT7.1)			0.577***(0.181)		
Log Survey Mean Income			-0.011(0.11)		
Log GDP_PC (PWT8.1)				0.31(0.202)	
Log Survey Mean Income				0.04(0.118)	
Log GDP_PC (PWT9)					.35*(.187)
Log Survey Mean Income					.04(.109)
R2	0.98	0.98	0.98	0.98	0.98
Number of Obs	680	680	680	680	680
Number of Clusters	115	115	115	115	115

Table 5.2 shows the coefficients of a bivariate regression of log light per capita on log GDP per capita from WDI, UQICD, PWT 7.1, PWT 8.1 and PWT 9.0 and corresponding log survey means. Standard errors are in parentheses and clustered by country.

5.2 Estimates of Relative Weights

In this section, I will use GDP per capita and consumption per capita from UQICD, PWT 7.1, PWT 8.1 and PWT 9.0 instead of GDP per capita and consumption per capita from WDI to replicates Table II (Pinkovskiy and Sala-i-Martin 2016a, page 39) and then compare the results with the original paper. The results are shown in Tables 5.3-5.7 in Results Appendix. For the comparisons and consistency between difference sources of data, I used a smaller sample size compared to the original paper. (For more detail, see Tables 5.3 - 5.7 in Results Appendix).

The true relative weight of log GDP per capita defined as

$$w_G^* = \frac{\gamma_G^*}{\gamma_G^* + \gamma_S^*}$$

while the relative weight of log survey means is given by, $w_S^* = 1 - w_G^*$

As mentioned in Chapter 4, the relative weight of log GDP per capita can be estimated by using the coefficients of equation (12)

$$y_{i,t}^L = b^0 + b^G y_{i,t}^G + b^S y_{i,t}^S$$

The best linear unbiased estimator is

$$\widehat{w}_G = \frac{b^G}{b^G + b^S}$$

and the corresponding relative estimated survey means ratio is, $\widehat{w}_S = 1 - \widehat{w}_G$.

Previous studies have calculated the relative weight of GDP per capita and survey means differently. While some researchers used solely GDP per capita by assuming that $\gamma_G^* = 1$, or equivalent $w_G^* = 1$ (see literature in Barro (1991)), others used solely survey means by assuming that $\gamma_G^* = 0$, or equivalent $w_G^* = 0$ (Chen and Ravallion 2004, Milanović 2005). A study by Chen and Ravallion (2010), used a simple regression of log survey mean consumption on log national accounts consumption and a constant to measure income per capita. They concluded that $w_G^* < 0.5 < w_S^*$ (Pinkovskiy and Sala-i-Martin 2016a, page 17)[8].

[8] Chen and Ravallion (2010) used national accounts consumption from WDI dataset

The main purpose of this study is to test the robustness of the null hypothesis that $w_G^* = 1$ that was found by P&Sa will not change if we alter GDP series from WDI by GDP series from UQICD, PWT 7.1, PWT 8.1 and PWT 9.0. In particular, I will test the hypothesis

$$w_G^*(WDI) = w_G^*(UQICD) = w_G^*(PWT7.1) = w_G^*(PWT8.1) = w_G^*(PWT9.0) = 1$$

Column 1 of Tables 5.3-5.7 in the Results Appendix show the coefficient estimates of w_G^* and w_S^* from the baseline regression equation (12) using aggregate data with four types of fixed effects. The results suggest that the UQICD and PWT 7.1 datasets do not show different results from those using WDI dataset when estimating the relative weights on log GDP per capita and log survey means. To be more precise, we always fail to reject the null hypothesis that $w_G^* = 1$, or that GDP per capita contribute 100% weight, while the weight for survey are equal to zero regardless of the specification of effects included. Moreover, we can also reject the null hypothesis that $w_G^* = 0.5$, or surveys and GDP per capita from WDI, UQICD and PWT 7.1 have the equal weights (see column 1 of Tables 5.1-5.3 in appendix). By contrast, using GDP series from PWT 8.1 and PWT 9.0 partially change P&S's findings. In particular, for the regressions without country fixed effects, we can reject the null hypothesis that $w_G^* = 1$. This result is opposite to P&S's. However, in the specifications with country fixed effects, the null hypothesis that $w_G^* = 1$ cannot be rejected, even though we fail to reject the null hypothesis that $w_G^* = 0.5$. For example, in column 1, row 2 and row 4 of Table 5.6, the estimates weight of w_G^* increases from 0.47 with the confidence interval of (0.23, 0.7) (no fixed effect) to 0.88 with the confidence interval of (0.2, 1.4) (country-year fixed effects)[9]. In this case, we fail to reject the null hypothesis that $w_G^* = 1$ as well as the null hypothesis $w_G^* = 0.5$.

To sum up, with aggregate data, PWT 7.1 and UQICD behave similarly to WDI, while PWT 8.1 and PWT 9.0 give results similar to each other, but different to those of the three other datasets.

5.3 Robustness check

In order to check the robustness of their inferences based on baseline regression equation (12), P&Sa have used different methods. In particular, P&Sa strengthens their conclusions in four directions. Firstly, they added new controls to make the baseline regression equation better.

[9] In the no fixed effect specification, we reject the null hypothesis that $w_G^* = 1$ because the confidence interval does not contain the number 1. Similarly, in the country fixed effect specification, we fail to reject the null hypothesis that $w_G^* = 1$ as well as the null hypothesis $w_G^* = 0.5$ because the confidence interval contains both these numbers.

These controls aim to investigate whether or not the assumption A2 is violated. Specifically, to check whether there is a relationship between GDP per capita and survey means with lights due to other causes than joint correlation with true income (see columns 2-4 of Tables 5.3-5.7 in the Results Appendix). Secondly, P&S changed the dependent variable (aggregate lights per capita) of the baseline regression equation (12) by three alternative measures of lights, including light density (aggregate lights per area), calibrated lights and fraction population light (see columns 5-7 of Tables 5.3-5.7 in the Results Appendix). Thirdly, after checking the dependent variables, P&S continued to check the independent variables by using an alternative to GDP per capita (WDI) and survey measures. They argued that surveys data are not homogeneous since some measure consumption and others measures income, and national accounts on consumption better capture disposable income than GDP per capita. Therefore, it is more comparable to use surveys consumption with national accounts consumption, and surveys income with national accounts GDP per capita.[10] The results from the estimation are presented in a graphical form through Figures 5.6 to 5.9.

Tables 5.3-5.7 shown in the Results Appendix show the detailed estimation output which has been used to obtained the graphs. In all cases the confidence interval from the base model is shown as the dotted lines. The computed weight for consumption for alternative data sets (PWT 7.1, WDI, PWT 8.1, PWT 9.0 and UQICD) is presented for each model specification. Figure 5.6 presents the results for models with no fixed effects. Figure 5.7 presents the results for models with year fixed effects; Figure 5.8 presents estimates for model with country fixed effects; and Figure 5.9 plots the models with both year and country fixed effects.

Overall, while WDI, PWT 7.1 and UQICD show a similar pattern, PWT 8.1 and PWT 9.0 indicate a different pattern. Figure 5.6 shows in the no fixed effects specification, the point estimates of w_G^* fluctuate around 1 and they lie inside or above baseline confidence interval. In addition, as can be seen in row 1 of Tables 5.3-5.5 in Results Appendix, almost all estimates of w_G^* are statistically significant at 1%, while almost all estimates of w_S^* are insignificant. Moreover, since the confidence intervals contain the number 1 in almost cases, we fail to reject the null hypothesis that $w_G^* = 1$. By contrast, for PWT 8.1 (row 1 of Table 5.6 in the Results Appendix), we can see that 5 out of 12 regression coefficients of w_G^* are insignificant, while seven out of 12 regression coefficients of w_S^* are significant. Figure 5.6 shows six out of 12 point

[10] For more detail, see Pinkovskiy and Sala-i-Martin (2016a, page 19-23)

estimates of w_G^* lie below the lower bound of baseline confidence interval. Hence, we can reject the null hypothesis that $w_G^* = 1$ in almost all cases. Likewise, the behavior of PWT 9.0 is similar to PWT 8.1 (see row 1 of Table 5.7 and Figure 5.6). When year fixed effects are used, Figure 5.7, the same pattern persists (see Figure 5.7 and row 2 of Tables 5.3-5.7 in the Results Appendix).

Figure 5.6 Models with no fixed effects

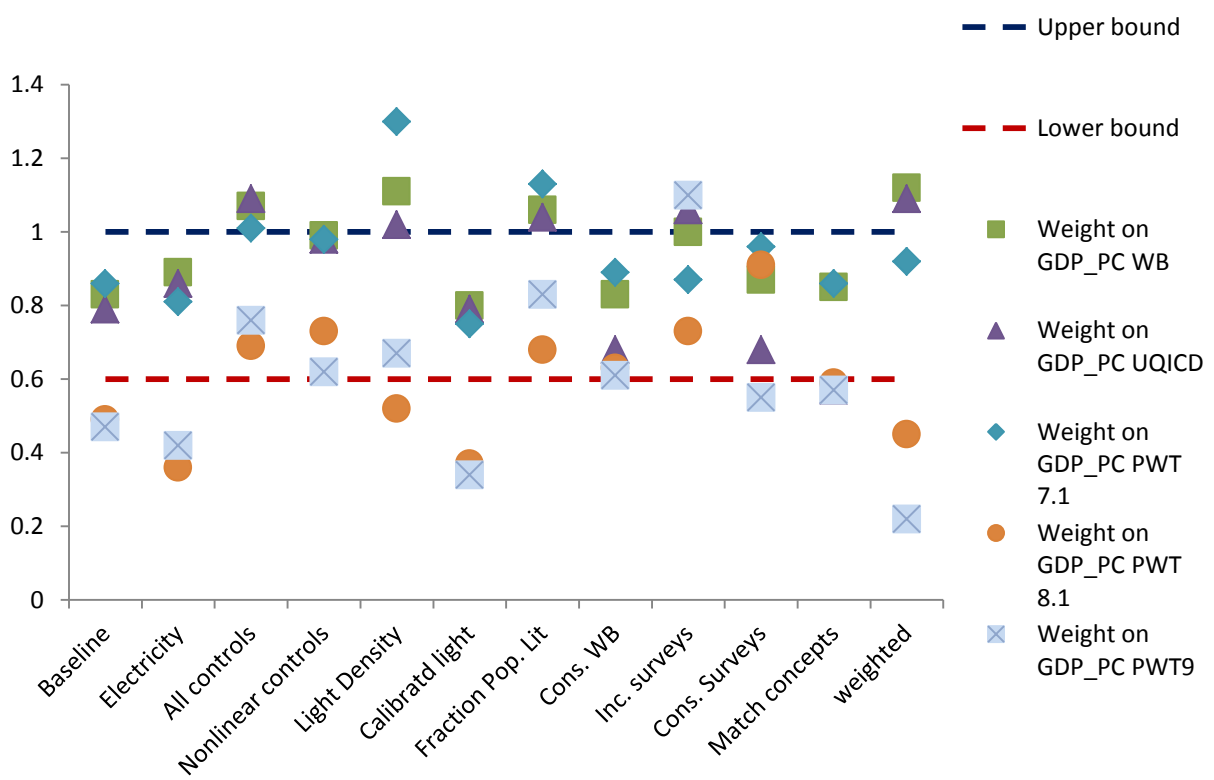


Figure 5.6 illustrates the results of statistics $\widehat{w}_G$ for 12 different regressions from row 1 of table 5.3-5.7 in the no fixed effects specification. In particular, we used 95% confidence interval of the estimates of w_G^* from baseline regressions equation 12 using GDP series from WDI (see row 1 and column 1 of table 5.3) as a benchmark to compare different sources of GDP per capita. Accordingly, the dashed blue line represents upper bound of confidence interval, while the dashed red line represents lower bound of confidence interval. In the next step, we plot the values of $\widehat{w}_G$ that are the coefficients of 12 different regressions. There are five colors, each color represents for a particular source of data.

Figure 5.7 Models with year fixed effects

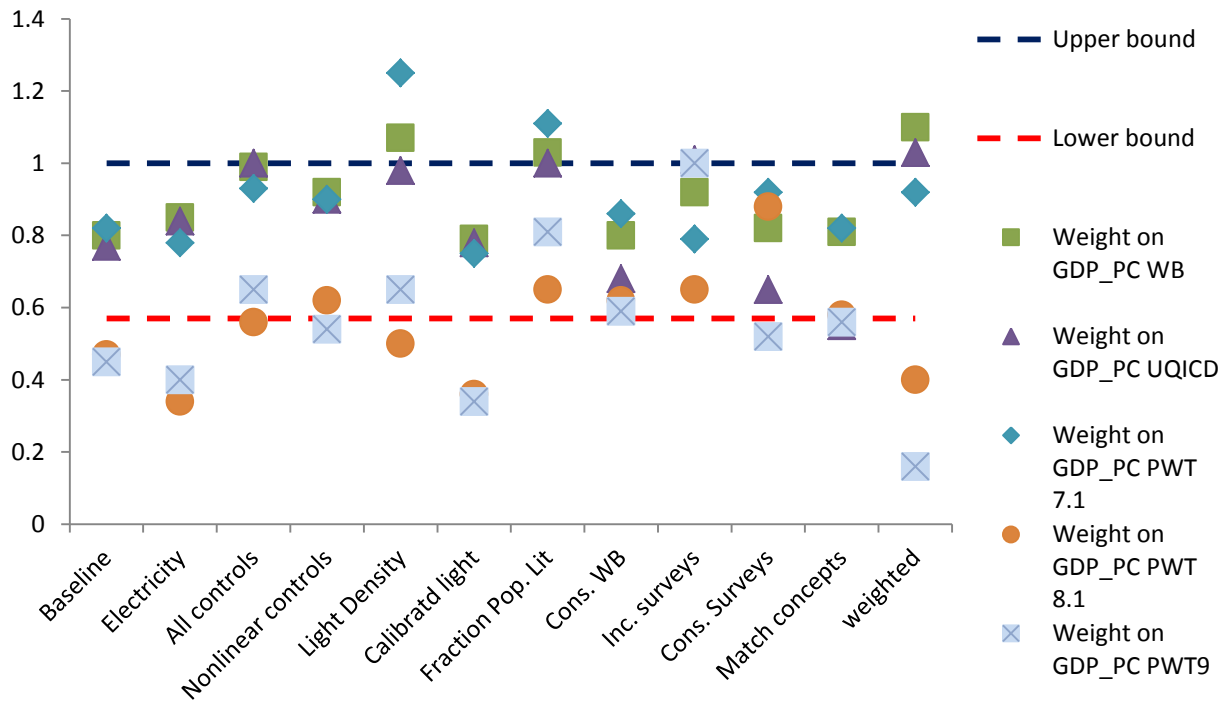


Figure 5.7 illustrates the results of statistics w_G^* for 12 different regressions from row 2 of Table 5.3-5.7 in the year fixed effects specification. In particular, we used 95% confidence interval of the estimates of w_G^* from baseline regressions equation 12 using GDP series from WDI (see row 2 and column 1 of Table 5.3) as a benchmark to compare different sources of GDP per capita. Accordingly, the dashed blue line represents upper bound of confidence interval, while the dashed red line represents lower bound of confidence interval. In the next step, we plot the values of w_G^* that are the coefficients of 12 different regressions. There are five colors, each color represents a particular source

However, specifications with country fixed effects show a completely different picture. Figure 5.8 shows the results are consistent across consumption measures and thus do not change P&Sa's findings. In particular, we nearly always fail to reject the null hypothesis that $w_G^* = 1$. The point estimates of w_G^* using the five different GDP series have a tendency to converge to one point. Moreover, these points are well inside the baseline confidence interval. As for PWT 8.1 and PWT 9.0, although in 3 out of 12 regressions, the point estimates lie below confidence interval, we still fail to reject the null hypothesis that $w_G^* = 1$ in these cases (see row 3 and columns 7, 9, 12 of Tables 5.6 and 5.7 in the Results Appendix). When we include both year and country fixed effects, the results reinforce P&Sa's finding. We can see in Figure 5.9, the estimates of w_G^* from all datasets are very close to 1 in almost regressions. In general, these results are therefore consistent with P&Sa's finding in the year-country specification.

In summary, unless country fixed effects are included in the model PWT 8.1 and PWT 9.0 behave differently and partially change P&S's findings, whereas WDI, PWT 7.1 and UQICD are consistent under all specifications.

Figure 5.8 Models with country fixed effects

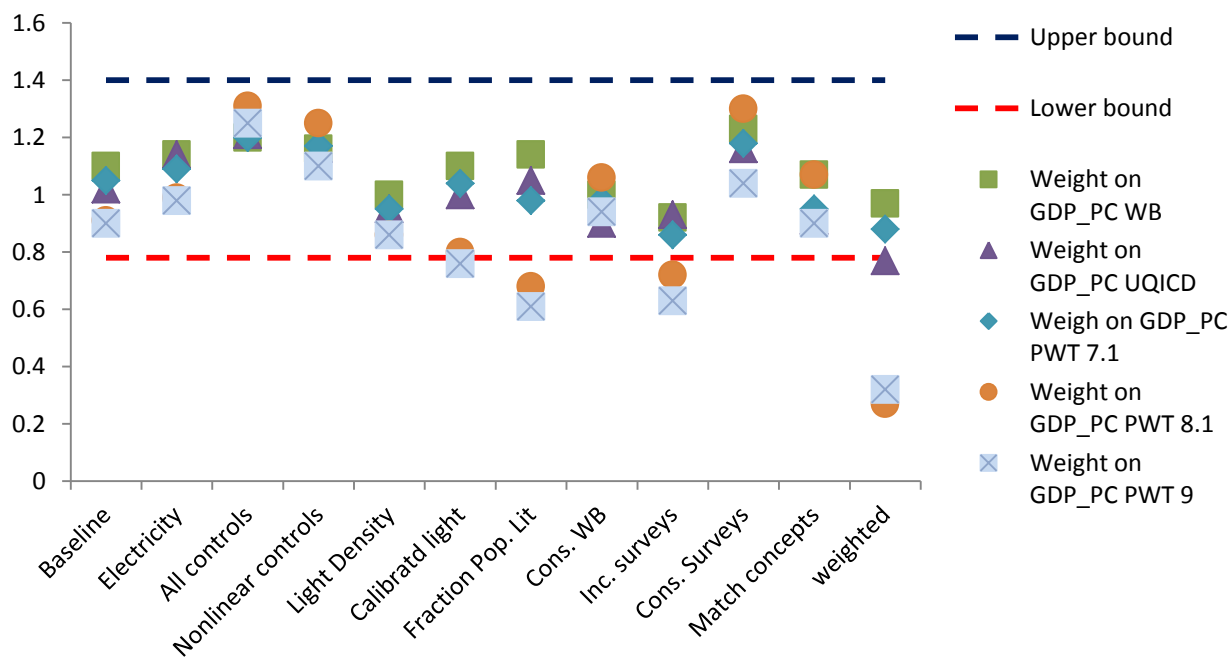


Figure 5.8 illustrates the results of statistics w_G^* for 12 different regressions from row 3 of table 5.3-5.7 in the country fixed effects specification. In particular, we used 95% confidence interval of the estimates of w_G^* from baseline regressions equation 12 using GDP series from WDI (see row 3 and column 1 of table 5.3) as a benchmark to compare different sources of GDP per capita. Accordingly, the dashed blue line represents upper bound of confidence interval, while the dashed red line represents lower bound of confidence interval. In the next step, we plot the values of w_G^* that are the coefficients of 12 different regressions. There are five colors, each color represents for a particular source of data.

Figure 5.9 Models with year and country fixed effects

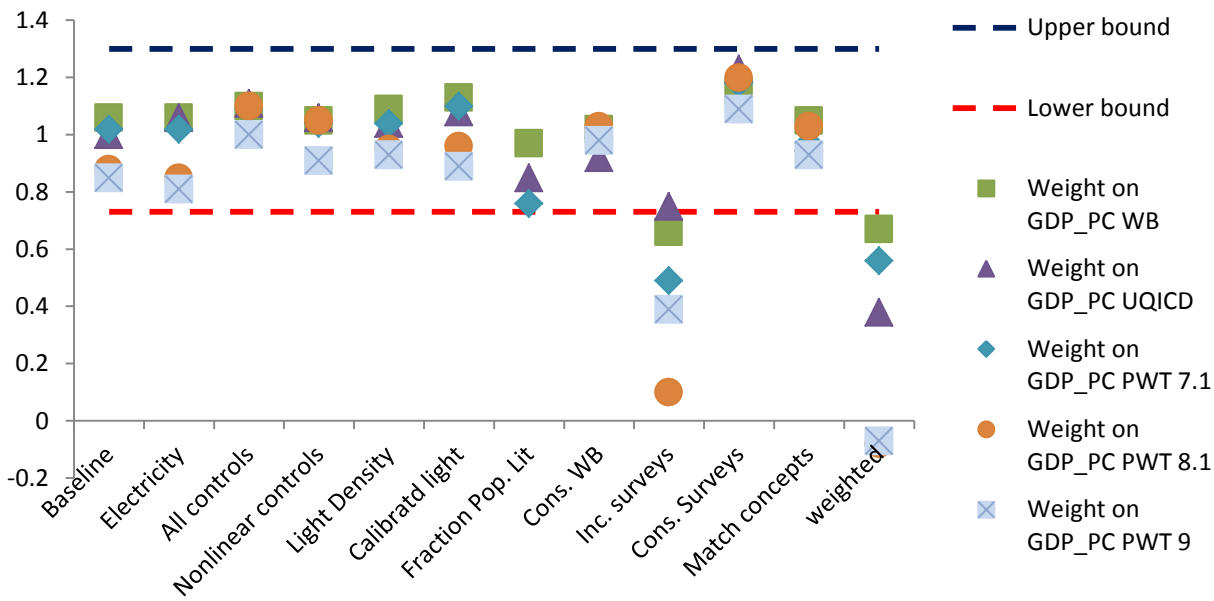


Figure 5.9 illustrates the results of statistics w_G^* for 12 different regressions from row 4 of table 5.3-5.7 in the year and country fixed effects specification. In particular, we used 95% confidence interval of the estimates of w_G^* from baseline regressions equation 12 using GDP series from WDI (see row 4 and column 1 of table 5.3) as a benchmark to compare different sources of GDP per capita. Accordingly, the dashed blue line represents upper bound of confidence interval, while the dashed red line represents lower bound of confidence interval. In the next step, we plot the values of w_G^* that are the coefficients of 12 different regressions. There are five colors, each color represents for a particular source of data.

5.4 Combination forecast approach

In this section, I will present the results of a combination forecast approach to estimating the relative weight of log GDP per capita and log survey means. The estimates of log survey means can be obtained from equation 31 in Chapter 4

$$y_{i,t}^{L} - y_{i,t}^{G} = (1 - \widehat{w}_G)(y_{i,t}^{S} - y_{i,t}^{G})$$

Or

$$y_{i,t}^{L} - y_{i,t}^{G} = \widehat{w}_S(y_{i,t}^{S} - y_{i,t}^{G})$$

and the corresponding estimates of log GDP per capita is $\widehat{w}_G = 1 - \widehat{w}_S$. The notation is the same as that in Sections 4.1-4.2 in Chapter 4.

In this chapter, I will use four types of measures of NTL including aggregate lights per capita, light density, calibrated lights and fraction population lights in the regressions[11]. Also, I will use GDP per capita data from five alternative sources: WDI, UQICD, PWT 7.1, PWT 8.1 and PWT 9.0. The results are provided in Tables 5.8-5.12 in the Results Appendix.

At the outset we note that combination forecast estimates do not change the findings of P&Sa if we use GDP per capita data from WDI, UQICD and PWT 7.1. As Figure 5.10 shows, in comparison to P&Sa's method, the point estimates of w_G^* are slightly higher for three out of four regressions in the specification with no fixed effects (for more details see row 1 of Table 5.8 in the Results Appendix). When country fixed effects are included, Figure 5.11 shows that if dependent variables are log light per capita and log light density, the results reverse between no fixed effects and country fixed effects, while the remaining two measures of NTL show the same pattern (for more details see row 3 of Table 5.8 in the Results Appendix). However, these differences do not influence our inferences. Table 5.13 summarizes the conclusions when we test the null hypothesis $w_G^* = 1$. In their approach, P&Sa drew conclusions based on the confidence intervals that are computed from bootstrapping technique. Since confidence intervals always contained the number 1, we always failed to reject the null hypothesis that $w_G^* = 1$ (see columns 1-4 of Table 5.8 in the Results Appendix). For the combination forecast approach we used t-test

[11] All these measures of lights are described clearly in Pinkovskiy and Sala-i-Martin (2016a, page 21).

to test the null hypothesis that $w_G^* = 1$[12]. Table 5.15 in the Results Appendix shows the p-value statistics when we use t-test to test the null hypothesis $w_G^* = 1$. We can see that p-value statistics are always larger than 0.05 (p-value > 0.05). Therefore, we always fail to reject the null hypothesis that $w_G^* = 1$, that is, the conclusion we draw is that GDP per capita alone should be used to predict the true income, or that survey means should have a weight of zero at 5% level of significance. In general, using GDP data from WDI, UQICD and PWT 7.1, either P&Sa or combination forecast yield the same findings (see more detail in Tables 5.9, 5.10, 5.16 and 5.17 in the Results Appendix).

By contrast, using GDP data from PWT 8.1 and PWT 9, the combination forecast method has shown some different results compared to P&Sa's approach. As indicated in Figures 5.12-5.13, the first difference between using WDI data and PWT 8.1 data is that the point estimates of $\widehat{w}_G$ are significantly smaller when using PWT 8.1 compared to WDI data. At this point, P&Sa's method and the combination forecast are the same. However, we can see certain differences in more detail in Table 5.14. With P&Sa's method we reject the null hypothesis $w_G^* = 1$ in eight out of 16 specifications (see columns 1-4 of Table 5.11 in the Results Appendix), while with the combination forecast, we reject the null hypothesis $w_G^* = 1$ in 10 out of 16 specifications (see Table 5.18 in the Results Appendix). Hence, using PWT 8.1, we can conclude that the combination forecast method and P&Sa's method provide inconsistent results and significantly change P&Sa's findings. Likewise, using GDP data from PWT 9.0, these results are similar as when using PWT 8.1 (see more detail in Table 5.12 and Table 5.19 in the Results Appendix).

To sum up, it is clear that using a combination forecast approach generally provides the same findings as P&Sa's method if we use GDP per capita from WDI, UQICD and PWT 7.1. Using PWT 8.1 and PWT 9.0, the results are inconsistent between the two methods and significantly change P&Sa's findings. Once again, these results confirm the fact that PWT 8.1 and PWT 9.0 behave differently from the three other sources of GDP per capita.

[12] We can also directly obtain the estimated weight of log GDP per capita from equation $y_{i,t}^L - y_{i,t}^S = \widehat{w}_G(y_{i,t}^G - y_{i,t}^S)$. Hence, we can use t-test to test the null hypothesis that $w_G^* = 1$.

Table 5.13 Summary of the results when we test the null hypothesis $w_G^* = 1$ using P&Sa's method and Combination forecast method – WDI data

	P&S method Dependent variables				Combination forecast Dependent variables			
	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit
No Fixed Effects	DNR	DNR	DNR	DNR	DNR	DNR	DNR	DNR
Year Fixed Effects	DNR	DNR	DNR	DNR	DNR	DNR	DNR	DNR
Country Fixed Effects	DNR	DNR	DNR	DNR	DNR	DNR	DNR	DNR
Year and Country Fixed Effects	DNR	DNR	DNR	DNR	DNR	DNR	DNR	DNR

Note: DNR = Do not reject. P&S method used confidence intervals that were constructed from bootstrapping method to draw conclusions, while combination forecast method used t-test to test the null hypothesis $w_G^* = 1$ (See p-value statistic in Table 5.15 in the Results Appendix and STATA code in the Code Appendix)

Table 5.14 Summary the results when we test the null hypothesis $w_G^* = 1$ using P&Sa's method and Combination forecast method – PWT 8.1 data

	P&S method Dependent variables				Combination forecast Dependent variables			
	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit
No Fixed Effects	R	R	R	DNR	R	DNR	R	R
Year Fixed Effects	R	R	R	DNR	R	DNR	R	R
Country Fixed Effects	DNR	DNR	DNR	DNR	R	DNR	DNR	R
Year and Country Fixed Effects	DNR	R	DNR	R	DNR	R	DNR	R

Note: R= Reject, DNR = Do not reject. P&S method used confidence intervals that were constructed from bootstrapping method to draw conclusions, while combination forecast method used t-test to test the null hypothesis $w_G^* = 1$ (See p-value statistic in Table 5.18 in the Results Appendix and STATA code in the Code Appendix)

Figure 5.10 WDI-No Fixed Effects

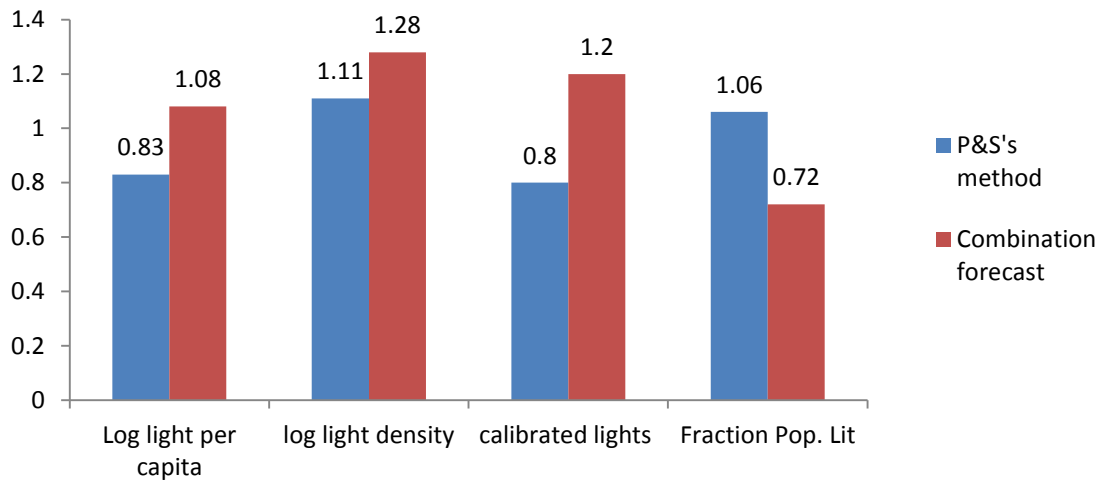


Figure 5.10 Comparision the estimates of relative weight on GDP per capita (WDI) between P&S's method and Combination forecast approach - No fixed effects

Figure 5.11 WDI-Country Fixed Effects

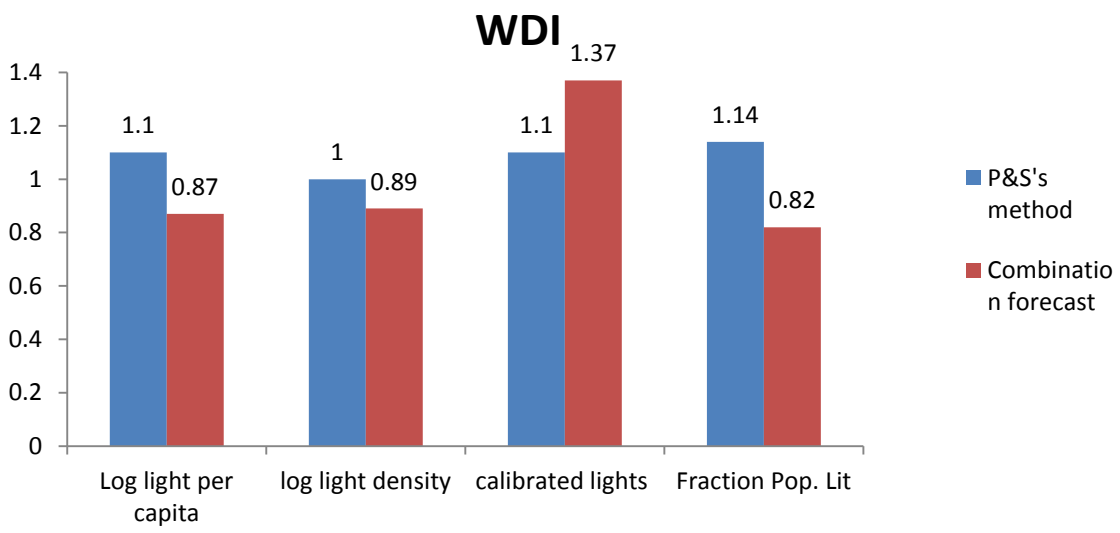


Figure 5.11 Comparision the estimates of relative weight on GDP per capita (WDI) between P&S's method and Combination forecast approach - Country fixed effects

Figure 5.12 PWT 8.1-No Fixed Effects

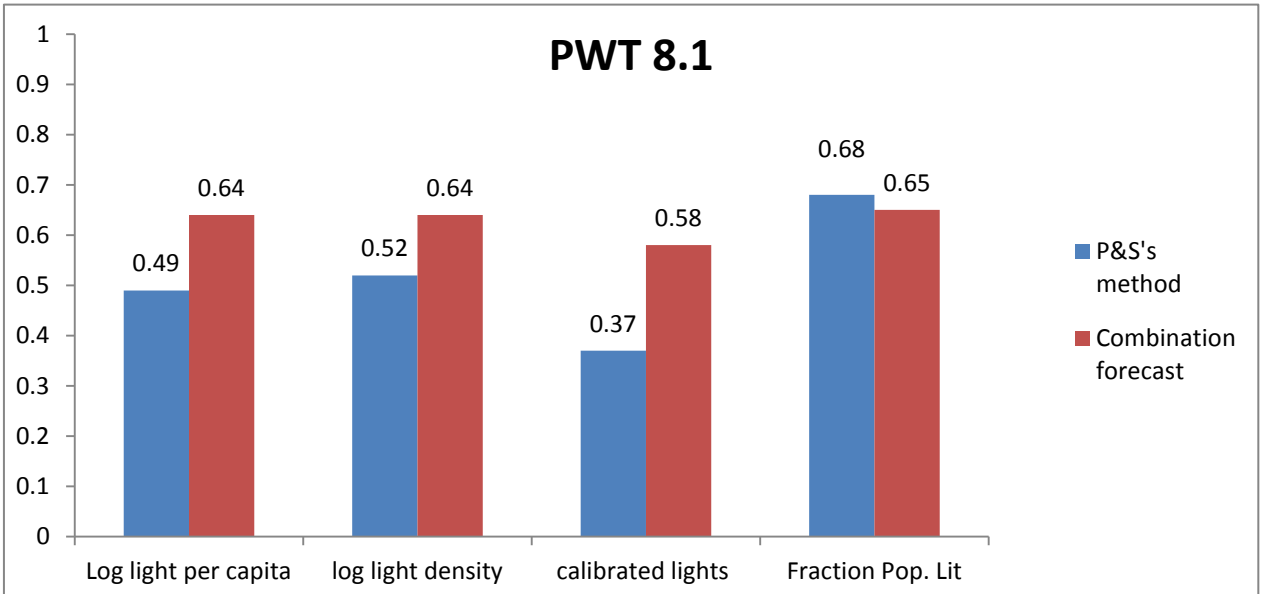


Figure 5.12 Comparision the estimates of relative weight on GDP per capita (PWT 8.1) between P&S's method and Combination forecast approach - No fixed effects

Figure 5.13 PWT 8.1-Country Fixed Effects

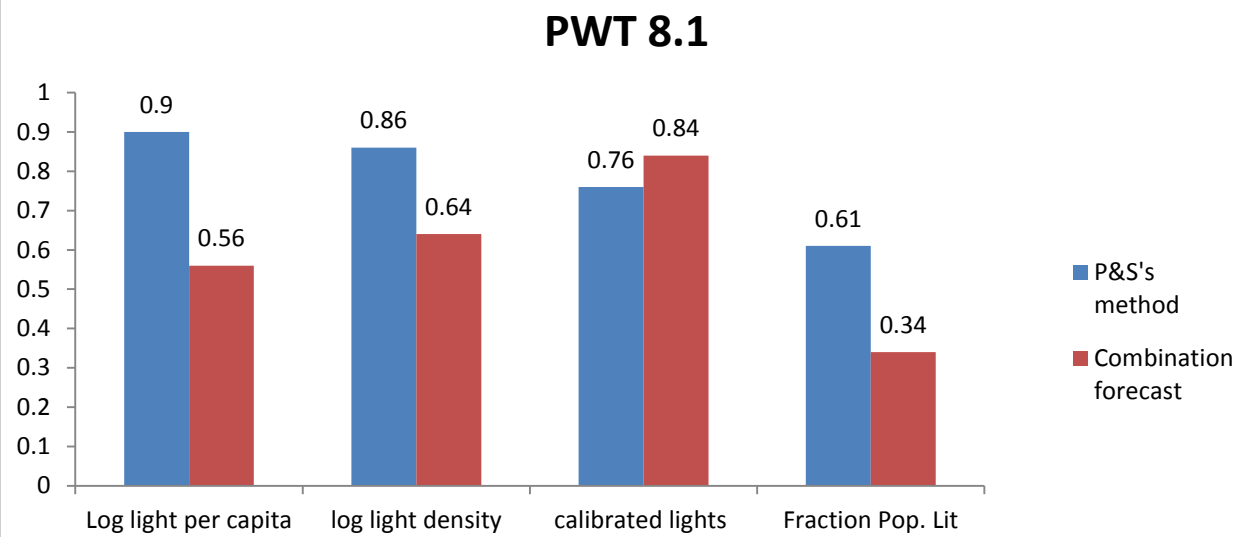


Figure 5.13 Comparision between P&S's method and Combination forecast approach using GDP per capita from PWT 8.1 - Country fixed effects

Chapter 6 Group Models Results

In this chapter, we re-estimate P&Sa's models by World Bank income groups and geographical regions. In particular, we test the null hypothesis that GDP per capita gets full weight ($w_G^* = 1$) or survey means gets zero weight ($w_S^* = 0$) in measuring true income per capita will not change across income groups or region groups.

Table 6.1: Re-estimate P&S's model by World Bank income groups and geographical regions – WDI data

Weights in the Optimal Proxy: Subsample (WDI)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Income groups			Region groups		
		Low income	Lower Middle income	Upper Middle income	Asia and Pacific	Africa and Middle East	Latin America and Caribbean
No Fixed Effects							
Log GDP_PC WDI	.83***	.91	.61**	.85***	2.9	.91***	.93**
	(.6, 1.06)	(-2.5, 4.3)	(.13, 1.08)	(.4, 1.2)	(-28, 34)	(.58, 1.2)	(.16, 1.7)
Log Survey Mean	.16	.08	.38	.14	-1.9	.08	.06
	(-.06, .39)	(-3.3, 3.5)	(-.08, .86)	(-.29, .59)	(-33, 29)	(-.25, .41)	(-.70, .83)
Year Fixed Effects							
Log GDP_PC WDI	.80***	.72	.52**	.85***	2.6	.88***	.86**
	(.57, 1.03)	(-1.8, 3.2)	(.00, 1.05)	(.44, 1.2)	(-10, 15)	(.58, 1.1)	(.16, 1.5)
Log Survey Mean	.19*	.27	.47*	.14	-1.6	.11	.13
	(-.03, .42)	(-2.2, 2.8)	(-.05, .99)	(-.26, .55)	(-14, 11)	(-.18, .41)	(-.57, .83)
Country Fixed Effects							
Log GDP_PC WDI	1.1***	2.1	1.12***	1.09***	1.1***	1.21	.70***
	(.78, 1.4)	(-28, 32)	(.46, 1.7)	(.79, 1.3)	(.85, 1.5)	(-3.0, 5.4)	(.35, 1.06)
Log Survey Mean	-.1	-1.1	-.12	-.09	-.19	-.21	.29
	(-.43, .21)	(-31, 29)	(-.78, .53)	(-.39, .2)	(-.53, .14)	(-4, 4)	(-.06, .64)
Country and Year Fixed Effects							
Log GDP_PC WDI	1.06***	1.7	1.07*	.92***	.95	1.07*	.48
	(.73, 1.3)	(-17, 21)	(-.03, 2)	(.6, 1.2)	(-.98, 2.8)	(-.05, 2.2)	(-3.3, 4.2)
Log Survey Mean	-.06	-.77	-.07	.07	.04	-.07	.51
	(-.39, .26)	(-20, 18)	(-1.1, 1.03)	(-.25, .39)	(-1.8, 1.9)	(-1.2, 1.05)	(-3.2, 4.3)
Number of Obs	680	79	215	296	82	136	227
Number of Clusters	115	26	37	39	17	48	22

Note: Table 6.1 shows estimates of the relative weight of GDP per capita (WDI) and survey means in measuring true income per capita using P&S's method across income groups and region groups. Block-bootstrapped 95% confidence interval in parentheses. Column 1 is baseline specification using aggregate data (the same with column 1 of Table 5.3 and Table 5.8). Columns 2-4 present the results across income groups. Columns 5-7 show the results across region groups.

Tables 6.1-6.5 illustrate the estimates of relative weight on GDP per capita and survey means income across three income groups and three region groups[13]. It is important to note that our sample size is small in some groups. This may affect the precision of our inferences. For

[13] Table 6.2-6.5 are in Results Appendix

example, the low income group has only 79 observations in 26 clusters, and Asia & Pacific region group has only 82 observations in 17 clusters (see column 2 and column 5 of Table 6.1). Hence, we can see that the confidence intervals are too wide and almost all the statistics $\widehat{w}_G$ and $\widehat{w}_S$ of two these groups are insignificant. In general, for those estimates of w_G^* and w_S^* which are significant, we almost fail to reject the null hypothesis that $w_G^* = 1$, but we can reject the null hypothesis that $w_G^* = 1$ in some particular specifications. The results are fairly similar when we replace GDP per capita data from WDI by four other sources of GDP data (see Tables 6.2-6.5 in the Results Appendix). However, it is hard to draw a robust conclusion if we rely only on these results. Therefore, we propose a method to test the null hypothesis that $w_G^* = 1$ for individual group using dummy variables and interaction terms by an asymptotic valid Wald statistic. We recall equation (32)

$$y_{i,t}^L = D_1 + D_2 + D_3 + b^{G1}D_1y_{i,t}^{G1} + b^{S1}D_1y_{i,t}^{S1} + b^{G2}D_2y_{i,t}^{G2} + b^{S2}D_2y_{i,t}^{S2} + b^{G3}D_3y_{i,t}^{G3} + b^{S3}D_3y_{i,t}^{S3} + \varepsilon_{i,t}$$

Our null hypothesis will be

$$w_G^*(group\ i) = 1 \qquad \text{with } i = 1, 2, 3$$

Or

$$\frac{b^{Gi}}{b^{Gi} + b^{Si}} = 1$$

Table 6.6 reports the p-value from STATA "testnl" command (see Code Appendix). We test the null hypothesis that $w_G^* = 1$ for individual groups with various types of fixed effects specifications[14]. We can see immediately that the p-value is always larger than 0.01, or we always fail to reject the null hypothesis $w_G^* = 1$ for each income or region group at 1% level of significance. But at 10% level of significance, we can sometimes reject the null hypothesis $w_G^* = 1$ (0.01 < p-value < 0.1). For example, in the specifications without country fixed effects, the estimate of relative weight on GDP per capita in the lower middle income group is not equal 1 at the 10% levels of significance. However, although the results slightly change at different levels of significances, this fact does not affect our inferences. In general, the results suggest that the optimal weights on GDP per capita in estimating true income proxy are equal 1 and are the

[14] We use GDP per capita from WDI – Code shown in Code Appendix

same across income groups and region groups. Again, this finding did not change if we alter GDP per capita from WDI by GDP per capita from UQICD and PWT 7.1 (see Tables 6.7-6.8 in the Results Appendix).

Table 6.6: The estimates of p-value using Wald test – WDI data to test the null hypothesis $\boldsymbol{w_G^* = 1}$ for individual group

	Income groups			Region groups		
	Low income	Lower Middle Income	Upper Middle Income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects	0.7597	0.0628	0.5363	0.2631	0.6204	0.8447
Year Fixed Effects	0.5546	0.0395	0.5583	0.2932	0.4733	0.6987
Country Fixed Effects	0.4486	0.6036	0.6419	0.3133	0.6594	0.1505
Country and Year Fixed Effects	0.4145	0.5884	0.8303	0.8545	0.6188	0.0995

By contrast, if we use PWT 8.1 or PWT 9.0 in place of WDI, our findings will change significantly. As shown in Table 6.10, we reject the null hypothesis $w_G^* = 1$ in Latin American and Caribbean group at 1% level of significance (p-value < 0.01) in all four fixed effect specifications. Similarly, the null hypothesis that $w_G^* = 1$ is rejected in three out of four fixed effects specifications (p-value < 0.01) for Lower Middle Income groups, and the null hypothesis that $w_G^* = 1$ is rejected in two out of four fixed effects specifications at 5% level of significance (p-value < 0.05) for Africa and Middle East group. By contrast, for the three other groups, we always fail to reject the null hypothesis that $w_G^* = 1$ (p-value > 0.1). Therefore, the relative weights of GDP per capita are different between different income groups or region groups. Once again, if GDP per capita from PWT 9.0 is replaced by PWT 8.1, this conclusion did not change (see Table 6.9 in Results Appendix).

Table 6.10: The estimates of p-value using Wald test – PWT 9.0 data to test the null hypothesis $\boldsymbol{w_G^* = 1}$ for individual group

	Income groups			Region groups		
	Low income	Lower Middle Income	Upper Middle Income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects	0.4252	0.0000	0.1825	0.3270	0.0139	0.0093
Year Fixed Effects	0.5509	0.0000	0.1908	0.4037	0.0069	0.0077
Country Fixed Effects	0.4856	0.0065	0.8180	0.3353	0.4640	0.0000
Country and Year Fixed Effects	0.7000	0.1881	0.9101	0.9509	0.9318	0.0000

In conclusion, it is clear that we cannot reject our null hypothesis at the beginning of this chapter if we use GDP per capita data from WDI, UQICD and PWT 7.1. In particular, the results show that our key finding from the previous chapter that $w_G^* = 1$ is remarkably robust across different income groups and region groups. However, our null hypothesis will be rejected if we use GDP per capita from PWT 8.1 and PWT 9.0. Again, PWT 8.1 and PWT 9.0 behave differently from WDI, UQICD and PWT 7.1.

Chapter 7 Discussion and Conclusion

The purpose of this study was to investigate the robustness of P&S's findings that national accounts GDP per capita should get full weight and income measured by survey means should get zero weight in measuring true income (Pinkovskiy and Sala-i-Martin 2016a). The findings from this study suggest that if we use GDP series from WDI, UQICD or PWT 7.1 datasets, the results are largely the same and do not change P&Sa's findings. However, if we use GDP series from PWT 8.1 or PWT 9.0 datasets, the results are considerably different and partly change P&Sa's findings. The results also point out that by using WDI, UQICD or PWT 7.1 datasets, our findings are consistent with whichever method we use to estimate the weights: combination forecast method or P&Sa's method, or if we re-estimate P&Sa's model for income or geographical groups. Again, using PWT 8.1 or PWT 9.0 datasets results in inconsistent findings between P&Sa's method and the combination forecast method. Results also differ for income and geographical group models. In general, PWT 8.1 and PWT 9.0 always behave differently compared to WB, UQICD and PWT 7.1.

These findings concur with other studies that show different alternative measures of GDP will significantly impact on research outcomes. This implies that the quality of different measures of GDP per capita is different. Similarly, Pinkovskiy and Sala-i-Martin (2016b) also found that PWT 7.1 GDP show a different relationship to Night Lights to that with PWT 8.0 and PWT 8.1 GDP series. In the same vein, in their research, Ciccone and Jarociski (2010) found that using different versions of PWT will significantly change research results. Therefore, our findings are consistent with previous findings.

This is the first study, to our knowledge, to compare UQICD constant GDP per capita series to other similar sources of internationally comparable income. Despite using different methodologies in comparison with PWT 7.1 and WDI datasets, UQICD always behave similar to these two sources in our research. This finding indicates that UQICD appear to be a good source of income estimate. This is because many previous findings show that income estimates from WDI or PWT 7.1 are always better in capturing true income than other datasets. Moreover, we know that PWT 8.1 and PWT 9.0 have been constructed using identical methodology but different data (2005 ICP and both 2005 and 2011 ICP respectively), while PWT 7.1 and PWT 8.1 were based on the same data (2005 ICP) but different methodology. However, the results from using PWT 7.1 are different from results using PWT 8.1 and PWT 9.0. Therefore, our findings

suggest that the different ways of estimating PPP-adjusted GDP per capita are responsible for the different behaviors of different versions of PWT dataset.

The study possesses some limitations. Firstly, the above analysis does not enable us to determine which source of GDP per capita data is the best. Although we find that there are two groups which behave differently, we do not know which group performs better. Secondly, we have not examined the robustness of NTL as a new indicator of income data in measuring economic activity.

Several questions remain to be resolved. In particular, although NTL data have been used broadly in economic research, there is no research to check the robustness of nighttime lights as an independent benchmark in measuring true income. At this point, we suggest one way to check the reliability of NTL. In particular, we suggest to study the relationship between sources of GDP and consumption using Carbon dioxide (CO_2) emissions per capita instead of NTL. In fact, CO_2 emissions and luminosity can be interchangeable because of two reasons. It is obvious that CO_2 emissions and economic activity are strongly corelated, especially in developing countries. For example, many previous studies have proved the relationship between CO_2 emissions and economic growth in Algeria (Bouznit and Pablo-Romero, 2016), in China (Meng et al., 2012), or in Korea (Kim et al., 2010). The P&Sa's study to determine whether national accounts or household survey based GDP are more relevant can be repeated using CO_2 instead of NTL. The measurement errors of GDP and survey are uncorrelated with the measurement errors of CO_2 emissions and thus P&Sa's assumptions could not be violated. This is a feasible assumption since CO_2 emissions are calculated from fossil fuel consumption and world cement manufacturing (The World Bank Group, 2016), "*whereas GDP or survey data are obtained primarily or largely by asking people*" (Pinkovskiy and Sala-i-Martin 2016a, pages 10). Therefore, CO_2 emissions data can be used as a potential alternative source for NTL data.

References

ALAN HESTON, ROBERT SUMMERS & ATEN, B. Penn World Table Version 7.1. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, November 2012.

BARRO, R. J. 1991. Economic growth in a cross section of countries. (includes lists of variables as appendices). *Quarterly Journal of Economics,* 106**,** 407.

BICKENBACH, F., BODE, E., NUNNENKAMP, P. & SÖDER, M. 2016. Night lights and regional GDP. *Weltwirtschaftliches Archiv,* 152**,** 425-447.

BLUHM, R. & KRAUSE, M. 2016. Top Lights - Bright Spots and their Contribution to Economic Development. *The 34th IARIW General Conference.* Dresden, Germany.

BOUZNIT, M. & PABLO-ROMERO, M. D. P. 2016. CO2 emission and economic growth in Algeria. *Energy Policy,* 96**,** 93-104.

CHEN, S. & RAVALLION, M. 2004. How Have the Worlds Poorest Fared since the Early 1980s? *The World Bank Research Observer,* 19**,** 141-169.

CHEN, S. & RAVALLION, M. 2010. The developing world is poorer than we thought, but no less successful in the fight against poverty.(Report). *Quarterly Journal of Economics,* 125**,** 1577.

CHEN, X. & NORDHAUS, W. D. 2011. Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 8589.

CICCONE, A. & JAROCISKI, M. 2010. Determinants of Economic Growth: Will Data Tell. *American Economic Journal: Macroeconomics,* 2**,** 222-246.

DAWSON, J. W., DEJUAN, J. P., SEATER, J. J. & STEPHENSON, E. F. 2001. Economic information versus quality variation in cross-country data. *Canadian Journal of Economics/Revue canadienne d'économique,* 34**,** 988-1009.

DEATON, A. & ATEN, B. 2014. Trying to Understand the PPPs in ICP2011: Why are the Results so Different? Version 2015. Princeton University.

EBENER, S., MURRAY, C., TANDON, A. & ELVIDGE, C. C. 2005. From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics,* 4**,** 5-5.

ELVIDGE, C. D., BAUGH, K. E., ANDERSON, S. J., SUTTON, P. C. & GHOSH, T. 2012. The Night Light Development Index (NLDI): A spatially explicit measure of human development from satellite data. *Social Geography,* 7**,** 23-35.

ELVIDGE, C. D., BAUGH, K. E., KIHN, E. A., KROEHL, H. W., DAVIS, E. R. & DAVIS, C. W. 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing,* 18**,** 1373-1379.

ELVIDGE, C. D., SUTTON, P. S., GHOSH, T., TUTTLE, B. T., BAUGH, K. E., BHADURI, B. L. & BRIGHT, E. A. 2009. A Global Poverty Map Derived from Satellite Data. *Computers & Geosciences,* 35.

FEENSTRA, R. C., INKLAAR, R. & TIMMER, M. M. P. 2016. *What is new in PWT 9.0?* [Online]. The University of Groningen Available: http://www.rug.nl/research/ggdc/data/pwt/pwt-9.0 [Accessed 3 November 2016].

FEENSTRA, R. C., INKLAAR, R. & TIMMER, M. P. 2015. The Next Generation of the Penn World Table. *American Economic Review,* 105**,** 3150-3182.

HENDERSON, J. V., STOREYGARD, A. & WEIL, D. N. 2012. Measuring Economic Growth from Outer Space. *American Economic Review,* 102**,** 994-1028.

HESTON, A., SUMMERS, R. & ATEN, B. 2012. PWT 7.1: Penn World Table Version 7.1. *Center for International Comparisons of Production and Income and Prices at the University of Pennsylvania. Available from: http://pwt. econ. upenn. edu/php_site/pwt_index. php [accessed 21 September 2012].*

JOHNSON, S., LARSON, W., PAPAGEORGIOU, C. & SUBRAMANIAN, A. 2012. Is newer better? Penn World Table Revisions and their impact on growth estimates. *Journal of Monetary Economics,* 60, 255-274.

KEOLA, S., ANDERSSON, M. & HALL, O. 2015. Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth. *World Development,* 66**,** 322-334.

KIM, S.-W., LEE, K. & NAM, K. 2010. The relationship between CO 2 emissions and economic growth: The case of Korea with nonlinear evidence. *Energy Policy,* 38**,** 5938-5946.

MELLANDER, C., LOBO, J., STOLARICK, K. & MATHESON, Z. 2015. Night-time light data: A good proxy measure for economic activity? *PLoS ONE,* 10**,** <xocs:firstpage xmlns:xocs=""/>.

MENG, M., NIU, D. & SHANG, W. 2012. CO2 emissions and economic development: China's 12th five-year plan. *Energy Policy,* 42**,** 468.

MILANOVIĆ, B. 2005. *Worlds apart : measuring international and global inequality / Branko Milanovic,* Princeton, NJ, Princeton, NJ : Princeton University Press.

PAUL, C. S., CHRISTOPHER, D. E., REBECCA, L. P. L. P., SHAROLYN, A. & TILOTTAMA, G. 2009. Estimation of Mexico's Informal Economy and Remittances Using Nighttime Imagery. *Remote Sensing,* 1**,** 418-444.

PINKOVSKIY, M. & SALA-I-MARTIN, X. 2016a. Lights, Camera … Income! Illuminating the National Accounts-Household Surveys Debate. *The Quarterly Journal of Economics,* 131**,** 579-631.

PINKOVSKIY, M. & SALA-I-MARTIN, X. 2016b. Newer need not be better: evaluating the Penn World Tables and the World Development Indicators using nighttime lights. National Bureau of Economic Research.

RAM, R. & URAL, S. 2014. Comparison of GDP Per Capita Data in Penn World Table and World Development Indicators. *An International and Interdisciplinary Journal for Quality-of-Life Measurement,* 116**,** 639-646.

RAO, D. S. P., RAMBALDI, A. & DORAN, H. 2010. Extrapolation of Purchasing Power Parities Using Multiple Benchmarks and Auxiliary Information: A New Approach. *Rev. Income Wealth,* 56**,** S59-S98.

RAO, D. S. P., RAMBALDI, A. N., GANEGODAGE, K. R., HUYNH, L. T. & DORAN, H. E. 2016. UQICD v2 User Guide. The University of Queensland School of Economics. Discuss Paper No 534.

RATI, R. 2009. International income distribution: comparing new ICP and the existing data. *International Journal of Social Economics,* 36**,** 652-658.

SUMMERS, R. & HESTON, A. 1991. The Penn World Trade (Mark 5): An Expanded Set of International Comparisons, 1950-1988. *The Quarterly Journal of Economics,* 106**,** 327.

SUTTON, P. C. & COSTANZA, R. 2002. Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics,* 41**,** 509-527.

SUTTON, P. C., ELVIDGE, C. D. & GHOSH, T. 2007. Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics,* 8**,** 5-21.

THE WORLD BANK GROUP. *World Bank Country and Lending Groups* [Online]. The World Bank. Available: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [Accessed 28 September 2016].

THE WORLD BANK GROUP. 2016. *CO2 emissions (metric tons per capita)* [Online]. Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.: The World Bank Available: http://data.worldbank.org/indicator/EN.ATM.CO2E.PC [Accessed 26 - October 2016].

TILOTTAMA, G., SHAROLYN, J. A., CHRISTOPHER, D. E. & PAUL, C. S. 2013. Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being. *Sustainability,* 5**,** 4988-5019.

TIMMERMANN, A. 2006. Chapter 4 Forecast Combinations. *In:* G. ELLIOTT, C. W. J. G. & TIMMERMANN, A. (eds.) *Handbook of Economic Forecasting.* Elsevier.

WANG, W., CHENG, H. & ZHANG, L. 2012. Poverty assessment using DMSP/OLS night-time light satellite imagery at a provincial scale in China. *Advances in Space Research,* 49**,** 1253-1264.

WOOLDRIDGE, J. M. 2002. *Econometric analysis of cross section and panel data,* Cambridge, Mass: MIT Press.

WOOLDRIDGE, J. M. 2012. *Introductory Econometrics,* Mason, OH, Mason, OH : Cengage Learning US.

Results Appendix

Figure 5.1 Plot log light per capita on log GDP per capita (WDI) and log survey means

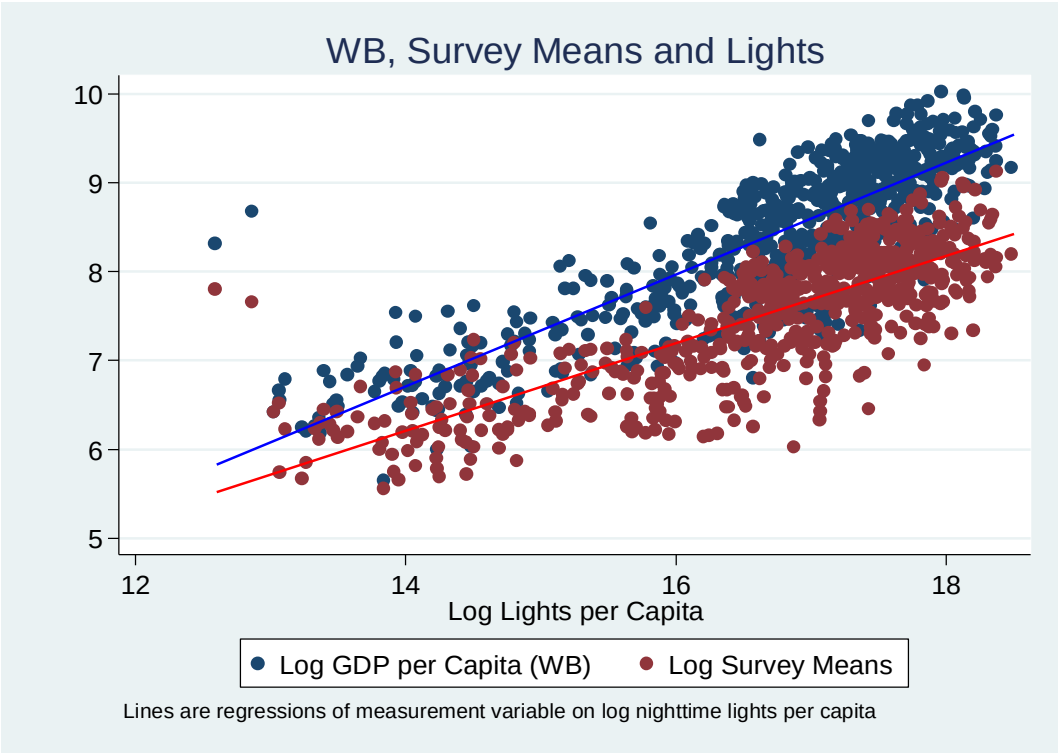


Figure 5.2 Plot log light per capita on log GDP per capita (UQICD) and log survey means

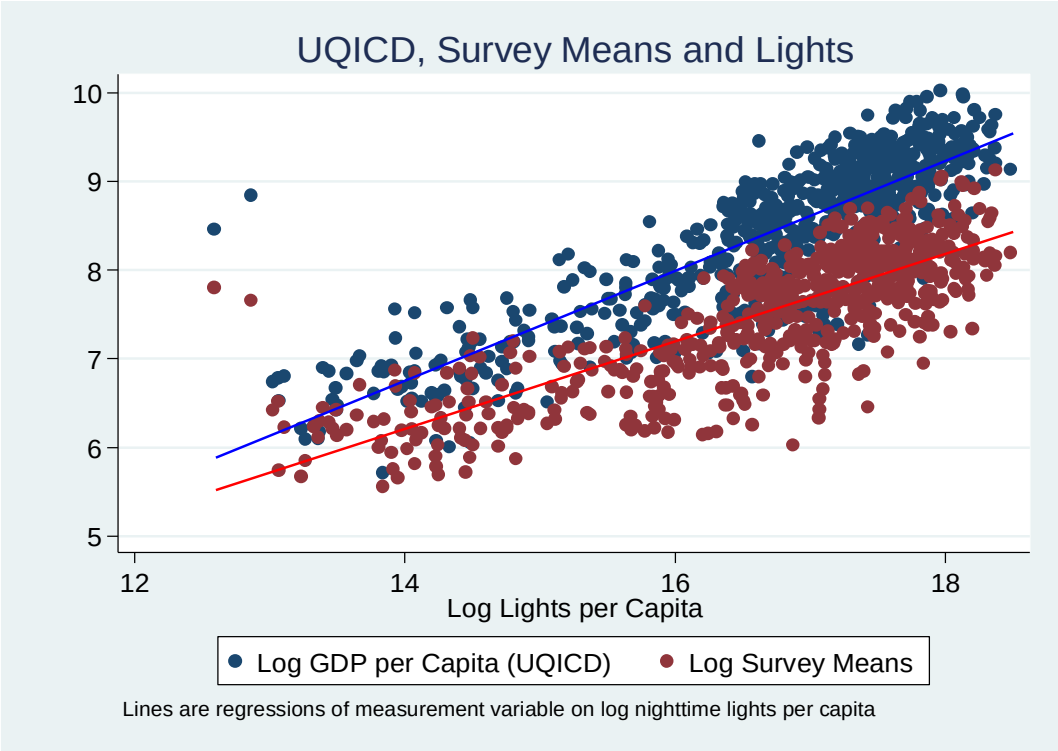


Figure 5.3 Plot log light per capita on log GDP per capita (PWT 7.1) and log survey means

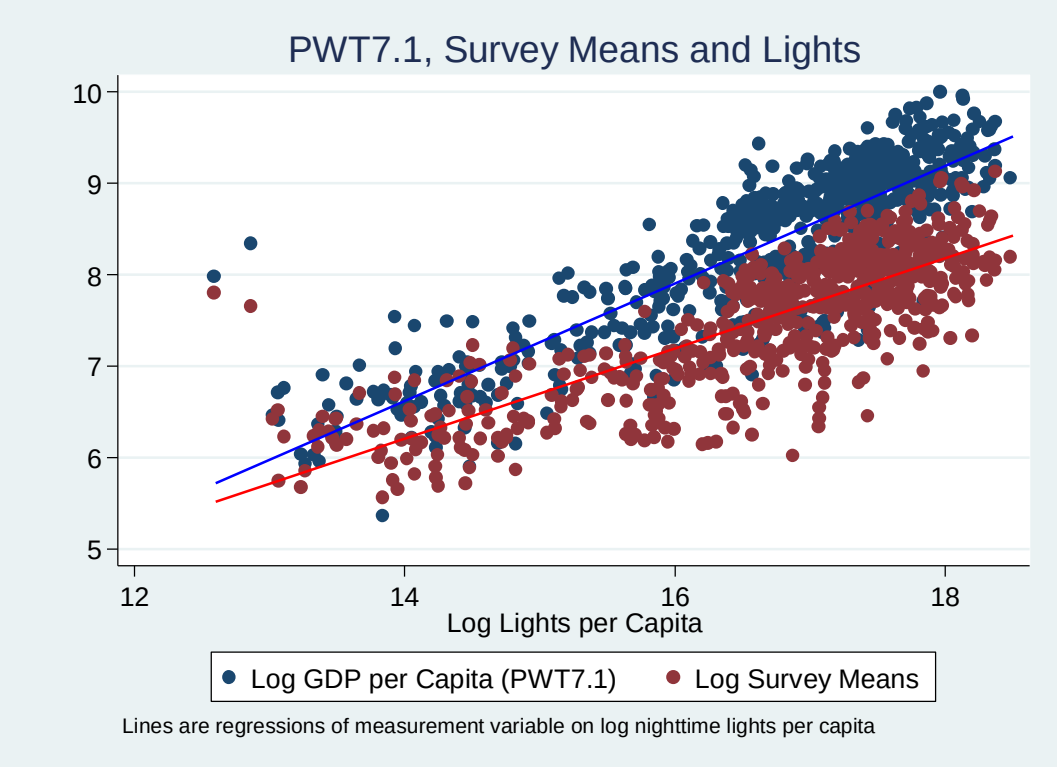


Figure 5.4 Plot log light per capita on log GDP per capita (PWT 8.1) and log survey means

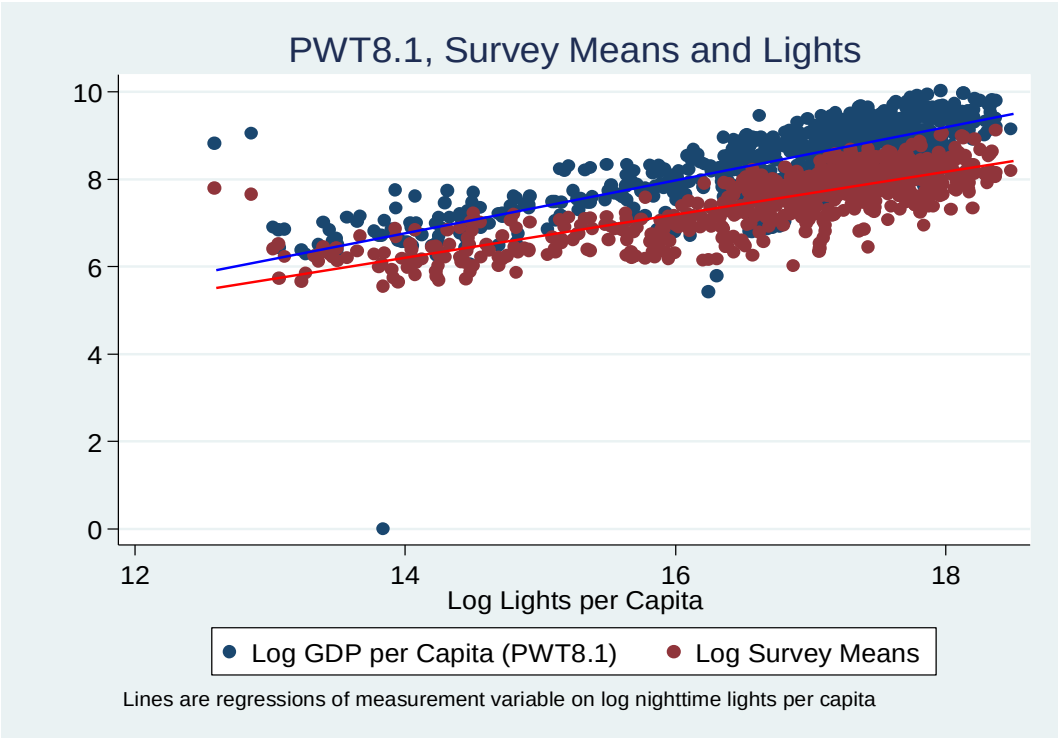


Figure 5.5 Plot log light per capita on log GDP per capita (PWT 9.0) and log survey means

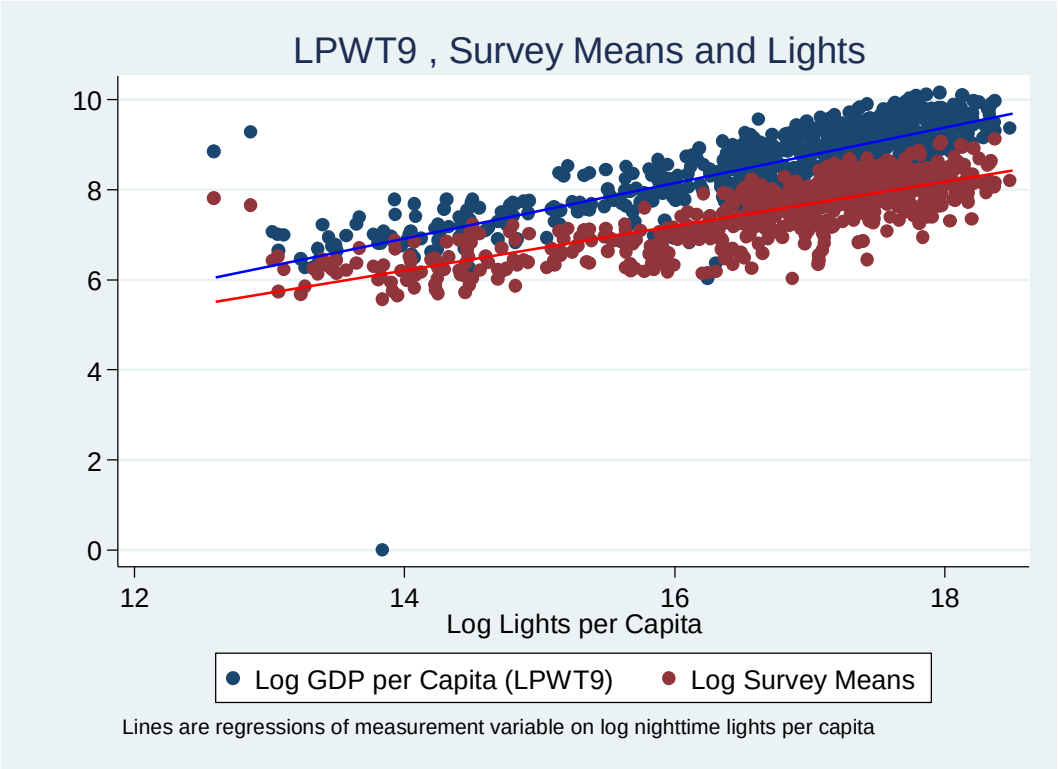


Table 5.3: Replicate Table II in Pinkovskiy and Sala-i-Martin (2016a, page 39) – WDI GDP data

Weights in the Optimal Proxy: Robustness Checks (WDI)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Baseline	Additional Covariates			Difrerent Dep.Var.			Different NAS Variables				
	Log Light per capita	Electricity	All Controls	Nonlinear Controls	Log Light Density	Calibrated Lights	Fraction Pop. Lit	Cons. WDI	Income Surveys	Cons. Surveys	Match Concepts	weighted
No Fixed Effects												
Log GDP_PC WDI	.83***	.89***	1.07	.99***	1.11***	.80***	1.06***	.83***	1.0***	.87***	.85***	1.12***
	(.6, 1.)	(.46, 1.3)	(-.2, 2.3)	(.3, 1.6)	(.56, 1.6)	(.44, 1.1)	(.59, 1.5)	(.59, 1.)	(.24, 1.7)	(.58, 1)	(.6, 1.1)	(.7, 1.5)
Log Survey	.16	.10	-.07	.00	-.11	.19	-.06	.16	-.001	.12	.14	-.12
	(-.06, .39)	(-.31, .5)	(-1.3, 1.2)	(-.68, .69)	(-.66, .4)	(-.17, .5)	(-.5, .4)	(-.06, .4)	(-.7, .7)	(-.15, .4)	(-.1, .39)	(-.5, .25)
Year Fixed Effects												
Log GDP_PC WDI	.80***	.85***	.99**	.92***	1.07***	.79***	1.03***	.80***	.92***	.82***	.81***	1.10***
	(.57, 1.)	(.42, 1.2)	(.04, 1.9)	(.43, 1.4)	(.5, 1.6)	(.42, 1.1)	(.55, 1.5)	(.57, 1.03)	(.2, 1.6)	(.55, 1)	(.56, 1)	(.7, 1.5)
Log Survey	.19*	.14	.008	.07	-.074	.20	-.03	.19*	.07	.17	.18	-.10
	(-.03, .4)	(-.28, .5)	(-.9, .9)	(-.4, .56)	(-.6, .47)	(-.1, .57)	(-.5, .44)	(-.03, .42)	(-.6, .7)	(-.1, .44)	(-.06, .43)	(-.5, .29)
Country Fixed Effects												
Log GDP_PC WDI	1.10***	1.14***	1.2***	1.16***	1.0***	1.10***	1.14***	1.0***	.92***	1.23***	1.07***	.97***
	(.78, 1.4)	(.48, 1.8)	(.69, 1.7)	(.67, 1.6)	(.73, 1.)	(.87, 1.3)	(.67, 1.6)	(.69, 1.3)	(.5, 1.2)	(.8, 1.6)	(.77, 1.3)	(.6, 1.3)
Log Survey	-.10	-.14	-.20	-.16	-.009	-.10	-.14	-.01	.07	-.23	-.07	.02
	(-.43, .2)	(-.8, .5)	(-.7, .3)	(-.64, .3)	(-.27, .26)	(-.33, .1)	(-.6, .3)	(-.34, .3)	(-.27, .4)	(-.6, .12)	(-.37, .2)	(-.3, .35)
Year and Country Fixed Effects												
Log GDP_PC WDI	1.06***	1.06***	1.10***	1.05***	1.09***	1.13***	.97*	1.02***	.66	1.15***	1.05***	.67***
	(.73, 1.3)	(.54, 1.5)	(.67, 1.5)	(.27, 1.8)	(.4, 1.7)	(.88, 1.3)	(-.0, 1.9)	(.5, 1.5)	(-.6, 1.9)	(.38, 1.9)	(.72, 1.3)	(.33, 1)
Log Survey	-.06	-.06	-.10	-.05	-.09	-.13	.027	-.02	.33	-.15	-.054	.32*
	(-.3, .26)	(-.57, .4)	(-.53, .3)	(-.84, .7)	(-.75, .56)	(-.39, .1)	(-.98, 1)	(-.5, .49)	(-.9, 1.6)	(-.9, .6)	(-.38, .27)	(-.01, .66)
No. of Obs	680	601	552	552	680	680	155	680	253	427	680	680
No. of Clusters	115	87	83	83	115	115	79	115	39	94	115	115

Note: Each column of Table II presents estimates of the relative weights of log GDP per capita and log survey means in the optimal lights-based proxy $z_{i;t}$ of the mean of the true income distribution. Block – bootstrapped 95% confidence intervals in parentheses. The baseline specification does not include covariate controls, and uses log aggregate lights per capita to measure light intensity. Column 2 controls for log electricity production per capita. Column 3 controls for log electricity production per capita, log total population , log % rural population , log % urban population , log area , latitude and longitude, the income share of the richest 10 % and the income share of the poorest 5 0% , log consumption share, log capital formation as % of GDP, log shares of GDP in agriculture, manufacturing and services, log export share, log import share, log government expenditure share of GDP, log GDP per energy unit consumed and log oil rents. Column 4 includes the controls in column 3 as well as their squares. Column 5 replaces the dependent variable with log light density. Columns 6 and 7 replace the dependent variable with log calibrated lights per capita, where the calibration is done to optimize fit to L IS data on Mexican state incomes (L IS 2013) and log fraction of the country is population that resides in lit areas, respectively. Column 8 replaces log GDP per capita with log national accounts consumption per capita. Column 9 considers only the sample of income survey s. Column 10 considers only the sample of consumption surveys, and replaces log GDP per capita with log national accounts consumption per capita. Column 11 rep laces log GDP per capita with log national accounts consumption per capita whenever the corresponding survey is a consumption survey. Column 12 weights all observations by average country population divided by the number of surveys for that country (Pinkovskiy and Sala-i-Martin 2016a, page 39).

Table 5.4: Replicate Table II in Pinkovskiy and Sala-i-Martin (2016a, page 39) – UQICD GDP data

Weights in the Optimal Proxy: Robustness Checks (UQICD)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Baseline	Additional Covariates			Difrerent Dep.Var.			Different NAS Variables				
	Log Light per capita	Electricity	All Controls	Nonlinear Controls	Log Light Density	Calibrated Lights	Fraction Pop. Lit	Cons. PWT7.1	Income Surveys	Cons. Surveys	Match Concepts	weighted
No Fixed Effects												
Log GDP_PC	.79***	.86***	1.09***	.98***	1.02***	.79***	1.04***	.68***	1.06***	.68***	.57***	1.09***
UCICD	(.58, 1)	(.46, 1)	(.2, 1.9)	(.3, 1.6)	(.4, 1.5)	(.47, 1)	(.6, 1.4)	(.4, .9)	(.4, 1.7)	(.3, .98)	(.27, .8)	(.7, 1.4)
Log Survey	.20*	.13	-.09	.014	-.02	.20	-.04	.31**	-.06	.31**	.42***	-.09
	(-.0, .4)	(-.27, .5)	(-.9, .7)	(-.6, .63)	(-.56, .5)	(-.11, .5)	(-.47, .3)	(.06, .5)	(-.7, .57)	(.01, .6)	(.12, .7)	(-.4, .28)
Year Fixed Effects												
Log GDP_PC	.77***	.84***	1.0**	.90***	.98***	.78***	1.0***	.68***	1.01***	.65***	.55***	1.03***
UCICD	(.5, .98)	(.43, 1)	(.2, 1.7)	(.4, 1.3)	(.4, 1.5)	(.45, 1)	(.5, 1.4)	(.4, .9)	(.4, 1.5)	(.37, .9)	(.2, .8)	(.6, 1.4)
Log Survey	.22**	.15	-.004	.09	.01	.21	-.00	.31**	-.01	.34**	.44***	-.03
	(.01, .4)	(-.2, .56)	(-.79, .7)	(-.3, .56)	(-.5, .5)	(-.1, .5)	(-.4, .4)	(.07, .5)	(-.5, .5)	(..05, .6)	(.1, .7)	(-.4, .3)
Country Fixed Effects												
Log GDP_PC	1.02***	1.14***	1.21	1.16***	.95***	1.0***	1.05***	.90***	.93***	1.16***	.91***	.77**
UCICD	(.68, 1.3)	(.49, 1.7)	(-1, 3.6)	(.7, 1.6)	(.69, 1.2)	(.7, 1.2)	(.5, 1.5)	(.4, 1.3)	(.6, 1.2)	(.4, 1.8)	(.4, 1.3)	(.16, 1.3)
Log Survey	-.02	-.14	-.21	-.16	.04	-.000	-.05	.09	.06	-.16	.08	.22
	(-.36, .3)	(-.79, .5)	(-2.6, 2.)	(-.6, .28)	(-.21, .3)	(-.29, .2)	(-.5, .4)	(-.39, .5)	(-.2, .3)	(-.87, .5)	(-.3, .5)	(-.37, .8)
Year and Country Fixed Effects												
Log GDP_PC	1.0***	1.06***	1.11***	1.06***	1.04**	1.08***	.85	.92*	.75***	1.23	.99***	.38
UCICD	(.5, 1.4)	(.5, 1.6)	(.7, 1.4)	(.5, 1.5)	(.1, 1.9)	(.6, 1.4)	(-.9, 2.6)	(-.17, 2)	(.43, 1)	(-16, 19)	(.2, 1.7)	(-4.3, 5.1)
Log Survey	-.00	-.06	-.11	-.06	-.04	-.08	.14	.07	.24	-.23	.00	.61
	(-.43, .4)	(-.6, .48)	(-.48, .25)	(-.58, .4)	(-.9, .8)	(-.48, .3)	(-1.6, 1.9)	(-1, 1.1)	(-.08, .5)	(-18, 17)	(-.7, .7)	(-4.1, 5.3)
No. of Obs	680	601	552	552	680	680	155	680	253	427	680	680
No. of Clusters	115	87	83	83	115	115	79	115	39	94	115	115

Note: Each column of Table II presents estimates of the relative weights of log GDP per capita and log survey means in the optimal lights-based proxy $z_{i;t}$ of the mean of the true income distribution. Block –bootstrapped 95% confidence intervals in parentheses. The baseline specification does not include covariate controls, and uses log aggregate lights per capita to measure light intensity. Column 2 controls for log electricity production per capita. Column 3 controls for log electricity production per capita, log total population , log % rural population , log % urban population , log area , latitude and longitude, the income share of the richest 10 % and the income share of the poorest 5 0% , log consumption share, log capital formation as % of GDP, log shares of GDP in agriculture, manufacturing and services, log export share, log import share, log government expenditure share of GDP, log GDP per energy unit consumed and log oil rents. Column 4 includes the controls in column 3 as well as their squares. Column 5 replaces the dependent variable with log light density. Columns 6 and 7 replace the dependent variable with log calibrated lights per capita, where the calibration is done to optimize fit to L IS data on Mexican state incomes (L IS 2013) and log fraction of the country is population that resides in lit areas, respectively. Column 8 replaces log GDP per capita with log national accounts consumption per capita. Column 9 considers only the sample of income survey s. Column 10 considers only the sample of consumption surveys, and replaces log GDP per capita with log national accounts consumption per capita. Column 11 rep laces log GDP per capita with log national accounts consumption per capita whenever the corresponding survey is a consumption survey. Column 12 weights all observations by average country population divided by the number of surveys for that country (Pinkovskiy and Sala-i-Martin 2016a, page 39).

Table 5.5: Replicate Table II in Pinkovskiy and Sala-i-Martin (2016a, page 39) – PWT 7.1 GDP data

Weights in the Optimal Proxy: Robustness Checks (PWT 7.1)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Baseline	Additional Covariates			Difrerent Dep.Var.			Different NAS Variables				
	Log Light per capita	Electricity	All Controls	Nonlinear Controls	Log Light Density	Calibrated Lights	Fraction Pop. Lit	Cons. PWT7.1	Income Surveys	Cons. Surveys	Match Concepts	weighted
No Fixed Effects												
Log GDP_PC	.86***	.81***	1.01	.98*	1.30***	.75***	1.13***	.89***	.87***	.96***	.86***	.92***
PWT 7.1	(.62, 1.)	(.38, 1.2)	(-3.1, 5.2)	(-.00, 1.9)	(.7, 1.8)	(.41, 1.)	(.67, 1.5)	(.6, 1.1)	(.2, 1.5)	(.6, 1.2)	(.5, 1.1)	(.58, 1.2)
Log Survey	.13	.18	-.019	.01	-.30	.24	-.13	.10	.12	.03	.13	.07
	(-.09, .37)	(-.2, .6)	(-4.2, 4.1)	(-.96, 1.0)	(-.8, .28)	(-.09, .58)	(-.59, .3)	(-.13, .3)	(-.5, .77)	(-.29, .3)	(-.16, .4)	(-.26, .4)
Year Fixed Effects												
Log GDP_PC	.82***	.78***	.93	.90***	1.25***	.75***	1.11***	.86***	.79***	.92***	.82***	.92***
PWT 7.1	(.59, 1.)	(.3, 1.2)	(-.9, 2.8)	(.36, 1.4)	(.6, 1.8)	(.40, 1.)	(.65, 1.5)	(.63, 1.1)	(.2, 1.3)	(.6, 1.2)	(.5, 1.1)	(.57, 1.2)
Log Survey	.17	.21	.061	.095	-.25	.24	-.11	.13	.20	.07	.17	.07
	(-.05, .4)	(-.21, .6)	(-1.8, 1.9)	(-.44, .63)	(-.8, .33)	(-.09, .59)	(-.57, .3)	(-.1, .3)	(-.38, .79)	(-.2, .39)	(-.12, .4)	(-.26, .4)
Country Fixed Effects												
Log GDP_PC	1.05***	1.09***	1.20***	1.17***	.95***	1.04***	.98***	.98***	.86***	1.18***	.95***	.88***
PWT 7.1	(.69, 1.4)	(.3, 1.8)	(.7, 1.6)	(.73, 1.6)	(.66, 1.2)	(.8, 1.2)	(.46, 1.5)	(.6, 1.4)	(.47, 1.2)	(.6, 1.7)	(.5, 1.3)	(.43, 1.3)
Log Survey	-.05	-.09	-.20	-.17	.04	-.04	.014	-.01	.13	-.18	.04	.11
	(-.40, .3)	(-.86, .6)	(-.69, .27)	(-.60, .26)	(-.25, .3)	(-.28, .19)	(-.5, .53)	(-.4, .3)	(-.24, .52)	(-.74, .38)	(-.34, .43)	(-.33, .56)
Year and Country Fixed Effects												
Log GDP_PC	1.02***	1.02***	1.09***	1.04***	1.04*	1.10***	.76	.99***	.49***	1.18***	.97***	.56***
PWT 7.1	(.64, 1.3)	(.4, 1.6)	(.5, 1.6)	(.55, 1.5)	(-.01, 2)	(.79, 1.4)	(-31, 32)	(.37, 1.6)	(-1.7, 2.7)	(-13, 15)	(.2, 1.6)	(.13, .99)
Log Survey	-.021	-.02	-.09	-.041	-.04	-.10	.23	.00	.50***	-.18***	-.02	.43**
	(-.39, .3)	(-.62, .5)	(-.6, .4)	(-.53, .44)	(-1.1, 1.)	(-.4, .2)	(-31, 32)	(-.61, .6)	(-1.7, 2.7)	(-14, 14)	(-.6, .71)	(.00, .86)
No. of Obs	680	601	552	552	680	680	155	680	253	427	680	680
No. of Clusters	115	87	83	83	115	115	79	115	39	94	115	115

Note: Each column of Table II presents estimates of the relative weights of log GDP per capita and log survey means in the optimal lights-based proxy $z_{i;t}$ of the mean of the true income distribution. Block – bootstrapped 95% confidence intervals in parentheses. The baseline specification does not include covariate controls, and uses log aggregate lights per capita to measure light intensity. Column 2 controls for log electricity production per capita. Column 3 controls for log electricity production per capita, log total population , log % rural population , log % urban population , log area , latitude and longitude, the income share of the richest 10 % and the income share of the poorest 5 0% , log consumption share, log capital formation as % of GDP, log shares of GDP in agriculture, manufacturing and services, log export share, log import share, log government expenditure share of GDP, log GDP per energy unit consumed and log oil rents. Column 4 includes the controls in column 3 as well as their squares. Column 5 replaces the dependent variable with log light density. Columns 6 and 7 replace the dependent variable with log calibrated lights per capita, where the calibration is done to optimize fit to L IS data on Mexican state incomes (L IS 2013) and log fraction of the country is population that resides in lit areas, respectively. Column 8 replaces log GDP per capita with log national accounts consumption per capita. Column 9 considers only the sample of income survey s. Column 10 considers only the sample of consumption surveys, and replaces log GDP per capita with log national accounts consumption per capita. Column 11 rep laces log GDP per capita with log national accounts consumption per capita whenever the corresponding survey is a consumption survey. Column 12 weights all observations by average country population divided by the number of surveys for that country (Pinkovskiy and Sala-i-Martin 2016a, page 39).

Table 5.6: Replicate Table II in Pinkovskiy and Sala-i-Martin (2016a, page 39) – PWT 8.1 GDP data

Weights in the Optimal Proxy: Robustness Checks (PWT 8.1)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Baseline	Additional Covariates			Difrerent Dep.Var.			Different NAS Variables				
	Log Light per capita	Electricity	All Controls	Nonlinear Controls	Log Light Density	Calibrated Lights	Fraction Pop. Lit	Cons. PWT7.1	Income Surveys	Cons. Surveys	Match Concepts	weighted
No Fixed Effects												
Log GDP_PC PWT 8.1	.49***	.36	.69	.73	.52	.37***	.68***	.63***	.73	.91***	.59***	.45*
	(.2, .74)	(-.07, .8)	(-.5, 1.9)	(-11, 13)	(-.19, 1.2)	(.10, .65)	(.20, 1.1)	(.33, .93)	(-10, 12)	(.6, 1.1)	(.25, .9)	(-.0, .91)
Log Survey	.50***	.63***	.30	.26**	.47	.62***	.31	.36**	.26	.08	.40**	.54**
	(.2, .76)	(.19, 1.0)	(-.9, 1.5)	(-12, 12)	(-.23, 1.1)	(.34, .89)	(-.17, .7)	(.06, .66)	(-11, 11)	(-.1, .3)	(.06, .74)	(.08, 1.0)
Year Fixed Effects												
Log GDP_PC PWT 8.1	.47***	.34	.56	.62	.50	.36***	.65***	.62***	.65	.88***	.58***	.40
	(.23, .7)	(-.08, .7)	(-18, 19)	(-.78, 2)	(-.2, 1.2)	(.08, .64)	(.17, 1)	(.33, .92)	(-1.9, 3)	(.6, 1.1)	(.25, .91)	(-.08, .8)
Log Survey	.52***	.65***	.43	.37	.49	.63***	.34	.37**	.34	.11	.41**	.59**
	(.28, .7)	(.22, 1.0)	(-18, 19)	(-1, 1.7)	(-.2, 1.2)	(.35, .91)	(-.13, .8)	(.07, .66)	(-2, 2)	(-.1, .3)	(.08, .74)	(.1, 1.0)
Country Fixed Effects												
Log GDP_PC PWT 8.1	.91***	.99**	1.31**	1.25***	.86***	.80***	.68*	1.06***	.72***	1.3***	1.07***	.27
	(.4, 1.3)	(.10, 1.8)	(.2, 2.3)	(.6, 1.8)	(.51, 1.2)	(.42 ,1.1)	(-.0, 1.3)	(.66, 1.4)	(.2, 1.1)	(.6, 1.9)	(.69, 1.4)	(-.4, .95)
Log Survey	.08	.00	-.31	-.25	.13	.19	.31	-.06	.27	-.31	-.07	.72**
	(-.35, .5)	(-.8, .89)	(-1.3, .7)	(-.85, .3)	(-.2, .48)	(-.19, .5)	(-.38, 1)	(-.46, .3)	(-.19, .74)	(-.9, .3)	(-.45, .3)	(.04, 1.4
Year and Country Fixed Effects												
Log GDP_PC PWT 8.1	.88***	.85	1.1***	1.05***	.95	.96***	12	1.03***	.10	1.2**	1.03***	-.08
	(.2, 1.4)	(-3.4, 5)	(.5, 1.7)	(.57, 1.5)	(-.8, 2.7)	(.3, 1.6)	(-6, 31)	(.45, 1.6)	(-3.8, 4.0)	(.2, 2.1)	(.51, 1.5)	(-13, 13)
Log Survey	.11	.14	-.13	-.05	.04	.03	-11	-.03	.89	-.21	-.03	1.08
	(-.49, .7)	(-4, 4.4)	(-.7, .49)	(-.54, .4)	(-1.7, 1.8)	(-.6, .69)	(-30, 7)	(-.6, .54)	(-3.0, 4.8)	(-1, .7)	(-.5, .48)	(-12, 14)
No. of Obs	680	601	552	552	680	680	155	680	253	426	680	680
No. of Clusters	115	87	83	83	115	115	79	115	39	93	115	115

Note: Each column of Table II presents estimates of the relative weights of log GDP per capita and log survey means in the optimal lights-based proxy $z_{i;t}$ of the mean of the true income distribution. Block – bootstrapped 95% confidence intervals in parentheses. The baseline specification does not include covariate controls, and uses log aggregate lights per capita to measure light intensity. Column 2 controls for log electricity production per capita. Column 3 controls for log electricity production per capita, log total population , log % rural population , log % urban population , log area , latitude and longitude, the income share of the richest 10 % and the income share of the poorest 5 0% , log consumption share, log capital formation as % of GDP, log shares of GDP in agriculture, manufacturing and services, log export share, log import share, log government expenditure share of GDP, log GDP per energy unit consumed and log oil rents. Column 4 includes the controls in column 3 as well as their squares. Column 5 replaces the dependent variable with log light density. Columns 6 and 7 replace the dependent variable with log calibrated lights per capita, where the calibration is done to optimize fit to L IS data on Mexican state incomes (L IS 2013) and log fraction of the country is population that resides in lit areas, respectively. Column 8 replaces log GDP per capita with log national accounts consumption per capita. Column 9 considers only the sample of income survey s. Column 10 considers only the sample of consumption surveys, and replaces log GDP per capita with log national accounts consumption per capita. Column 11 rep laces log GDP per capita with log national accounts consumption per capita whenever the corresponding survey is a consumption survey. Column 12 weights all observations by average country population divided by the number of surveys for that country (Pinkovskiy and Sala-i-Martin 2016a, page 39).

Table 5.7: Replicate Table II in Pinkovskiy and Sala-i-Martin (2016a, page 39) – PWT 9.0 GDP data

Weights in the Optimal Proxy: Robustness Checks (PWT 9.0)												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Baseline	Additional Covariates			Difrerent Dep.Var.			Different NAS Variables				
	Log Light per capita	Electricity	All Controls	Nonlinear Controls	Log Light Density	Calibrated Lights	Fraction Pop. Lit	Cons. PWT7.1	Income Surveys	Cons. Surveys	Match Concepts	weighted
No Fixed Effects												
Log GDP_PC PWT 9	.47***	.42**	.76	.62	.67***	.34**	.83***	.61***	1.1	.55***	.57***	.22
	(.22, .71)	(.07, .77)	(-1, 2.8)	(-.79, 2.0)	(.17, 1.1)	(.05, .64)	(.41, 1.2)	(.26, .96)	(-1.3, 3.5)	(.1, .97)	(.25, .88)	(-.22, .68)
Log Survey	.52***	.57***	.23	.37	.32	.65***	.16	.38**	-.10	.44**	.42***	.77***
	(.28, .77)	(.22, .92)	(-1.8, 2)	(-1.0, 1.7)	(-.17, .82)	(.35, .94)	(-.25, .58)	(.03, .73)	(-2.5, 2.3)	(.0, .87)	(.11, .74)	(.31, 1.2)
Year Fixed Effects												
Log GDP_PC PWT 9	.45***	.40**	.65	.54	.65***	.34**	.81***	.59***	1.0*	.52**	.56***	.16
	(.21, .70)	(.05, .75)	(-4, 5.5)	(-.30, 1.3)	(.17, 1.1)	(.06, .62)	(.40, 1.2)	(.24, .94)	(-.07, 2.1)	(.1, .94)	(.24, .87)	(-.34, .67)
Log Survey	.54***	.59***	.34	.45	.34	.65***	.18	.40**	-.02	.47**	.43***	.83***
	(.29, .78)	(.24, .94)	(-4.5, 5)	(-.39, 1.3)	(-.13, .82)	(.37, .93)	(-.23, .59)	(.05, .75)	(-1.1, 1)	(.0, .89)	(.12,.75)	(.32, 1.3)
Country Fixed Effects												
Log GDP_PC PWT 9	.90***	.98**	1.25	1.1	.86***	.76***	.61	.94***	.63***	1.04**	.90***	.32
	(.47, 1.3)	(.07, 1.8)	(-1, 3.5)	(-1.0, 3.4)	(.54, 1.1)	(.39, 1.1)	(-.32, 1.5)	(.4, 1.4)	(.2, 1.07)	(.2, 1.8)	(.45, 1.3)	(-.38, 1.0)
Log Survey	.09	.01	-.25	-.18	.13	.23	.38	.05	.36	-.04	.09	.67*
	(-.33, .52)	(-.89, .92)	(-2.5, 2)	(-2.4, 2.0)	(-.18, .45)	(-.13, .60)	(-.55, 1.3)	(-.4, .54)	(-.07, .79)	(-.8,.77)	(-.34, .54)	(-.03, 1.3)
Year and Country Fixed Effects												
Log GDP_PC PWT 9	.85***	.81	1.0***	.91*	.93	.89**	10	.98***	.39	1.09	.93***	-.07
	(.32, 1.3)	(-1.2, 2.9)	(.3, 1.6)	(-.14, 1.9)	(-6.6, 8)	(.19, 1.5)	(-6.5, 26)	(.37, 1.5)	(-.41, 1.2)	(-.5, 2.7)	(.45, 1.4)	(-2.6, 2.4)
Log Survey	.14	.18	-.01	.08	.06	.10	-9	.01	.60	-.09	.06	1.07
	(-.37, .67)	(-1.9, 2.2)	(-.69, .6)	(-.97, 1.1)	(-7.5, 7.6)	(-.58, .80)	(-25, 7.5)	(-.59, .6)	(-.20, 1.4)	(-1, 1.5)	(-.40, .54)	(-1.4, 3.6)
No. of Obs	680	601	552	552	680	680	155	680	253	426	680	680
No. of Clusters	115	87	83	83	115	115	79	115	39	93	115	115

Note: Each column of Table II presents estimates of the relative weights of log GDP per capita and log survey means in the optimal lights-based proxy $z_{i;t}$ of the mean of the true income distribution. Block – bootstrapped 95% confidence intervals in parentheses. The baseline specification does not include covariate controls, and uses log aggregate lights per capita to measure light intensity. Column 2 controls for log electricity production per capita. Column 3 controls for log electricity production per capita, log total population , log % rural population , log % urban population , log area , latitude and longitude, the income share of the richest 10 % and the income share of the poorest 5 0% , log consumption share, log capital formation as % of GDP, log shares of GDP in agriculture, manufacturing and services, log export share, log import share, log government expenditure share of GDP, log GDP per energy unit consumed and log oil rents. Column 4 includes the controls in column 3 as well as their squares. Column 5 replaces the dependent variable with log light density. Columns 6 and 7 replace the dependent variable with log calibrated lights per capita, where the calibration is done to optimize fit to L IS data on Mexican state incomes (L IS 2013) and log fraction of the country is population that resides in lit areas, respectively. Column 8 replaces log GDP per capita with log national accounts consumption per capita. Column 9 considers only the sample of income survey s. Column 10 considers only the sample of consumption surveys, and replaces log GDP per capita with log national accounts consumption per capita. Column 11 rep laces log GDP per capita with log national accounts consumption per capita whenever the corresponding survey is a consumption survey. Column 12 weights all observations by average country population divided by the number of surveys for that country (Pinkovskiy and Sala-i-Martin 2016a, page 39).

Table 5.8: Combination forecast approach (WDI data)

Weights in the Optimal Proxy : Combination forecast (WDI)								
	P&S method Dependent variables				Combination forecast Dependent variables			
	(1) Baseline	(2) Log Light Density	(3) Calibrated Lights	(4) Fraction Pop. Lit	(5) Baseline	(6) Log Light Density	(7) Calibrated Lights	(8) Fraction Pop. Lit
No Fixed Effects								
Log GDP_PC WDI	.83***	1.11***	.80***	1.06***	1.08***	1.28***	1.2***	.72***
	(.6, 1.)	(.56, 1.6)	(.44, 1.1)	(.59, 1.5)	(.8, 1.3)	(.81, 1.7)	(.82, 1.7)	(.42, .1.0)
Log Survey	.16	-.11	.19	-.06	-.08	-.28	-.27	.27*
	(-.06, .39)	(-.66, .4)	(-.17, .5)	(-.5, .4)	(-.37, .2)	(-.75, .19)	(-.72, .18)	(-.03, .58)
Year Fixed Effects								
Log GDP_PC WDI	.80***	1.07***	.79***	1.03***	1.06***	1.23***	1.24***	.69***
	(.57, 1.)	(.5, 1.6)	(.42, 1.1)	(.55, 1.5)	(.77, 1.3)	(.76, 1.7)	(.79,1.7)	(.4, 1)
Log Survey	.19*	-.074	.20	-.03	-.06	-.23	-.24	.30*
	(-.03, .4)	(-.6, .47)	(-.16, .57)	(-.5, .44)	(-.35, .23)	(-.71, .24)	(-.70, .21)	(00, .60)
Country Fixed Effects								
Log GDP_PC WDI	1.10***	1.0***	1.10***	1.14***	.87***	.89***	1.37***	.82***
	(.78, 1.4)	(.73, 1.)	(.87, 1.3)	(.67, 1.6)	(.65, 1.0)	(.67, 1.1)	(.86, 1.8)	(.48, 1.1)
Log Survey	-.10	-.009	-.10	-.14	.12	.10	-.37	.17
	(-.43, .2)	(-.27, .26)	(-.33, .1)	(-.6, .3)	(-.09, .35)	(-.11, .33)	(-.89, .14)	(-.16. .52)
Country and Year Fixed Effects								
Log GDP_PC WDI	1.06***	1.09***	1.13***	.97*	.94***	.88***	1.29***	.79***
	(.73, 1.3)	(.4, 1.7)	(.88, 1.3)	(-.04, 1.9)	(.72, 1.1)	(.64, 1.1)	(.84, 1.7)	(.5, 1.1)
Log Survey	-.06	-.09	-.13	.027	.05	.11	-.29	.20
	(-.3, .26)	(-.75, .56)	(-.39, .1)	(-.98, 1)	(-.17, .28)	(-.13, .36)	(-.75, .16)	(-.10, .50)
No. of Obs	680	680	680	155	680	680	680	155
No. of Clusters	115	115	115	79	115	115	115	79

Table 5.8 shows estimates of the relative weight of GDP per capita (WDI) and survey means in measuring true income per capita using P&S's method and combination method. Columns 1-4 are taken from columns 1, 5, 6 and 7 of Table 5.3. Columns 5-8 present the results of combination forecast approach using the same data as column 1-4. Note that the numbers inside parentheses of columns 1-4 are block-bootstrapped 95% confidence intervals, while the numbers inside parentheses of columns 5-8 are true confidence interval. Columns 1 and 5 are baseline which means that these columns use aggregate light per capita. Column 2-4 and 6-8 use three different types of light data which are described in detail in Pinkovskiy and Sala-i-Martin (2016a, page 21).

Table 5.9: Combination forecast approach (UQICD data)

Weights in the Optimal Proxy : Combination forecast (UQICD)								
	P&S method Dependent variables				Combination forecast Dependent variables (no restrict)			
	(1) Baseline	(2) Log Light Density	(3) Calibrated Lights	(4) Fraction Pop. Lit	(5) Baseline	(6) Log Light Density	(7) Calibrated Lights	(8) Fraction Pop. Lit
No Fixed Effects								
Log GDP_PC UQICD	.79***	1.02***	.79***	1.04***	1.04***	1.20***	1.24***	.73***
	(.58, 1)	(.4, 1.5)	(.47, 1)	(.6, 1.4)	(.76, 1.33)	(.71, 1.6)	(.83, 1.6)	(.45, 1.0)
Log Survey	.20*	-.02	.20	-.04	-.04	-.20	-.24	.26*
	(-.0, .4)	(-.56, .5)	(-.11, .5)	(-.47, .3)	(-.33, .24)	(-.69, .29)	(-.66, .17)	(-.02, .55)
Year Fixed Effects								
Log GDP_PC UQICD	.77***	.98***	.78***	1.0***	1.02***	1.15***	1.22***	.70***
	(.5, .98)	(.4, 1.5)	(.45, 1)	(.5, 1.4)	(.74, 1.3)	(.66, 1.6)	(.8, 1.6)	(.42, 1)
Log Survey	.22**	.01	.21	-.00	-.02	-.15	-.22	.29*
	(.01, .4)	(-.5, .5)	(-.1, .5)	(-.4, .4)	(-.32, .26)	(-.65, .34)	(-.65, .20)	(.00, .58)
Country Fixed Effects								
Log GDP_PC UQICD	1.02***	.95***	1.0***	1.05***	.77***	.81***	1.19***	.71***
	(.68, 1.3)	(.69, 1.2)	(.7, 1.2)	(.5, 1.5)	(.53, 1.0)	(.59, 1.0)	(.63, 1.7)	(.33, 1.1)
Log Survey	-.02	.04	-.000	-.05	.22	.18	-.19	.28
	(-.36, .3)	(-.21, .3)	(-.29, .2)	(-.5, .4)	(-.01, .47)	(-.04, .41)	(-.76, .37)	(-.10, .67)
Country and Year Fixed Effects								
Log GDP_PC UQICD	1.0***	1.04**	1.08***	.85	.84***	.80***	1.12***	.69***
	(.5, 1.4)	(.1, 1.9)	(.6, 1.4)	(-.9, 2.6)	(.6, 1.1)	(.55, 1.0)	(.61, 1.6)	(.35, 1.0)
Log Survey	-.00	-.04	-.08	.14	.15	.19	-.12	.30*
	(-.43, .4)	(-.9, .8)	(-.48, .3)	(-1.6, 1.9)	(-.10, .40)	(-.06, .45)	(-.63, .39)	(-.04, .65)
No. of Obs	680	680	680	155	680	680	680	155
No. of Clusters	115	115	115	79	115	115	115	79

Table 5.9 shows estimates of the relative weight of GDP per capita (UQICD) and survey means in measuring true income per capita using P&S's method and combination method. Columns 1-4 are taken from columns 1, 5, 6 and 7 of Table 5.4. Columns 5-8 present the results of combination forecast approach using the same data as columns 1-4. Note that the numbers inside parentheses of column 1-4 are block-bootstrapped 95% confidence intervals, while the numbers inside parentheses of columns 5-8 are true confidence interval. Columns 1 and 5 are baseline which means that these columns use aggregate light per capita. Columns 2-4 and 6-8 use three different types of light data which are described in detail in Pinkovskiy and Sala-i-Martin (2016a, page 21).

Table 5.10: Combination forecast approach (PWT 7.1 data)

Weights in the Optimal Proxy : Combination forecast (PWT 7.1)								
	P&S method Dependent variables				Combination forecast Dependent variables (no restrict)			
	(1) Baseline	(2) Log Light Density	(3) Calibrated Lights	(4) Fraction Pop. Lit	(5) Baseline	(6) Log Light Density	(7) Calibrated Lights	(8) Fraction Pop. Lit
No Fixed Effects								
Log GDP_PC PWT 7.1	.86***	1.30***	.75***	1.13***	1.1***	1.4***	1.2***	.74***
	(.62, 1.)	(.7, 1.8)	(.41, 1.)	(.67, 1.5)	(.83, 1.3)	(.97, 1.9)	(.77, 1.6)	(.45, 1.0)
Log Survey	.13	-.30	.24	-.13	-.1	-.4*	-.2	.25*
	(-.09, .37)	(-.8, .28)	(-.09, .58)	(-.59, .3)	(-.38, .17)	(-.92, .03)	(-.64, .23)	(-.03, .55)
Year Fixed Effects								
Log GDP_PC PWT 7.1	.82***	1.25***	.75***	1.11***	1.08***	1.4***	1.18***	.71***
	(.59, 1.)	(.6, 1.8)	(.40, 1.)	(.65, 1.5)	(.81, 1.3)	(.93, 1.8)	(.74, 1.6)	(.43, 1.0)
Log Survey	.17	-.25	.24	-.11	-.08	-.4*	-.18	.28*
	(-.05, .4)	(-.8, .33)	(-.09, .59)	(-.57, .3)	(-.35, .19)	(-.88, .07)	(-.64, .26)	(-.01, .57)
Country Fixed Effects								
Log GDP_PC PWT 7.1	1.05***	.95***	1.04***	.98***	.83***	.84***	1.26***	0.73***
	(.69, 1.4)	(.66, 1.2)	(.8, 1.2)	(.46, 1.5)	(.6, 1.0)	(.61, 1.0)	(.76, 1.7)	(.39, 1.0)
Log Survey	-.05	.04	-.04	.014	.16	.15	-.26	.26
	(-.40, .3)	(-.25, .3)	(-.28, .19)	(-.5, .53)	(-.06, .40)	(-.09, .39)	(-.78, .24)	(-.09, .61)
Country and Year Fixed Effects								
Log GDP_PC PWT 7.1	1.02***	1.04*	1.10***	.76	.88***	.83***	1.19***	.69***
	(.64, 1.3)	(-.01, 2)	(.79, 1.4)	(-31, 32)	(.65, 1.1)	(.58, 1.0)	(.75, 1.6)	(.34, 1.0)
Log Survey	-.021	-.04	-.10	.23	.11	.16	-.19	.30*
	(-.39, .3)	(-1.1, 1.)	(-.4, .2)	(-31, 32)	(-.13, .35)	(-.09, .42)	(-.64, .25)	(-.05, .66)
No. of Obs	680	680	680	155	680	680	680	155
No. of Clusters	115	115	115	79	115	115	115	79

Table 5.10 shows estimates of the relative weight of GDP per capita (PWT 7.1) and survey means in measuring true income per capita using P&S's method and combination method. Columns 1-4 are taken from column 1, 5, 6 and 7 of Table 5.5. Columns 5-8 present the results of combination forecast approach using the same data as column 1-4. Note that the numbers inside parentheses of column 1-4 are block-bootstrapped 95% confidence intervals, while the numbers inside parentheses of column 5-8 are true confidence interval. Column 1 and 5 are baseline which means that these columns use aggregate light per capita. Column 2-4 and 6-8 use three different types of light data which are described in detail in (Pinkovskiy and Sala-i-Martin 2016a, page 21).

Table 5.11: Combination forecast approach (PWT 8.1 data)

Weights in the Optimal Proxy : Combination forecast (PWT 8.1)								
	P&S method Dependent variables				Combination forecast Dependent variables (no restrict)			
	(1) Baseline	(2) Log Light Density	(3) Calibrated Lights	(4) Fraction Pop. Lit	(5) Baseline	(6) Log Light Density	(7) Calibrated Lights	(8) Fraction Pop. Lit
No Fixed Effects								
Log GDP_PC PWT 8.1	.49***	.52	.37***	.68***	0.64***	.64	.58***	.54***
	(.2, .74)	(-.19, 1.2)	(.10, .65)	(.20, 1.1)	(.32, .97)	(-.2, 1.51)	(.17, .99)	(.18, .9)
Log Survey	.50***	.47	.62***	.31	.35**	.36	.42**	.46**
	(.2, .76)	(-.23, 1.1)	(.34, .89)	(-.17, .7)	(.03, .68)	(-.51, 1.2)	(.01, .83)	(.10, .82)
Year Fixed Effects								
Log GDP_PC PWT 8.1	.47***	.50	.36***	.65***	.62***	.61	.56***	.52***
	(.23, .7)	(-.2, 1.2)	(.08, .64)	(.17, 1)	(.31, .95)	(-.2, 1.47)	(.15, .96)	(.16, .87)
Log Survey	.52***	.49	.63***	.34	.37**	.39	.44**	.48***
	(.28, .7)	(-.2, 1.2)	(.35, .91)	(-.13, .8)	(.05, .69)	(-.47, 1.2)	(.04, .85)	(.13, .84)
Country Fixed Effects								
Log GDP_PC PWT 8.1	.91***	.86***	.80***	.68*	.64***	.7***	.91***	.43*
	(.4, 1.3)	(.51, 1.2)	(.42 ,1.1)	(-.0, 1.3)	(.3, .99)	(.39, 1.01)	(.27, 1.53)	(-.08, .94)
Log Survey	.08	.13	.19	.31	.35**	.30*	.09	.57***
	(-.35, .5)	(-.2, .48)	(-.19, .5)	(-.38, 1)	(.01, .70)	(-.01, .61)	(-.53, .73)	(.06, 1.08)
Country and Year Fixed Effects								
Log GDP_PC PWT 8.1	.88***	.95	.96***	12	.67***	.65***	.84***	.42
	(.2, 1.4)	(-.8, 2.7)	(.3, 1.6)	(-6, 31)	(.3, 1.03)	(.31, .99)	(.24, 1.43)	(-.08, .93)
Log Survey	.11	.04	.03	-11	.33*	.35**	.16	.58**
	(-.49, .7)	(-1.7, 1.8)	(-.6, .69)	(-30, 7)	(-.03, .70)	(.007, .69)	(-.43, .76)	(.07, 1.08)
No. of Obs	680	680	680	155	680	680	680	155
No. of Clusters	115	115	115	79	115	115	115	79

Table 5.11 shows estimates of the relative weight of GDP per capita (PWT 8.1) and survey means in measuring true income per capita using P&S's method and combination method. Columns 1-4 are taken from columns 1, 5, 6 and 7 of Table 5.6. Columns 5-8 present the results of combination forecast approach using the same data as columns 1-4. Note that the numbers inside parentheses of columns 1-4 are block-bootstrapped 95% confidence intervals, while the numbers inside parentheses of columns 5-8 are true confidence interval. Columns 1 and 5 are baseline which means that these columns use aggregate light per capita. Columns 2-4 and 6-8 use three different types of light data which are described in detail in (Pinkovskiy and Sala-i-Martin 2016a, page 21).

Table 5.12: Combination forecast approach (PWT 9.0 data)

Weights in the Optimal Proxy : Combination forecast (PWT 9)								
	P&S method Dependent variables				Combination forecast Dependent variables			
	(1) Baseline	(2) Log Light Density	(3) Calibrated Lights	(4) Fraction Pop. Lit	(5) Baseline	(6) Log Light Density	(7) Calibrated Lights	(8) Fraction Pop. Lit
No Fixed Effects								
Log GDP_PC PWT 9.0	.47***	.67***	.34**	.83***	.64***	.82***	.58***	.65***
	(.22, .71)	(.17, 1.1)	(.05, .64)	(.41, 1.2)	(.31, .96)	(.28, 1.35)	(.16, 1)	(.36, .94)
Log Survey	.52***	.32	.65***	.16	.36**	.18	.42**	.35**
	(.28, .77)	(-.17, .82)	(.35, .94)	(-.25, .58)	(.04, .69)	(-.35, .72)	(.00, .84)	(.06, .64)
Year Fixed Effects								
Log GDP_PC PWT 9.0	.45***	.65***	.34**	.81***	.62***	.8***	.56***	.64***
	(.21, .70)	(.17, 1.1)	(.06, .62)	(.40, 1.2)	(.29, .94)	(.26, 1.32)	(.16, .97)	(.35, .92)
Log Survey	.54***	.34	.65***	.18	.38**	.20	.44**	.36**
	(.29, .78)	(-.13, .82)	(.37, .93)	(-.23, .59)	(.04, .71)	(-.32, .74)	(.03, .84)	(.08, .65)
Country Fixed Effects								
Log GDP_PC PWT 9.0	.90***	.86***	.76***	.61	.56***	.64***	.84***	.34*
	(.47, 1.3)	(.54, 1.1)	(.39, 1.1)	(-.32, 1.5)	(.26, .86)	(.37, .91)	(.27, 1.4)	(-.02, .69)
Log Survey	.09	.13	.23	.38	.44***	.36***	.16	.66***
	(-.33, .52)	(-.18, .45)	(-.13, .60)	(-.55, 1.3)	(.14, .74)	(.09, .63)	(-.40, .73)	(.31, 1.02)
Country and Year Fixed Effects								
Log GDP_PC PWT 9.0	.85***	.93	.89**	10	.64***	.63***	.75***	.36*
	(.32, 1.3)	(-6.6, 8.5)	(.19, 1.5)	(-6.5. 26)	(.32, .97)	(.33, .94)	(.21, 1.28)	(-.01, .73)
Log Survey	.14	.06	.10	-9	.36**	.37**	.25	.64***
	(-.37, .67)	(-7.5, 7.6)	(-.58, .80)	(-25, 7.5)	(.03, .68)	(.06, .67)	(-.28, .79)	(.27, 1.01)
No. of Obs	680	680	680	155	680	680	680	155
No. of Clusters	115	115	115	79	115	115	115	79

Table 5.12 shows estimates of the relative weight of GDP per capita (PWT 9.0) and survey means in measuring true income per capita using P&S's method and combination method. Columns 1-4 are taken from columns 1, 5, 6 and 7 of Table 5.7. Columns 5-8 present the results of combination forecast approach using the same data as columns 1-4. Note that the numbers inside parentheses of columns 1-4 are block-bootstrapped 95% confidence intervals, while the numbers inside parentheses of column 5-8 are true confidence interval. Column 1 and 5 are baseline which means that these columns use aggregate light per capita. Columns 2-4 and 6-8 use three different types of light data which are described in detail in (Pinkovskiy and Sala-i-Martin 2016a, page 21).

Table 5.15 The estimates of p-value using t-test –WDI data to test the null hypothesis $w_G^* = 1$ for the results obtained from combination forecast method

	Dependent variables			
	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit
No Fixed Effects	0.5790	0.2482	0.2369	0.0838
Year Fixed Effects	0.6754	0.3306	0.2930	0.0537
Country Fixed Effects	0.2546	0.3465	0.1600	0.3054
Year and Country Fixed Effects	0.6143	0.3536	0.2079	0.1895

Table 5.16 The estimates of p-value using t-test –UQICD data to test the null hypothesis $w_G^* = 1$ for the results obtained from combination forecast method

	Dependent variables			
	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit
No Fixed Effects	0.7763	0.4197	0.2431	0.0769
Year Fixed Effects	0.8433	0.5381	0.2960	0.0470
Country Fixed Effects	0.0636	0.1128	0.5045	0.1541
Year and Country Fixed Effects	0.2444	0.1391	0.6470	0.0914

Table 5.17 The estimates of p-value using t-test –PWT 7.1 data to test the null hypothesis $w_G^* = 1$ for the results obtained from combination forecast method

	Dependent variables			
	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit
No Fixed Effects	0.4497	0.0704	0.3626	0.0889
Year Fixed Effects	0.5694	0.0999	0.4158	0.0586
Country Fixed Effects	0.1629	0.2195	0.3045	0.1434
Year and Country Fixed Effects	0.3757	0.2063	0.3925	0.0964

Table 5.18 The estimates of p-value using t-test –PWT 8.1 data to test the null hypothesis $w_G^* = 1$ for the results obtained from combination forecast method

	Dependent variables			
	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit
No Fixed Effects	0.0309	0.4141	0.0413	0.0120
Year Fixed Effects	0.0207	0.3736	0.0317	0.0076
Country Fixed Effects	0.0426	0.0578	0.7607	0.0284
Year and Country Fixed Effects	0.0728	0.0450	0.5850	0.0248

Table 5.19 The estimates of p-value using t-test –PWT 9 data to test the null hypothesis $w_G^* = 1$ for the results obtained from combination forecast method

	Dependent variables			
	Baseline	Log Light Density	Calibrated Lights	Fraction Pop. Lit
No Fixed Effects	0.0271	0.5014	0.0461	0.0157
Year Fixed Effects	0.0261	0.4429	0.0319	0.0128
Country Fixed Effects	0.0039	0.0084	0.5685	0.0003
Year and Country Fixed Effects	0.0298	0.0171	0.3526	0.0009

Table 6.2: Re-estimate P&S's model by World Bank income groups and geographical regions – UQICD data

Weights in the Optimal Proxy: Subsample (UQICD)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Poverty Level			Regions		
		Low income	Lower Middle income	Upper Middle income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects							
Log GDP_PC	.79***	.68	.53**	.82***	2.8	.87***	1.01***
UQICD	(.58, 1)	(-17, 19)	(.00, 1.06)	(.40, 1.2)	(-10E4, 10E4)	(.57, 1.1)	(.36, 1.6)
Log Survey	.2*	.31	.46*	.17	-1.8	.12	-.01
	(-.00, .41)	(-18, 18)	(-.06, .99)	(-.24, .59)	(-10E4, 10E4)	(-.18, .42)	(-.65, .63)
Year Fixed Effects							
Log GDP_PC	.77***	.41	.42	.84***	2.4	.84***	.96***
UQICD	(.55, .98)	(-5.8, 6.6)	(-.18, 1.02)	(.45, 1.2)	(-26, 31)	(.55, 1.1)	(.38, 1.5)
Log Survey	.22**	.58	.57*	.15	-1.4	.15	.03
	(.01, .44)	(-5.6, 6.8)	(-.02, 1.1)	(-.22, .54)	(-30, 27)	(-.13, .44)	(-.53, .61)
Country Fixed Effects							
Log GDP_PC	1.02***	1.7	1.08***	1.08***	1.3***	.67	.75***
UQICD	(.68, 1.36)	(-34, 37)	(.47, 1.6)	(.77, 1.3)	(.89, 1.7)	(-60, 61)	(.49, 1.01)
Log Survey	-.02	-.71	-.08	-.08	-.30	.32	.24*
	(-.36, .31)	(-36, 35)	(-.68, .52)	(-.39, .22)	(-.70, .10)	(-60, 61)	(-.01, .50)
Country and Year Fixed Effects							
Log GDP_PC	1.0***	4.2	1.05*	.93***	1.01*	.59	.66***
UQICD	(.57, 1.4)	(-18, 27)	(-.13, 2.2)	(.6, 1.2)	(-.03, 2.05)	(-4.2, 5.4)	(.30, 1.01)
Log Survey	-.004	-3	-.05	.06	-.01	.40	.33*
	(-.43, .42)	(-26, 19)	(-1.2, 1.1)	(-.26, .39)	(-1.05, 1.03)	(-4, 5)	(-.01, .69)
Number of Obs	680	79	215	296	82	136	227
Number of Clusters	115	26	37	39	17	48	22

Note: Table 6.2 shows estimates of the relative weight of GDP per capita (UQICD) and survey means in measuring true income per capita using P&S's method across income groups and region groups. Block-bootstrapped 95% confidence interval in parentheses. Column 1 is baseline specification using aggregate data (the same as column 1 of Table 5.4 and Table 5.9). Columns 2-4 present the results across income groups. Columns 5-7 show the results across region groups.

Table 6.3: Re-estimate P&S's model by World Bank income groups and geographical regions – PWT 7.1 data

Weights in the Optimal Proxy: Subsample (PWT 7.1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Poverty Level			Regions		
		Low income	Lower Middle income	Upper Middle income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects							
Log GDP_PC PWT 7.1	.86***	.99	.51**	.92***	2.5	.94***	.81***
	(.62, 1.09)	(-17, 19)	(.06, .97)	(.49, 1.3)	(-9, 14)	(.6, 1.28)	(.19, 1.4)
Log Survey	.13	.003	.48**	.07	-1.5	.05	.18
	(-.09, .37)	(-18, 18)	(.02, .93)	(-.36, .5)	(-13, 10)	(-.28, .39)	(-.42, .80)
Year Fixed Effects							
Log GDP_PC PWT 7.1	.82***	.77	.41*	.90***	2.3	.91***	.75**
	(.59, 1.05)	(-69, 70)	(-.04, .87)	(.48, 1.3)	(-10, 15)	(.62, 1.2)	(.16, 1.3)
Log Survey	.17	.22	.58**	.09	-1.3	.08	.24
	(-.05, .40)	(-69, 70)	(.12, 1.04)	(-.32, .51)	(-14, 11)	(-.20, .37)	(-.33, .83)
Country Fixed Effects							
Log GDP_PC PWT 7.1	1.05***	1.9	1.04**	1.08***	1.11***	1.04	.66***
	(.69, 1.4)	(-28, 32)	(.09, 1.9)	(.76, 1.4)	(.78, 1.43)	(-4.8, 6.9)	(.24, 1.08)
Log Survey	-.05	-.97	-.04	-.08	-.11	-.04	.33
	(-.40, .30)	(-31, 29)	(-.99, .90)	(-.40, .23)	(-.43, .21)	(-5.9, 5.8)	(-.08, .75)
Country and Year Fixed Effects							
Log GDP_PC PWT 7.1	1.02***	1.7	1.02	.94***	1.09	1.03	.30
	(.64, 1.3)	(-74, 78)	(-5.9, 8)	(.63, 1.2)	(-4.4, 6.6)	(-17, 19)	(-18, 18)
Log Survey	-.02	-.79	-.02	.05	-.09	-.032	.69
	(-.39, .35)	(-77, 75)	(-7.01, 6.9)	(-.24, .36)	(-5.6, 5.4)	(-18, 18)	(-17, 19)
Number of Obs	680	79	215	296	82	136	227
Number of Clusters	115	26	37	39	17	48	22

Note: Table 6.3 shows estimates of the relative weight of GDP per capita (PWT 7.1) and survey means in measuring true income per capita using P&S's method across income groups and region groups. Block-bootstrapped 95% confidence interval in parentheses. Column 1 is baseline specification using aggregate data (the same as column 1 of Table 5.5 and Table 5.10). Columns 2-4 present the results across income groups. Columns 5-7 show the results across region groups.

Table 6.4: Re-estimate P&S's model by World Bank income groups and geographical regions – PWT 8.1 data

Weights in the Optimal Proxy: Subsample (PWT 8.1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Poverty Level			Regions		
		Low income	Lower Middle income	Upper Middle income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects							
Log GDP_PC PWT 8.1	.49***	1.2	.27	.55**	1.8	.61***	.65
	(.2, .74)	(-1.2, 3.7)	(-.06, .61)	(.00, 1.09)	(-10, 14)	(.22, .99)	(-2.5, 3.8)
Log Survey Mean	.50***	-.22	.72***	.44	-.87	.38**	.34
	(.2, .76)	(-2.7, 2.2)	(.38, 1.06)	(-.09, .99)	(-13, 11)	(.00, .77)	(-2.8, 3.5)
Year Fixed Effects							
Log GDP_PC PWT 8.1	.47***	1.02	.23	.55*	1.8	.62***	.60
	(.23, .7)	(-54, 56)	(-.10, .58)	(-.01, 1.1)	(-6.9, 10)	(.23, 1.01)	(-1.8, 3.0)
Log Survey Mean	.52***	-.02	.76***	.44	-.80	.37*	.39
	(.28, .7)	(-55, 55)	(.41, 1.1)	(-.12, 1.01)	(-9.5, 7.9)	(-.01, .76)	(-2.04, 2.8)
Country Fixed Effects							
Log GDP_PC PWT 8.1	.91***	1.7	.62	1.17***	1.07***	.50	.52**
	(.4, 1.3)	(-18, 22)	(-.3, 1.5)	(.80, 1.5)	(.70, 1.4)	(-5, 6)	(-.00, 1.05)
Log Survey Mean	.08	-.78	.37	-.17	-.07	.49	.47*
	(-.35, .5)	(-21, 19)	(-.56, 1.3)	(-.54, .19)	(-.45, .29)	(-5, 6)	(-.054, 1.0)
Country and year Fixed Effects							
Log GDP_PC PWT 8.1	.88***	1.71	.69	.90***	1.09	1.6	-.05
	(.2, 1.4)	(-7.4, 10)	(-49, 50)	(.57, 1.2)	(-4.4, 6.6)	(-2, 5)	(-4.6, 4.5)
Log Survey Mean	.11	-.71	.30	.09	-.09	-.65	1.05
	(-.49, .7)	(-9.8, 8.4)	(-49, 50)	(-.23, .42)	(-5.6, 5.4)	(-4.7, 3.4)	(-3.5, 5.6)
Number of Obs	680	79	215	296	82	136	227
Number of Clusters	115	26	37	39	17	48	22

Note: Table 6.4 shows estimates of the relative weight of GDP per capita (PWT 8.1) and survey means in measuring true income per capita using P&S's method across income groups and region groups. Block-bootstrapped 95% confidence interval in parentheses. Column 1 is baseline specification using aggregate data (the same as column 1 of Table 5.6 and Table 5.11). Columns 2-4 present the results across income groups. Columns 5-7 show the results across region groups.

Table 6.5: Re-estimate P&S's model by World Bank income groups and geographical regions – PWT 9.0 data

Weights in the Optimal Proxy: Subsample (PWT 9)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Poverty Level			Regions		
		Low income	Lower Middle income	Upper Middle income	Asia and Pacific	Africa and Middle East	Latin America and Caribbean
No Fixed Effects							
Log GDP_PC PWT 9	.47***	1.2	.29*	.61**	2.07	.67***	1.03
	(.22, .71)	(-6.2, 8.8)	(-.03, .61)	(.06, 1.1)	(-59, 63)	(.39, .95)	(-.71, 2.7)
Log Survey Mean	.52***	-.2	.70***	.38	-1.07	.32**	-.03
	(.28, .77)	(-7.8, 7.2)	(.38, 1.0)	(-.17, .93)	(-62, 60)	(.04, .60)	(-1.7, 1.7)
Year Fixed Effects							
Log GDP_PC PWT 9	.45***	1.0	.27*	.64**	1.9	.67***	.98*
	(.21, .70)	(-7.3, 9.5)	(-.05, .60)	(.08, 1.1)	(-14, 18)	(.38, .97)	(-.11, 2.0)
Log Survey Mean	.54***	-.09	.72***	.35	-.91	.32**	.01
	(.29, .78)	(-8.5, 8.3)	(.39, 1.05)	(-.19, .91)	(-17, 15)	(.02, .61)	(-1.0, 1.1)
Country Fixed Effects							
Log GDP_PC PWT 9	.90***	1.7	.35	1.07***	1.2***	.49	.45**
	(.47, 1.3)	(-13, 17)	(-.11, .82)	(.50, 1.6)	(.7, 1.7)	(-37, 38)	(.00, .90)
Log Survey Mean	.09	-.74	.64***	-.07	-.22	.50	.54**
	(-.33, .52)	(-16, 14)	(.17, 1.1)	(-.65, .49)	(-.72, .26)	(-37, 38)	(.09, .99)
Country and year Fixed Effects							
Log GDP_PC PWT 9	.85***	2.7	.59	.79**	1.0	1.2	.36
	(.32, 1.3)	(-59, 64)	(-1E4, 1E4)	(.11, 1.4)	(-2.5, 4.7)	(-49, 52)	(-.63, 1.3)
Log Survey Mean	.14	-1.7	.40	.20	-.09	-.22	.63
	(-.37, .67)	(-63, 60)	(-1E4, 1E4)	(-.47, .88)	(-3.7, 3.5)	(-51. 50)	(-.35, 1.6)
Number of Obs	680	79	215	296	82	136	227
Number of Clusters	115	26	37	39	17	48	22

Note: Table 6.5 shows estimates of the relative weight of GDP per capita (PWT 9.0) and survey means in measuring true income per capita using P&S's method across income groups and region groups. Block-bootstrapped 95% confidence interval in parentheses. Column 1 is baseline specification using aggregate data (the same as column 1 of Table 5.7 and Table 5.12). Columns 2-4 present the results across income groups. Columns 5-7 show the results across region groups.

Table 6.7: The estimates of p-value using Wald test – UQICD data to test the null hypothesis $w_G^* = 1$ for individual group

	Income groups			Region groups		
	Low income	Lower Middle Income	Upper Middle Income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects	0.3239	0.0568	0.4278	0.2870	0.4235	0.9712
Year Fixed Effects	0.2309	0.0405	0.4651	0.3221	0.3265	0.9045
Country Fixed Effects	0.7857	0.7615	0.6313	0.1710	0.6977	0.1074
Country and Year Fixed Effects	0.8381	0.7350	0.9484	0.5009	0.8008	0.0747

Table 6.8: The estimates of p-value using Wald test – PWT 7.1 data to test the null hypothesis $w_G^* = 1$ for individual group

	Income groups			Region groups		
	Low income	Lower Middle Income	Upper Middle Income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects	0.9857	0.0166	0.7639	0.2378	0.7428	0.5221
Year Fixed Effects	0.7156	0.0078	0.7629	0.2898	0.6177	0.3405
Country Fixed Effects	0.4091	0.8695	0.6117	0.5479	0.9373	0.1389
Country and Year Fixed Effects	0.3828	0.8596	0.9599	0.9607	0.8151	0.1133

Table 6.9: The estimates of p-value using Wald test – PWT 8.1 data to test the null hypothesis $w_G^* = 1$ for individual group

	Income groups			Region groups		
	Low income	Lower Middle Income	Upper Middle Income	Asia and Pacific	Africa & Middle East	Latin America & Caribbean
No Fixed Effects	0.4381	0.0000	0.1115	0.3460	0.0383	0.3174
Year Fixed Effects	0.6362	0.0000	0.1210	0.4248	0.0228	0.1840
Country Fixed Effects	0.3847	0.4256	0.4543	0.6721	0.4832	0.0714
Country and Year Fixed Effects	0.4739	0.5787	0.8780	0.8876	0.9772	0.0475

Code Appendix

STATA CODE – FIGURE 1.1 (SCATTER PLOT OF LOG GDP PER CAPITA (WDI) AND SURVEY MEANS WITH LOG NTL)

```
Scatter lrgdpchWDI lrgdpchsurveys lrgdpchlight if lrgdpchlight>=12.5 & lrgdpchlight<=18.5 || lfit lrgdpchWDI lrgdpchlight, lc(blue) range(12.6 18.5) || lfit lrgdpchsurveys lrgdpchlight , lc(red)  range(12.6 18.5) title("WDI, Survey Means and Lights") ytitle("Log GDP per Capita (WDI) / Log Survey Mean") ylabel(,angle(horiz)) xtitle("Log Lights per Capita") legend(order(1 2 ) label(1 "Log GDP per Capita (WDI)") label(2 "Log Survey Means")) note("Lines are regressions of measurement variable on log nighttime lights per capita")
```

STATA CODE – TABLE 5.1 (WDI)

Column 1

```
import excel "C:\Users\thaonguyen\Desktop\thesis\data chuan\data Table1 full.xlsx", sheet("Table1") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

reg lrgdpchlight lrgdpchWDI, cluster(iso_n)

tab year, gen(t)

reg lrgdpchlight lrgdpchWDI t*, cluster(iso_n)

encode cname, gen(C)

tab c, gen(country)

reg lrgdpchlight lrgdpchWDI t* country*, cluster(iso_n)
```

Column 6

```
ivreg lrgdpchsurveys  (lrgdpchWDI =lrgdpchlight), cluster(iso_n)

ivreg lrgdpchsurveys  (lrgdpchWDI =lrgdpchlight) t*, cluster(iso_n)

ivreg lrgdpchsurveys  (lrgdpchWDI =lrgdpchlight) country*, cluster(iso_n)

ivreg lrgdpchsurveys  (lrgdpchWDI =lrgdpchlight t* country*, cluster(iso_n)
```

STATA CODE – TABLE 5.2 (WDI)

```
import excel "C:\Users\thaonguyen\Desktop\thesis\data chuan\data Table1 full.xlsx", sheet("Table1") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

reg lrgdpchlight lrgdpchWDI lrgdpchsurveys, cluster(iso_n)
```

```
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, cluster(iso_n)

reg lrgdpchlight lrgdpchWDI lrgdpchsurveys dcountry*, cluster(iso_n)

reg lrgdpchlight lrgdpchWDI lrgdpchsurveys dcountry* t*, cluster(iso_n)
```

STATA CODE – TABLE 5.3 (WDI)

Column 1

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Tableand scatter plot\Table2 official\data Table2.xlsx",
sheet("column1") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps($(Rati))
cluster(iso_n): areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 2

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Tableand scatter plot\Table2 official\data Table2.xlsx",
sheet("column2") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)
```

```
cap gen lSelectp=lSelect-lSpoptotal

global B=200

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): reg lrgdpchlight lrgdpchWDI lrgdpchsurveys lSelectp, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): areg lrgdpchlight lrgdpchWDI lrgdpchsurveys lSelectp, absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): areg lrgdpchlight lrgdpchWDI lrgdpchsurveys lSelectp, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): areg lrgdpchlight lrgdpchWDI lrgdpchsurveys lSelectp t*,  absorb(iso_n) cluster(iso_n)
```

Column 3

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column3") firstrow

global iso_n iso_n

global year year

global controls21 "lSpoptotal lSgdpnrg lSarea lSararea Slat Slon lSoil lSelect lSpoprural lSpopurban lSservices lSagr
lSexports lSimports lSmanuf lShcfa lSgovexp lSgrosscapform lSrichshare lSpoorshare"

global B=200

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps($(Alan Heston et
al.)) cluster(iso_n): reg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} , absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} t* , absorb(iso_n) cluster(iso_n)
```

Column 4

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column4") firstrow

global controls21 "lSpoptotal lSgdpnrg lSarea lSararea Slat Slon lSoil lSelect lSpoprural lSpopurban lSservices lSagr lSexports lSimports lSmanuf lShcfa lSgovexp lSgrosscapform lSrichshare  lSpoorshare"

global controls22 "lSpoptotal2 lSgdpnrg2 lSarea2 lSararea2 Slat2 Slon2 lSpoprural2 lSoil2 lSelect2 lSpopurban2 lSservices2 lSagr2 lSexports2 lSimports2 lSmanuf2 lShcfa2 lSgovexp2 lSgrosscapform2  lSrichshare2 lSpoorshare2"

global B=200

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} ${controls22}, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} ${controls22}, absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} ${controls22}, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controls21} ${controls22} t*, absorb(iso_n) cluster(iso_n)
```

Column 5

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column5") firstrow

global B=200

*preserve

*replace lrgdpchWDI=lrgdpchWDI+lSpoptotal-lSarea

*replace lrgdpchsurveys=lrgdpchsurveys+lSpoptotal-lSarea

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year
```

```
tsset, clear

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlightdens lrgdpchWDI lrgdpchsurveys, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightdens lrgdpchWDI lrgdpchsurveys, absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightdens lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightdens lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 6

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column6") firstrow

global B=200

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year,  gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlightcal lrgdpchWDI lrgdpchsurveys, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightcal lrgdpchWDI lrgdpchsurveys, absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightcal lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightcal lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 7

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column7") firstrow

global B=200
```

```
global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlightfrac lrgdpchWDI lrgdpchsurveys,  cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightfrac lrgdpchWDI lrgdpchsurveys, absorb(year)  cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightfrac lrgdpchWDI lrgdpchsurveys, absorb(iso_n)  cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlightfrac lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n)  cluster(iso_n)
```

Column 8

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column8") firstrow

global B=200

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)

preserve

replace lrgdpchWDI=lrgdpchWDIc

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controlset}, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controlset}, absorb(year) cluster(iso_n)
```

```
noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controlset}, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys ${controlset} t*, absorb(iso_n) cluster(iso_n)
```

Column 9

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column9") firstrow

global B=200

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column10

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column10") firstrow

global B=200

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)
```

```
replace lrgdpchWDI=lrgdpchWDIc

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 11

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column11") firstrow

global B=200

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)

replace lrgdpchWDI=lrgdpchWDIc if indinc==0

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys, cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])),  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 12

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column12") firstrow

global iso_n iso_n
```

```
global year year

sort iso_n year

xtset iso_n year

tsset, clear

tab year, gen(t)

global B=200

by isoc, so: egen nsurv=count(lrgdpchsurveys)

*gen sigma=sqrt(2)*invnormal((1+gini/100)/2)

*gen povsurv=popWDI*normal((ln(1.25*365.25)-lrgdpchsurveys)/sigma)

by isoc, so: egen avgpop=mean(popWDI)

gen weight=avgpop/nsurv

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI] +_b[lrgdpchsurveys])),  level(95) seed(777) reps(200) cluster(iso_n)
force: reg lrgdpchlight lrgdpchWDI lrgdpchsurveys [aw=weight], cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI] +_b[lrgdpchsurveys])),  level(95) seed(777) reps(200) cluster(iso_n)
force: areg lrgdpchlight lrgdpchWDI lrgdpchsurveys [aw=weight], absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI] +_b[lrgdpchsurveys])),  level(95) seed(777) reps(200) cluster(iso_n)
force: areg lrgdpchlight lrgdpchWDI lrgdpchsurveys [aw=weight], absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI] +_b[lrgdpchsurveys])),  level(95) seed(777) reps(200) cluster(iso_n)
force: areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t* [aw=weight], absorb(iso_n) cluster(iso_n)
```

STATA CODE – TABLE 5.8

Column 5

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column1") firstrow

gen YLG= lrgdpchlight – lrgdpchWDI

gen YSG = lrgdpchsurveys -  lrgdpchWDI

gen YLS= lrgdpchlight – lrgdpchsurveys

gem YGS = lrgdpchWDI - lrgdpchsurveys

tab year, gen(t)

encode cname, gen©

tab c, gen(country)

reg YLG YSG, cluster(iso_n)
```

```
reg YLG YSG t*, cluster(iso_n)
reg YLG YSG country*, cluster(iso_n)
reg YLG YSG t* country*, cluster(iso_n)
reg YLS YGS, cluster(iso_n)
reg YLS YGS t*, cluster(iso_n)
reg YLS YGS country*, cluster(iso_n)
reg YLS YGS t* country*, cluster(iso_n)
```

Column 6

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column6") firstrow
gen YLG= lrgdpchlightdens - lrgdpchWDI
gen YSG = lrgdpchsurveys -  lrgdpchWDI
gen YLS= lrgdpchlightdens - lrgdpchsurveys
gem YGS = lrgdpchWDI - lrgdpchsurveys
tab year, gen(t)
encode cname, gen©
tab c, gen(country)
reg YLG YSG, cluster(iso_n)
reg YLG YSG t*, cluster(iso_n)
reg YLG YSG country*, cluster(iso_n)
reg YLG YSG t* country*, cluster(iso_n)
reg YLS YGS, cluster(iso_n)
reg YLS YGS t*, cluster(iso_n)
reg YLS YGS country*, cluster(iso_n)
reg YLS YGS t* country*, cluster(iso_n)
```

Column 7

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column7") firstrow
gen YLG= lrgdpchlightcal - lrgdpchWDI
gen YSG = lrgdpchsurveys -  lrgdpchWDI
gen YLS= lrgdpchlightcal - lrgdpchsurveys
gem YGS = lrgdpchWDI - lrgdpchsurveys
tab year, gen(t)
```

```
encode cname, gen©
tab c, gen(country)
reg YLG YSG, cluster(iso_n)
reg YLG YSG t*, cluster(iso_n)
reg YLG YSG country*, cluster(iso_n)
reg YLG YSG t* country*, cluster(iso_n)
reg YLS YGS, cluster(iso_n)
reg YLS YGS t*, cluster(iso_n)
reg YLS YGS country*, cluster(iso_n)
reg YLS YGS t* country*, cluster(iso_n)
```

Column 8

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column7") firstrow
gen YLG= lrgdpchlightfrac - lrgdpchWDI
gen YSG = lrgdpchsurveys -  lrgdpchWDI
gen YLS= lrgdpchlightfrac - lrgdpchsurveys
gem YGS = lrgdpchWDI - lrgdpchsurveys
tab year, gen(t)
encode cname, gen©
tab c, gen(country)
reg YLG YSG, cluster(iso_n)
reg YLG YSG t*, cluster(iso_n)
reg YLG YSG country*, cluster(iso_n)
reg YLG YSG t* country*, cluster(iso_n)
reg YLS YGS, cluster(iso_n)
reg YLS YGS t*, cluster(iso_n)
reg YLS YGS country*, cluster(iso_n)
reg YLS YGS t* country*, cluster(iso_n)
```

STATA CODE – TABLE 5.15

Column 5

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column1") firstrow
gen YLS= lrgdpchlight – lrgdpchsurveys
```

```
gem YGS = lrgdpchWDI - lrgdpchsurveys

tab year, gen(t)

encode cname, gen©

tab c, gen(country)

reg YLS YGS, cluster(iso_n)

test YGS=1

reg YLS YGS t*, cluster(iso_n)

test YGS=1

reg YLS YGS country*, cluster(iso_n)

test YGS=1

reg YLS YGS t* country*, cluster(iso_n)

test YGS=1
```

Column 6

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column6") firstrow

gen YLS= lrgdpchlightdens - lrgdpchsurveys

gem YGS = lrgdpchWDI - lrgdpchsurveys

tab year, gen(t)

encode cname, gen©

tab c, gen(country)

reg YLS YGS, cluster(iso_n)

test YGS=1

reg YLS YGS t*, cluster(iso_n)

test YGS=1

reg YLS YGS country*, cluster(iso_n)

test YGS=1

reg YLS YGS t* country*, cluster(iso_n)

test YGS=1
```

Column 7

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column7") firstrow

gen YLS= lrgdpchlightcal - lrgdpchsurveys

gem YGS = lrgdpchWDI - lrgdpchsurveys
```

```
tab year, gen(t)

encode cname, gen©

tab c, gen(country)

reg YLS YGS, cluster(iso_n)

test YGS=1

reg YLS YGS t*, cluster(iso_n)

test YGS=1

reg YLS YGS country*, cluster(iso_n)

test YGS=1

reg YLS YGS t* country*, cluster(iso_n)

test YGS=1
```

Column 8

```
import excel "C:\Users\thaonguyen\Desktop\data Table2.xlsx", sheet("column7") firstrow

gen YLS= lrgdpchlightfrac - lrgdpchsurveys

gem YGS = lrgdpchWDI - lrgdpchsurveys

tab year, gen(t)

encode cname, gen©

tab c, gen(country)

reg YLS YGS, cluster(iso_n)

test YGS=1

reg YLS YGS t*, cluster(iso_n)

test YGS=1

reg YLS YGS country*, cluster(iso_n)

test YGS=1

reg YLS YGS t* country*, cluster(iso_n)

test YGS=1
```

STATA CODE – TABLE 6.1

Column 1

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Tableand scatter plot\Table2 official\data Table2.xlsx",
sheet("column1") firstrow

global iso_n iso_n
```

```
global year year

sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 2

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx", sheet("low
income ") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)
```

```
noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 3

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx", sheet("lower
middle income") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 4

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx",
sheet("upper middle incomme") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)
```

```
noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 5

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx", sheet("Asia
Pacific ") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 6

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx",
sheet("Africa Middle East filter")  firstrow

global iso_n iso_n

global year year
```

```
sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

Column 7

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx", sheet("Latin
American Carribean ") firstrow

global iso_n iso_n

global year year

sort iso_n year

xtset iso_n year

tsset, clear

global B=200

tab year, gen(t)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
reg lrgdpchlight lrgdpchWDI lrgdpchsurveys , cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys , absorb(year) cluster(iso_n)

noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys, absorb(iso_n) cluster(iso_n)
```

```
noi bootstrap NA=(_b[lrgdpchWDI]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys]))
Surveys=(_b[lrgdpchsurveys]/(_b[lrgdpchWDI]+_b[lrgdpchsurveys])) ,  level(95) seed(777) reps(${B}) cluster(iso_n):
areg lrgdpchlight lrgdpchWDI lrgdpchsurveys t*, absorb(iso_n) cluster(iso_n)
```

STATA CODE – TABLE 6.6

Income group

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx", sheet("Wald test income ") firstrow

gen y1= lrgdpchWDI *D1

gen y3= lrgdpchWDI *D2

gen y5= lrgdpchWDI *D3

gen y2= lrgdpchsurveys *D1

gen y4= lrgdpchsurveys *D2

gen y6= lrgdpchsurveys *D3

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6, cluster (iso_n)

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1

tab year, gen(t)

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6 t*, cluster (iso_n)

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1

encode cname, gen(c)

tab c, gen(country)

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6 country*, cluster (iso_n )

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6 country* t*, cluster (iso_n )

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1
```

Region group

```
import excel "C:\Users\thaonguyen\Desktop\thesis\Table3 income and region level\Table3 data.xlsx", sheet("Wald test region") firstrow

gen y1= lrgdpchWDI *D1

gen y3= lrgdpchWDI *D2

gen y5= lrgdpchWDI *D3

gen y2= lrgdpchsurveys *D1

gen y4= lrgdpchsurveys *D2

gen y6= lrgdpchsurveys *D3

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6, cluster (iso_n)

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1

tab year, gen(t)

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6 t*, cluster (iso_n)

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1

encode cname, gen(c)

tab c, gen(country)

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6 country*, cluster (iso_n )

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1

reg lrgdpchlight D1 D2 D3 y1 y3 y5 y2 y4 y6 country* t*, cluster (iso_n )

testnl _b[y1]/(_b[y1]+ _b[y2])=1

testnl _b[y3]/(_b[y3]+ _b[y4])=1

testnl _b[y5]/(_b[y5]+ _b[y6])=1
```