Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates

Simon Johnson[a], William Larson[b], Chris Papageorgiou[c], Arvind Subramanian[d*]

[a] MIT, Peterson Institute for International Economics, and NBER

[b] George Washington University

[c] IMF

[d] Peterson Institute for International Economics, Center for Global Development, and Johns Hopkins University

Received Date; Received in Revised Form Date; Accepted Date

Abstract

This paper sheds light on two problems in the Penn World Table (PWT) GDP estimates. First, we show that these estimates vary substantially across different versions of the PWT despite being derived from very similar underlying data and using almost identical methodologies; that the methodology deployed to estimate growth rates leads to systematic variability, which is greater: at higher data frequencies, for smaller countries, and the farther the estimate from the benchmark year. Moreover, this variability matters for the cross-country growth literature. While growth studies that use low frequency data remain robust to data revisions, studies that use annual data are less robust. Second, the PWT methodology leads to GDP estimates that are not valued at purchasing power parity (PPP) prices. This is surprising because the raison d'être of the PWT is to adjust national estimates of GDP by valuing output at common international (purchasing power

*Special thanks go to Angus Deaton, Alan Heston, David Romer, and David Weil, for comments, encouragement and suggestions. We also benefited from discussions with Daron Acemoglu, Craig Burnside, Francesco Caselli, Michael Clemens, Mario Crucini, Erwin Diewert, Robert Feenstra, Benjamin Jones, Chad Jones, Aart Kraay, Benjamin Olken, Valerie Ramey, Mick Silver, Antonio Spilimbergo, Nancy Stokey, Ping Wang, Alwyn Young and seminar participants at Brown University, the Bureau of Economic Analysis, IMF, Oxford University, the University of Pennsylvania (2008 Penn World Table workshop), Princeton University, Tilburg University, the World Bank, the Econometric Society Summer Meeting (Boston 2009), the NBER Summer Institute (Economic Growth workshop, and Wealth and Inequality workshop 2009), the Society for Economic Dynamics Meeting (Istanbul 2009), the conference on Measuring and Analyzing Economic Development, (University of Chicago, 2010), and the Midwest Macro Meetings (Vanderbilt University, 2011). The views expressed here are those of the authors.

parity [PPP]) prices so that the resulting PPP-adjusted estimates of GDP are comparable across countries. We propose an approach to address these two problems of variability and valuation.

Keywords: Penn World Table, GDP estimates, purchasing power parity, revisions, methodology, growth

JEL classification: O11, O40, O47

---

1. Introduction

How fast did Equatorial Guinea grow over the two and a half decades beginning in 1975? The natural place to turn to answer such a question is data from the Penn World Table (PWT), which is the most widely used source for cross-country comparisons for the level and growth rate of GDP.[1] According to one version (PWT 6.2), Equatorial Guinea is the second-fastest growing country among 40 African countries. However, according to the previous version (PWT 6.1), which was released four years before, Equatorial Guinea was the slowest growing country. Indeed, as table 1 shows, if one were to compile the list of the 10 fastest and slowest growing countries in Africa between 1975 and 1999, PWT 6.1 and PWT 6.2 would produce almost disjoint lists. Both would agree that Botswana has done best, and they would also agree that Egypt, Cape Verde, Lesotho, Mauritius, Morocco, and Tunisia belong in the list of fast growing countries. But there are six countries that are not identified by both versions as fast growing. The disagreement is even more severe for the "worst performers." There are a total of 10 countries that appear to have the slowest growth according to one version but not according to the other. Such data variability is especially surprising because PWT 6.1 and 6.2 use very similar underlying data and methodologies.[2] The concerns about GDP data quality and the variability it engenders have recently led researchers to explore

[1] Roughly two-thirds of all cross-country empirical work is based on PWT. Second place is held by the World Bank's World Development Indicators (WDI), which were originally based on the PWT but have subsequently diverged. The IMF's World Economic Outlook (WEO) dataset places a distant third.

[2] There are similarly dramatic differences if one were to compare estimates in the latest version (PWT 7) and those in PWT 6.1. In addition to the switch on Equatorial Guinea's status noted above, Malawi moves from being among the bottom ten growth performers according to PWT 6.1 to being among the top ten performers according to PWT 7.

alternatives to GDP data, and Henderson, Storeygard, and Weil (2009) and Young (2009) are noteworthy examples.

Before we examine this problem of data variability, we need to describe briefly what the PWT was designed to do. The pioneering work of Irving Kravis, Alan Heston, and Robert Summers, which led to the Penn World Table data, was aimed at converting national measures of GDP and income into internationally comparable estimates. Cross-country comparisons could not be based on national GDP data because these were valued at domestic prices. Since some goods and especially services were known to be cheaper in poor countries compared to rich countries, adjustments needed to be made to the valuation of these goods and services so that they could be made internationally comparable. These adjustments were made by calculating common international prices—the so-called purchasing power parity (PPP) prices—for all goods and services. With these PPP adjustments, GDP could then be compared across countries.

A large literature has assessed the basic methodology employed by the PWT for determining these PPPs. In addition to the series of papers by Kravis, Heston, and Summers (1978) and Summers and Heston (1980, 1991, and 1996), notable contributions include Ciccone and Jarocinski (2008), Deaton (2006), Dowrick (2005), Dowrick and Quiggin (1997), Feenstra, Heston, Timmer and Deng (2009), Heston (1994), Neary (2004), Nuxoll (1994), Rao and Selvanathan (1992), Samuelson (1994), Srinivasan (1994), and van Veelen (2002).

However, much of the discussion and criticism of the PWT methodology, including most recently by Deaton (2010), Deaton and Heston (2010) and Almas (2012), has focused on the PPP and GDP estimates for the benchmark year for which disaggregated data are collected from the countries. Only recently has attention been given to the intertemporal dimension of the PWT and its methodology, namely its estimates for the *growth* rates of PPP-adjusted GDP, which then has implications for PPP-adjusted GDP levels for nonbenchmark years. Feenstra, Heston, Timmer and Deng (2009) focus on the problems for PWT created by revisions to net exports. Feenstra, Ma, and Rao (2009) derive theory-consistent intertemporal data for consumption (but not for GDP) while Rao, Rambaldi and Doran (2009) use an econometric estimation procedure for deriving intertemporal estimates for the PPPs and GDP.

Most notably Katayama and Ponomareva (2009) were the first to explore the variability of growth rates resulting from revisions in PWT. More precisely, this paper documents how growth rates change across four versions of the PWT – 5.0, 5.5, 5.6 and 6.1 – and consequently analyzes the effects of these revisions on the main empirical finding of a negative relationship between volatility and growth in Ramey and Ramey (1995). Katayama and Ponomareva report that across the four different versions of PWT, 13 percent of their sample of 110 countries show a sign change in their annual average growth rates for the period 1962-1985. They also show that the Ramey and Ramey finding supported in some versions of the PWT but not in others.

The focus of this paper, as in Katayama and Ponomareva (2009), is the time dimension of the PWT, specifically the data revisions across versions of the PWT. Unlike Katayama and Ponomareva, however, we concentrate on the sources of variation in growth rates between versions 6.1 and 6.2. The main focus of our paper is the PWT methodology for constructing growth rates of GDP, and the resulting estimates for the levels of GDP for nonbenchmark years. It draws attention to two problems—variability and valuation—in the PWT's estimates of GDP (growth and level) and PPPs.

Our summary findings are: First, data revisions are substantial. This is particularly surprising because, as we describe below, versions 6.1 and 6.2 differ in relatively small ways in terms of data and methodology. Second, revisions are systematic and inherent to the PWT's methodology for computing growth rates and estimates for the nonbenchmark years. Revisions are more pronounced at high frequency (annual data are much more variable than longer averages); for small countries (i.e., countries with small total GDP); and for historical data, so that the further the data are from the benchmark year in the PWT, the more variable they are.

Third, the variability of growth data has implications for the cross-country growth literature. Results are indeed robust in the long-run (30-year averages) across versions of PWT which is a positive and hopeful result, given the magnitude of both PWT and the ICP and related literature. However, results based on annual data prove to be less robust across versions of the PWT than are results based on 10-year averages and/or levels of GDP. Results are also sensitive to sample, especially the inclusion of small countries.

Fourth, the PWT methodology raises a more basic question about valuation. The rationale for the PWT is to come up with GDP level and growth data that are at common international (the so-called PPP) prices so that the data are comparable across countries. The methodology, however, leads to the construction of GDP growth estimates that are based not on common international prices but on a mixture of international and domestic prices; as a result GDP level estimates for years other than the benchmark year are also not at international or PPP prices. In this case, it is not obvious that the data are comparable across countries.

Finally, we propose an alternative way of using the PWT data that might address the problems of both variability and valuation. Essentially, we urge the construction of a new chained series, with all data valued at common international prices, and based on greater use of the disaggregated data collected for the different benchmark years.

This paper is organized as follows: Section 2 briefly reviews the background and history of the PWT and illustrates key aspects of the methodology used in the PWT calculations of GDP growth rates. Section 3 shows the variability of growth data and the underlying patterns to this variability, focusing on versions 6.2 and 6.1 of the Table. In section 4, we summarize the results of our robustness/replication studies for leading growth studies. In section 5, we explain why not all PWT data fully reflect PPP prices. Section 6 describes our proposal that addresses the problems with PWT methodology relating to data variability and to data not being at PPP prices. Section 7 draws some conclusions, including offering some practical suggestions for researchers. Technical appendix 1 contains methodological details on the construction of the PWT data. Technical appendix 2 illustrates the valuation problem, namely why the current PWT methodology leads to GDP level and growth estimates that are not at international prices (both technical appendixes are included in an online supplement to this paper that is publicly available).

2. Penn World Table: Background and Methodology

A. Background and history

The PWT has reported on seven rounds of data, starting in 1970 (table 2). The latest published version is version 7, which was released in May 2011. The three previous versions were 6.3, 6.2, and 6.1. In this paper we focus mainly on the comparisons between the estimates in PWT 6.2 and 6.1. The reason is that the underlying price data used for the construction of PPP-adjusted GDP growth rates are the same, and the methodology deployed is very similar. Our demonstration of variability is thus rendered sharper because so many underlying factors are held constant.

We have, however, undertaken all the comparisons for the two newer versions of the data released after 6.2: PWT 6.3 and PWT 7. If we were to present all possible comparisons, the exposition risks becoming intractable. Therefore, we present all results for the 6.2 vs 6.1 comparisons, indicating as appropriate, where the results are different for estimates based on PWT 7.[3]

[3]It is worth pointing out that PWT version 6.3, which was released in August 2009, was quickly followed by PWT version 7, which was released in May 2011. The former was still based on ICP benchmark data for 1996, like versions 6.1 and 6.2, while PWT 7 used data from the ICP benchmark data for 2005. It thus seems more useful and relevant to add PWT 7 to the comparison exercise than PWT 6.3. It is worth noting that version 6.3 changed some of the methodology used for computing growth rates in large part to respond to the findings in the working paper version of this paper. Results not presented here (including comparisons involving PWT 6.3) are available from the authors upon request.

The core purpose of the PWT is to produce measures of real GDP from the corresponding relative price levels across countries and over time based on prices for the same or similar goods in different countries. The price collection operation, which is at the heart of PWT, is known as the International Comparison Programme/Project (ICP).[4] Each "generation" of the Table is based on a different round of the ICP; so, for example, versions 5.6 and 6.1 use different prices. The benchmark year of the PWT corresponds to the year for which the ICP exercise is carried out. The number of countries participating in the ICP has increased steadily, starting from 10 countries in 1970 to 146 countries in 2005. From table 2, it can be seen that the number of countries for which GDP estimates are provided by the PWT exceeds, often considerably, the number of countries participating in the ICP. Within a generation of the PWT, the ICP remains the same, e.g., this is the case for 6.2 and 6.1, but there are other revisions—the nature of which varies.

The ICP is a massive undertaking, requiring a vast amount of resources—mostly in terms of people's time, but the computer resources, at least until recently, were also significant. The only organizations that can sustain such an investment are government-funded, and even international organizations, such as the United Nations, have a hard time coming up with all the money required.[5] A substantial amount of the needed investment has come directly from governments.

As a result, control of the ICP has shifted over time. "After phase III [1975], the role of the University of Pennsylvania, which had until then been the main engine of the ICP, was gradually transformed into that of adviser on methodological issues. Another notable change in ICP responsibility was the increasing role of the Statistical Office of the European Communities (Eurostat). Eurostat, in fact, became not only the organizer of the European

[4] The International Comparison Project began in 1968, although its antecedents date back to the 1950s. Irving Kravis was the first director; http://unstats.un.org/unsd/methods/icp/ipco_htm.htm. In 1989 the "P" became Programme, rather than Project. International price comparisons have been completed for 1970, 1975, 1980, 1985, 1990, 1996, and 2005.

[5] The Ford Foundation did provide critical early financial support through grants to the University of Pennsylvania.

Community comparison, but also, with its experienced staff, it has provided substantial technical assistance to a number of regional comparisons and to the work on establishing links among the various regions."[6]

As control has shifted, so have preferences regarding methodology both for particular regions and for how these are aggregated to global estimates. The most important change may have been regionalization: "In phase IV [1980] and onward, countries participated through regions or country groups; first regional (e.g., African, OECD, etc.) comparisons were carried out and then the world comparison was built up by linking across these groups."[7] Deaton and Heston (2008) focus on problematic methodological points, all of which arise because particular organizations control part of the methods. Their concern suggests that the methodology changes significantly such that it is harder to say if the PWT is improving or not in its accuracy. But there are other reasons to worry about this exact same issue.

B. Methodology

The innovation and great contribution of the PWT was to convert national measures of GDP and income into internationally comparable PPP estimates. This is done—in principle—by collecting prices for the same or similar goods in different countries and deriving price indices that can be used to compare what people can actually buy. The PWT obtains local currency data from the national income accounts of countries. Then, based on international price comparisons, it converts these local currency data into PPP-based figures, which should be comparable across countries.

But how are estimates derived in practice by the PWT? The PWT has two distinct approaches and steps for calculating the estimates of domestic absorption and its constituents and GDP. The first step (with a distinct methodology) relates to estimates for the benchmark year. And the second step (involving a different methodology) relates to estimates for growth and to the level estimates for all nonbenchmark years.

[6] Quotation is from http://unstats.un.org/unsd/methods/icp/ipco_htm.htm.
[7] Ibid.

We describe both steps of the PWT methodology in detail in technical appendix 1. In this paper we focus on the second step, which involves estimation of GDP growth and hence the level of GDP in nonbenchmark years.

Once the PWT calculates the level of PPP-adjusted GDP (and the associated domestic absorption DA, which is the sum of consumption, investment and government expenditures) for the benchmark year (say 1996), it calculates the levels for nonbenchmark years according to the following equations. GDP for a nonbenchmark year, say 1995, is calculated as:

$$Y_{95} = DA_{95} + NFB_{95} \tag{1}$$

where Y is PPP-adjusted GDP, *DA* is domestic absorption and *NFB* is net foreign balance (which is the difference between exports and imports of goods and nonfactor services). By definition,

$$DA_{95} = DA_{96} / \left(1 + \widehat{DA}_{95,96}\right) \tag{2}$$

where the hat sign over a variable denotes the growth rate.

Now DA_{96} is domestic absorption for the benchmark year which is, by definition, in PPP terms. Consider next how $\widehat{DA}_{95,96}$ is computed:

$$\widehat{DA}_{95,96} = a_{95}^{C}\hat{C}_{95,96} + a_{95}^{I}\hat{I}_{95,96} + a_{95}^{G}\hat{G}_{95,96}\ , \tag{3}$$

where a_t^C , a_t^I , a_t^G are the time varying shares of consumption, investment, and government spending, respectively, in domestic absorption.

The growth rate of domestic absorption is calculated as the weighted average growth rate of its three components, C, I, and G. The weights assigned to each of these components are the shares of each component in domestic absorption measured in 1995 and measured at international prices. These shares are obtained from step 1 described in technical appendix 1. It should be stressed, however, that the growth rates of real C, I, and G, are from the national income accounts (constant price series), so that they are by definition computed at domestic prices and *not* at international prices, a key issue that we will explore subsequently.

In sum, the PWT calculates growth rates of DA and GDP using equation 3 above and uses these growth rates to derive estimates for the level of GDP in nonbenchmark years. One important corollary of this procedure is worth highlighting. Take versions 6.1 and 5.6 of the PWT, which had 1996 and 1985, respectively, as the benchmark years. In PWT 6.1 the estimate for the level of per capita GDP for 1985 is derived from the level estimate for 1996 (the benchmark year for PWT 6.1) to which the growth rate, as calculated in equation 3, is applied. The information on PPP-adjusted GDP per capita in PWT 5.6, based on the disaggregated data compiled in the ICP for 1985, is almost completely ignored.[8] This occurs for every new version so that potentially valuable information from previous ICPs is not used. Of course this is at the heart of the problem of reconciling benchmark studies with national accounts data, explored in early work by Summers and Heston (e.g. Summers and Heston, 1984), in what was described in the 1980s as the 'consistentization' problem and more recent studies by Eurostat and the OECD on reconciling their annual benchmarks and national accounts growth rates.[9]

3. Data Variability: Patterns and Associations

We focus on the variability of estimates between PWT version 6.1 and 6.2. These versions differ in some relatively small ways.[10] Both versions are within the same generation of the Table, and therefore have almost identical methodologies and, more importantly, between

[8] This information is used to a small extent as described on page 9 of the Data Appendix http://pwt.econ.upenn.edu/Documentation/append61.pdf.

[9] Although consistentization did reconcile national growth rates with implied growth rates of multiple benchmarks by using information of past benchmarks and current benchmarks, it proved difficult to implement. For details see the unpublished appendix to PWT 5.

[10] There are new PPPs for the OECD countries, 20 additional countries (mainly transition economies), coverage was extended from 2000 to 2004, and the reference year for calculation of the purchasing power parities has been moved from 1996 to 2000.

these versions there is no new ICP and hence no new international price data.[11] Despite that the changes are small, estimates of GDP levels and growth vary substantially between PWT 6.1 and 6.2. Of course, one substantive difference is that PWT 6.2 uses more "updated" national income accounts data than PWT 6.1, but in practice this updating does not lead to large revisions.[12]

The top panel of figure 1 illustrates that the 29-year average annual growth rates differ between PWT 6.2 and 6.1.[13] The average difference in the growth rate (for the period 1970–1999) generated by the two versions is close to zero (0.1 percent). But the standard deviation of the differences in growth is about 1.1 percent. This is quite substantial when compared with the fact that the average growth rate (in 6.2) is 1.56 percent. Put differently, growth is more than 1 percent per annum different in more than half the countries and more than 2 percent different nearly a quarter of all cases. The bottom panel of figure 1 depicts the same comparison between PWT 7 and 6.1 and the variability is broadly similar to that between 6.1 and 6.2.

Since the PPPs are critical to the PWT methodology, we show in figure 2 the variation in the estimates of the level of the PPPs (averaged over 29 years). There is a great deal of variation across versions in the calculated change in prices, measured in PPP terms.

One basic aspect of data variability between PWT versions—namely that it increases when the data are at higher frequency—becomes evident when we compute the growth

[11] In addition, an important point emerges from a careful examination of the methodological history of the tables (e.g., Deaton and Heston 2010). We usually think that a revised data series is better than the original series due to various kinds of corrections. But in the case of the tables, the methodology has not necessarily improved over time. There have been innovations, but some of these happened for bureaucratic or even political reasons as international organizations became involved in the data collection, preparation, and presentation. As a result, it is not reasonable to assume that one version of the tables (e.g., the latest) is necessarily better than other versions. This is another reason why results that hold in one version of the tables but not other versions should probably be viewed with greater skepticism.

[12] The correlation between the 29-year annual average growth rate in the national income accounts data used in the two versions is 0.99.

[13] This 29-year period is the longest for which the sample size can be maximized. Unless otherwise specified, all the growth rates are for the RGDPCH series in the PWT. This is a chained series and is the one endorsed by the authors. In technical appendix 1, we explain briefly the difference between this series and the RGDPL series. Appendix table 1 in *Supplementary Materials* available online lists the countries in the sample.

rate over 1-, 10-, and 29-year periods. These are presented in the three panels of figure 3 (where the third panel is the same as the top panel of figure 1) with the same scale for the growth rates. Figure 3 dramatically illustrates that growth computed on an annual horizon is considerably more variable across versions than growth computed over 10- and 29-year horizons. For example, the standard deviation of the growth rates across the two versions is 5.39 percent (relative to the average growth rate of 1.86 percent over that horizon) compared with a standard deviation of 1.08 percent (relative to average growth of 1.51 percent) for the 29-year horizon. The striking differences in this figure intuitively explain the results we find in the next section about the robustness of leading growth studies. Results based on annual data prove to be less robust across versions of the PWT than are results based on 10-year or 29-year averages. Evidently, the errors we are seeing get averaged out over longer horizons.

A. Associations of variability in GDP estimates across versions

Next, we examine systematically—based on simple regressions—the associations of data variability in the PWT. Our aim is not to establish causality but to understand the possible influences on data variability. In this spirit, we run regressions for three different and important variables estimated by the PWT—the PPP prices, the level of PPP GDP, and the growth rate of PPP GDP.

Why should the PWT estimates vary across revisions? Four possible factors may be at work. First, and most obviously, PWT estimates could change because the underlying National Income Accounts (NIA) data—which are key inputs for the PWT system—change.

Second, even apart from changes in the underlying data, revisions could be systematically related to the quality of the data. The PWT has always been quite transparent about this data quality issue, and it prominently assigned "quality grades" for each country (e.g., to version 5.0). These grades are subjective assessments made by the authors of the PWT, based on a number of factors described in the technical appendix to PWT 6.1, pages 13–18. Strikingly, only 32 countries received a grade of A or B. Grades of C or D were received by 147 countries. Plotting the data suggests that data quality might matter for revisions. The left-hand panel in figure 4 shows differences in 29-year annual average growth rates (1970–1999) for countries with data quality grades of A or B. The right-hand panel shows the same for countries with grades of C or D. All the major variation across versions of the Table occurs in the countries with lower grades. The same comparison also holds across higher frequency growth data.

Two other factors—size of a country and time—are potentially important and suggested by the PWT methodology for calculating the PPPs. The importance of size in affecting data variability was formally highlighted by Rao and Selvanathan (1992). They show that the PPPs and international prices can be seen as weighted averages, which makes it possible to interpret them as estimators of parameters from appropriately specified regression models. It then also becomes possible to assess the reliability of the estimates of these parameters. They show that the standard error of the estimates of the PPPs are inversely related to a country's size (more specifically, its total consumption expenditure).[14] A proposition that we will test in the data is: the smaller the country, the less reliable is the estimate of its PPP.

The final factor is time.[15] The GDP estimates for the nonbenchmark years in the PWT are based on *extrapolated* PPPs. The methodology works as follows. Take 1995, the first nonbenchmark year in PWT 6.1. The calculations of the international prices for 1995 are estimates based on the actual price data for 1996. So, some error is added to the 1995 estimates. If we take this back one more year to 1994, we know that 1994 international prices are in turn estimated from 1995 numbers. This adds one more layer of error. Relative to the benchmark year 1996, there are two sources of error for the 1994 estimates, and so on. This error structure going back in time is analogous to the error structure going forward

[14] See equation 5 in Rao and Selvanathan (1992).

[15] We should note that it is not the case that revisions to the Table change only when using more recent data. Appendix figure 3 (available online in the *Supplementary Materials*) shows that there are large revisions for the 1970s, the 1980s, and the 1990s looking at 10-year average annual growth rates.

imposed by many forecasting exercises. It might therefore be expected that estimates are more variable farther away from the benchmark year—a proposition we test in the data.[16]

B. Regressions results

Regressions for revisions to the PPPs, level of GDP, and GDP growth are reported in tables 3–5 respectively. The first variable we need to measure is the change in the underlying national income accounts data across versions. But there is a complication here that merits explanation. In the PWT, the underlying national income accounts growth data change for two reasons. First, national authorities can, and occasionally do, revise data. To capture the revisions done by the national authorities, we take GDP growth data from the World Development Indicators (which reflect data submitted by authorities) for the years corresponding to PWT 6.1 and 6.2, which are 2002 and 2005, respectively. We use the absolute value of the differences in the growth of GDP (at constant domestic prices) between these two datasets as our measure of changes in national income accounts data due to national authorities' revisions. But there is a second source of change to national income accounts data emanating from the PWT procedures. That is, the PWT further revises the data submitted by the country authorities before using it for computing PPP-based estimates described in technical appendix 1.[17] We use the log of the absolute value of differences in the growth of GDP (at constant domestic prices) between these two PWT-generated datasets as our measure of changes in national income accounts data due to the PWT revisions.[18]

The measure we use for data quality is the grading by PWT itself. We convert the four letter grades for data quality into an index, from one to four, where four is the best ("A"). We also included a dummy for countries that participated in the 1996 ICP study as an

[16] Rao and Selvanathan (1992) show that that the standard error of the Genebralized Least Squares (GLS) estimator for the PPPs is not only related to the size of a country (noted above) but also to the variance of the error term. Since, the PPPs are extrapolated sequentially, random error is added at each stage, so that the variance increases as we move away from the benchmark year.

[17] This layer of revision by the PWT appears to involve merging and splicing national income accounts data from multiple sources to derive a long series and appears to be done for low-income countries.

[18] In tables 3-5, we call the first source of change, "NIA-GDP growth" and the second source of change "PWT-NIA GDP growth."

additional measure of data quality. For size, we use the log of a country's PPP-adjusted total GDP obtained from PWT6.2. Since the benchmark year for PWT 6.1 is 1996, our measure of time is the number of decades from 1996 (with each year being 1/10 of a decade, the datapoint for 1995 will therefore be 0.1). A dummy for China is included as well because of the substantial, China-specific revisions between PWT 6.1 and PWT 6.2 (see Heston, 2001).

The basic regression we run is:

$$\begin{aligned} dif(x_{it}^{PPP}) &= \alpha_0 + \alpha_1 dif(g_{it}^{NIA}) + \alpha_2 dif(g_{it}^{PWT}) + \beta_1 grade_i^{PWT} + \\ &\quad \beta_2 GDP_{it} + \beta_3 dist_{it} + \beta_4 \left(dist_{it} * grade_i^{PWT}\right) + \varepsilon_{it}, \qquad (4) \end{aligned}$$

where $dif(.)$ refers to the absolute value of the difference in the estimate between PWT 6.2 and 6.1 of the relevant variable within brackets; i and t subscripts refer to country and time, respectively. x is either the log of PPP prices (in table 3), the log per capita PPP-adjusted GDP (in table 4), or the growth rate of per capita PPP-adjusted GDP (in table 5). g^{NIA} is the growth rate of per capita GDP in constant domestic (i.e., not PPP-adjusted) prices as reported by the national income authorities; g^{PWT} is the growth rate of per capita GDP in constant domestic (i.e., not PPP-adjusted) prices as revised by the PWT; $grade^{PWT}$ is the measure of country data quality as evaluated by the PWT. GDP is PPP-adjusted total GDP and is obtained from PWT6.2. $dist$ is calculated as the absolute difference between the year of the observation and 1996 (i.e., $abs(t - 1996)$), and is divided by 10 to express the variable in decades. $dist * grade$ is the interaction between the distance variable and the quality rating of the PWT, and ε is a random error term.

In table 3, for example, the dependent variable is the absolute value of the difference in log of international price between version 6.2 and version 6.1.[19] Columns 1 and 2 show that

[19] All the regressions are done at annual frequency but we have also replicated them at decadal and 25-year frequencies and obtained similar results with regards to coefficient signs and significance levels. Consistent with figure 3, when using higher frequency data, the coefficient magnitudes are considerably larger for all

the two sources of changes in the national income accounts data affect revisions to the PPP series. In column 3, it can be seen that a higher grade is negatively related to differences in the data, even when controlling for 1996 ICP participation, which confirms that the PWT team warnings on data quality were appropriate. Log GDP (column 4) has a similar effect, although this may be a proxy for data quality or the existence of a benchmark study.

Perhaps more novel is the result in column 5. This shows that the further a data point is from the benchmark year, measured in absolute value of years from 1996, the more likely its data is to change between versions 6.1 and 6.2, i.e., the coefficient is positively signed. These results on the variability of the PPPs and their relationship to economy size (log total GDP) and distance are graphically represented in figures 2 and 5, respectively. While individually these variables explain the revisions to the PWT estimates, we want to know which matters most when all potential determinants are included, which we do in column 6. All the variables above retain their significance except for the PWT data quality variable and the variable capturing the revisions to the national income data created by the PWT methodology. Even though the PWT grading seems not to matter, other measures of data quality such as size and whether a country participated in the ICP study do matter. The variable measuring the distance from the benchmark year (as well as its interaction with the data quality variable) continues to remain significance. Figure 5 strikingly depicts the relationship between variability and distance from benchmark year, particularly for countries with grades C and D. Data variability increases as the data point moves away from 1996 both forward and backward in time.[20]

Table 4, which reports results for the level of GDP per capita, suggests that the determinants of the revisions for GDP are similar to those for international prices.

Table 5, which explains revisions to the growth of GDP per capita, shows similar results

regressors.

[20] We re-estimated the specifications in table 3 by substituting data from PWT 7 for PWT 6.2. While the results were broadly similar, the interpretation of the distance from benchmark year is more complicated because the benchmark years are different in versions 6 and 7.

with one exception. It turns out the revisions to the PPP-adjusted growth rates computed by the PWT are largely determined by the revisions to the underlying national income accounts data. Specifically, the revisions matter mostly due to revisions by the PWT procedures illustrated in column 6, where the variable "PWT-NIA GDP growth" is large in magnitude and significance level. Other aspects of the PWT methodology such as data quality and country size remain significant determinants of PPP-based growth revisions.[21]

4. Replication Exercises for the Cross-Country Growth Literature

Thus far, we have shown that PWT data vary across revisions and do so systematically. But do they matter? To assess this we turn now to examine some of the most prominent studies in the growth literature that have used PWT data.

Most papers in the empirical growth literature use as a dependent variable either the level of real per capita GDP in PPP terms or the growth rate in the same. Some papers use the level of income or growth rates as a right-hand side control variable. Given the considerable variation in the data on growth rates (and levels) and the fact that there is no one "best" set of PWT to use, a natural question is: Which results in the literature are robust to changing versions of the Table?

A. Criteria for considering papers

We have examined many of the leading papers in the growth literature based on PWT 5.6 or 6.1.[22] In each case, we attempted to run exactly the same specifications and samples, but using version 6.2 of the Table instead. This approach cannot prove that a particular set of results is right or wrong, but it may illustrate patterns in terms of what kind of results

[21] We have also reproduced tables 3-5 for the subsamples of countries with data quality A&B and C&D, respectively. As expected, most of the variation in the data comes from the C&D subsample. Notably, the *dist* coefficient estimate is large, positive and highly significant in the C&D variable, while close to zero and insignificant in the A&B countries. The complete set of regressions results is available upon request.

[22] We focused our attention on "high impact" papers, measured either in terms of citations (using Google Scholar) or based on discussion with active researchers or papers that we think will prove influential. We sought to examine a range of papers, in terms of the frequency of data and methodology. However, our sample is not comprehensive, and the results are intended as illustrations only.

are more or less robust.

We had four criteria for inclusion. First, the list had to include papers that studied the level of GDP, the growth rate of GDP, and the volatility of GDP. Second, since data revisions can be thought of as measurement error, and since estimation depends upon whether measurement error is on the left- or right-hand side, we included papers in which the various GDP measures (level, growth, and volatility) featured as regressors and as regressands. Third, since our preliminary analysis of the data showed that measurement error varies significantly between high and low frequency data, we chose papers that adopted a pure cross-sectional approach (very low frequency); a panel approach, with data averaged over four to five years (high frequency); and a panel approach with annual data (very high frequency). Table 6 lists the various papers in the appropriate analytical category. After having decided the universe of potential papers based on these criteria, we narrowed the list to papers that we considered influential. Finally, for inclusion here, we had to be able to obtain the original data and to be able to replicate the paper's main results with those data.

In the replication exercise, we first replicate the authors' core result based on the original data. Next we show the same core results using the original data but changing the sample to match what is available in PWT 6.2. We then report the core result using PWT 6.2 data (holding the sample constant). Finally, we reproduce results for a more truncated sample that excludes countries with a data quality rating of D.

B. Papers with Table invariant results

In all, we tested the robustness of 13 papers in the growth literature. Note that we did not check all specifications in all papers. Rather we concentrated on what appeared to us—or to others citing the work—as the "main" results. The lower part of table 6 lists nine papers for which we found basically no or small changes in results. In addition, there were more substantial changes for four papers: Ramey and Ramey (1995), Jones and Olken (2005), Hausmann, Pritchett, and Rodrik (2005), and Aghion, Howitt, and Mayer-Foulkes

(2005). We discuss these results below, and report the detailed replication estimates in *Supplementary Materials* available online.

Before that, we should emphasize that many prominent papers were unaffected by the robustness checks. Whatever else is right or wrong with the growth literature, the bulk of it is not afflicted by the problem of sensitivity to changes in PWT GDP data. This list includes work such as Acemoglu et al. (2003), Barro (1999), Burnside and Dollar (2000), DeLong and Summers (1991), Demirguc-Kunt, Laeven, and Levine (2004), Easterly et al. (1993), and Sachs and Warner (1995). The minor changes that we found in results from these papers are mentioned briefly in table 6. Miguel, Satyanath, and Sergenti (2004) proved robust in a particularly interesting way, which we expand on below. And we also discuss Mankiw, Romer, and Weil (1992) in more detail below because their results come close to being affected despite involving level GDP regressions that proved very robust for other papers.

C. Papers with changing results

Ramey and Ramey (1995)

This paper, published in the *American Economic Review* in 1995, tests the link between growth and volatility. The specification involves running an annual panel with growth rate of per capita GDP on the left-hand side, a set of country- and time-varying controls on the right-hand side, and a time fixed effect.[23] There is no country fixed effect; instead there is a country-specific and time-invariant measure of volatility of growth, proxied by the standard deviation of the country-specific residuals over the period covered by the growth data. Residuals and the volatility measure are simultaneously estimated using maximum likelihood.

The paper used growth data from the PWT 5.6. In the original paper, there are 2,208 observations from 92 countries, while in the balanced sample (i.e., where there are data from

[23]The equation estimated is: $\Delta y_{it} = \lambda\sigma_i + \theta X_{it} + \varepsilon_{it}$, where y is the log level of per capita GDP, σ_i is the standard deviation of the residuals of a country across time, and X is a vector of controls, including initial investment share of GDP, initial population growth rate, initial human capital, initial per capita GDP, lagged GDP, and several time trend and dummy variables.

both PWT 5.6 and 6.2) there are 1,776 observations from 74 countries.

In the paper, the coefficient on volatility is negative and significant at the 1 percent level. The magnitude of the coefficient is -0.177. When we re-estimated their core specification on the balanced sample, the coefficient on the volatility term is smaller and barely significant. When we switch to using PWT 6.2, the volatility coefficient becomes even smaller and quite far from being significant. The same is true if we use the original dataset but drop countries with a data quality grade of D or if we drop the same countries while using version 6.2. This key result, namely the statistical insignificance of the volatility term, holds when we replace the original data with that from PWT 7.[24]

Jones and Olken (2005)

This paper, published in the *Quarterly Journal of Economics* in 2005, relates growth to leadership. Specifically, it estimates the effect of random leader deaths on a country's growth rate.[25] The two key findings are: such deaths have a significant impact on growth on average; and second, the death of leaders has a significant impact in autocracies but not in democracies. This paper used annual growth data from PWT 6.1 and we successfully replicated its baseline results.

When we re-estimated the core specification using data from PWT 6.2, we obtained the following differences compared with the original: First, in the original paper the coefficients of the random leader death are significant for the year of the leader death and for the two subsequent years. When we re-estimated it, using new data, the contemporaneous effect remained significant, but the effect for the two subsequent years ceased to be significant (p-value of 0.12 and 0.13, respectively). Second, the striking differences are related to the disaggregation of the results by type of political regime. In Jones and Olken (2005), random

[24] Dawson et al. (2001), and Katayama and Ponomareva (2010) also show that variability in growth data across different versions of PWT renders the volatility-growth relationship unstable.

[25] The equation estimated is: $g_{it} = \alpha_z PRE_z + \beta_z POST_z + \nu_i + \nu_t + \varepsilon_{it}$, where g is growth of per capita GDP and ν_i and ν_t are country fixed effects and time effects, respectively. For each leader death at t, there are location-specific time dummy variables equal to one in one of $t-5, t-4, \ldots, t-1, t+1, t+2, \ldots, t+5$ and equal to zero otherwise. These vectors of dummy variables are denoted PRE and $POST$.

leader deaths had a significant impact in autocracies in the year of the leader death and in the two subsequent years. In the robustness check in the same panel, they were not significant for any of these three time horizons. Even more interestingly, Jones and Olken find no significant effects in democracies. In the re-estimation using version 6.2 of the Table, leader deaths are significant in the year of the leader death and the following year. Thus, not only the average effect but the pattern across political regimes seems to vary depending on whether data from PWT 6.1 or 6.2 are used.

The same pattern held when we dropped from the sample those countries categorized as very poor quality (grade D) from a data point of view, i.e., their original results hold (although the coefficients are smaller) with PWT 6.1 but the reversal of results is still the case with version 6.2.[26]

When we substituted data from PWT 7 instead of the original date (6.2) used by the authors, there was a further twist. The effect of leader deaths on growth was found to be significant in autocracies and democracies. In effect, three versions gave three different results: 6.1 yielded significance for leader deaths only in autocracies; 6.2 for leader deaths only in democracies; 7 for leader deaths in both regimes.

Hausmann, Pritchett, and Rodrik (2005)

This paper, published in the *Journal of Economic Growth* in 2005, identifies a set of countries that are deemed to have sustained growth over a long period of time. The paper used data from PWT 6.1. The criteria used to define sustained growth are that countries must have experienced: an *improvement* in growth rates of at least 2 percentage points per capita (this captures the idea of acceleration); *sustained* growth of at least 3.5 percent per capita for seven years; and a higher post-acceleration income level than the pre-acceleration peak (this is to ensure that accelerations are not simply a rebound from a prior period of

[26] In private communication, Jones and Olken showed that their results continue to hold when nonparametric methods are used to estimate their basic specification, or when the original sample is maintained by filling in missing values for countries for which PWT 6.2 did not report estimates. We focused on their core linear estimation specification.

very bad performance, for example, due to wars or conflict or other shocks). In addition, *growth per capita* must remain *above 3 percent* after seven years, which captures the sense that good performance is sustained.

On this basis, Hausmann, Pritchett, and Rodrik (2005) identified 82 country-periods (see their table 3) that met these criteria.[27] Our exercise was simply to see how the list of countries changes when the data from PWT 6.2 are used. Between PWT 6.1 and 6.2, 100 country-periods were identified as sustained growers under this definition. However, only 65 cases were common to both data sets. In other words, there is a discrepancy of 35 percent between the two versions. There were 17 cases that Hausmann, Pritchett, and Rodrik (2005), using PWT 6.1 (for example, India in 1982 and Sri Lanka in 1979), identified as sustained growers but that did not show up as such using PWT 6.2. Conversely, there were 18 instances that were identified by PWT 6.2 as sustained growers but that were not so categorized by Hausmann, Pritchett, and Rodrik (2005) using PWT 6.1. One way of illustrating this discrepancy is to note that only in about 65 percent of the cases was there agreement between the two datasets; in other words, in about 4 out of 10 instances, the answer depends on which data set is used.

When we substituted PWT 7 data for the data used in the paper (from PWT 6.1), the extent of discrepancies increased. In 44 out of 110 cases (or 40 percent of the time), there is a discrepancy on the identity of fast-growers.

Miguel, Satyanath, and Sergenti (2004)

This paper, published in the *Journal of Political Economy* in 2004, attempts to identify and quantify the impact of economic factors on civil wars. The estimation is restricted to 41 countries in Africa and uses an annual panel framework using GDP data from PWT 6.1.

The key results of the paper are that in an ordinary least squares (OLS) or probit framework, contemporaneous and one-period lagged growth does not have a statistically signif-

[27] They actually identified 83 growth transitions but one of them has been excluded because data for country-regionplaceBotswana does not go farther back than 1970 in PWT 6.2.

icant effect on civil conflict in Africa.[28] However, when economic growth is instrumented with rainfall, one-period lagged growth has a statistically and economically significant effect on conflict—see column 1 (OLS) and column 4 in their table 9 (IV) from the original.

When we re-ran these regressions using data from PWT 6.2, we found that the IV results of Miguel, Satyanath, and Sergenti (2004) remained unchanged. However, the OLS results changed. Specifically in the OLS framework, contemporaneous economic growth, which did not have a significant effect in the OLS, becomes borderline statistically significant (the coefficient on contemporaneous growth has a t-statistic of 1.67).

Our explanation of these results is that the OLS and probit estimations are more prone to measurement error (in this case on the right-hand side) which is substantial in the annual panel framework, especially for Africa. This is cause for substantial differences in results. On the other hand, a good instrument addresses measurement error, resulting in a more robust estimation.

Interestingly, when we used PWT 7 data, the core result on the importance of growth for conflict was established even without the need for instrumentation. In the original paper, this core result does not hold without instrumentation.

Aghion, Howitt, and Mayer-Foulkes (2005)

This paper, published in the *Quarterly Journal of Economics* in 2005, examines the effect of financial development on income (or technological) convergence. The paper's two core findings are that income convergence will depend on the level of financial development; and that there will be a threshold level of financial development above which countries will converge and below which they will diverge.

Aghion, Howitt, and Mayer-Foulkes (2004) run a simple cross-country growth regression for a sample of 63 countries, where the key innovation is an interaction term between the

[28] The estimated equation is: $conflict_{it} = \alpha_i + X_{it}'\beta + \gamma_0 growth_{it} + \gamma_1 growth_{it-1} + \delta_i year_t + \varepsilon_{it}$, where X represents a vector of controls, and $growth_{it}$ is instrumented by $rainfall_{it}$ and $rainfalli_{t-1}$ in the 2SLS specifications.

level of financial development and the initial level of income divergence with the frontier country, the United States.[29] One important prediction of the model for which the estimation provides confirmation is that the coefficient on this interaction is negative. When we re-run the Aghion, Howitt, and Mayer-Foulkes (2004) core regression using PWT 6.2 data, we find that while the magnitude of this coefficient drops by about 40 percent, it remains statistically significant.

Although the primary prediction of the model is robust to changing data, the second implication of the Aghion, Howitt, and Mayer-Foulkes (2004) paper is not robust to changing the GDP data. Aghion, Howitt, and Mayer-Foulkes (2004) derive a threshold value of financial development, which is the ratio of the coefficient on the income convergence term and the coefficient on the interaction between income convergence and financial development. In the Aghion, Howitt, and Mayer-Foulkes (2004) specifications this value is about 25 percent.

In the revised estimations, using PWT 6.2 data, the coefficient on the income convergence term is economically close to zero and statistically insignificant from zero (in one of the core specifications, the coefficient on this term switches signs relative to that in the paper). This yields a critical value for financial development of zero, suggesting that all countries will converge, which runs counter to the spirit of the main results implied by Aghion, Howitt, and Mayer-Foulkes (2004).

When we used PWT 7 in the replication exercise, the results in the original paper were restored, suggesting a difference between PWT 7 and PWT 6.2.

Mankiw, Romer, and Weil (1992)

This paper, published in the *Quarterly Journal of Economics* in 1992, tests an augmented version of the Solow growth model, with the augmentation consisting of adding human capital as an additional input (apart from capital and labor) into production. The test comprises

[29] The equation estimated is: $g_i - g_1 = \beta_0 + \beta_f F_i + \beta_y (y_i - y_1) + \beta_{fy} F_i (y_i - y_1) + \varepsilon_i$, where $y_i - y_1$ is the gap in output between country i and the country at the technology frontier, $g_i - g_1$ is the gap in per capita GDP growth, and F is the level of financial development.

two elements: first, checking for the significance of the coefficients on savings (proxied by investment), human capital, and population growth (with the latter augmented to reflect technical progress and adjusted for depreciation); and second, importantly, checking for the magnitudes of these coefficients. In particular, the Solow model yields the result that the sum of the coefficients on savings, population, and human capital should add to zero; or, in practice, the coefficient on the savings term adjusted for population growth should be equal in sign and magnitude to the coefficient on the human capital term also adjusted for population growth.

Mankiw, Romer, and Weil (1992) estimate a simple cross-section regression in which the left-hand side is the level of log per capita PPP GDP.[30] They estimate it for three samples, one consisting of 98 countries for which data are available, another for a sample of 75 countries in which countries with PWT data quality grade D are dropped, and a third where non-oil countries are dropped from the sample.

When we re-estimate their equations for the Solow model using PWT 6.2 data (and the non-oil sample), we find that the coefficients on savings, population growth, and human capital are all correctly signed and statistically significant. However, we find that the key test of the Solow model—that the coefficients on savings and human capital (both adjusted for population) should be equal in magnitude—comes close to being rejected by the data (the p-value for the test of the Solow restriction is 0.166). This result is interesting because it contrasts with the general finding that long-horizon cross-sectional growth studies as well as those that use the level of GDP do not see significant changes in the results. In this study, the results change presumably because the regressions use the investment-GDP ratio, which is particularly prone to variability.

The biggest difference between using PWT 6.2 in the replication exercise and using

[30] The equation estimated is: $\ln[Y/L] = \pi_0 + \pi_1[\ln(I/GDP) - \ln(n+g+\delta)] + \pi_2[\ln(school) - \ln(n+g+\delta)] + \varepsilon$, where Y/L is GDP per worker, I/GDP is the investment share of GDP, n is the population growth rate, g is the technology growth rate, δ is the rate of capital depreciation, and *school* is average number of working age adults in school.

PWT 7 arises in this paper. Replacing data in the paper by data from PWT 7, results in the coefficient on investment share dropping close to zero. And the key Solow restriction is strongly rejected. The reason is that the investment series is much more variable in PWT 7 for reason that are still unclear.

D. Discussion

Is there an explanation for the pattern observed? The least robust papers are those that use annual panel data, for which measurement error is large. Conversely, all the papers that survive the robustness are those that use low frequency data (i.e., cross-sectional estimations) or five-year panel data. Evidently, averaging reduces "data revision error" and makes the estimations more robust.

But this raises another question. As long as this "data revision error"—which is in principle a particular kind of measurement error—is random, why isn't it captured by the random error term in the growth regression? The answer is twofold: First, when the PPP-adjusted GDP appears as the independent variable (in level or growth) in the regression estimation, revision errors to it are not random and suffer from the biases imposed by the PWT methodology as discussed in section 3. For example, whether the countries used in the estimation sample are small or large or whether the data used extends to the early 1960s biases the PPP-adjusted GDP measures and therefore cannot be appropriately captured by the random error term. Second, a typical growth regression usually incorporates regressors that are either directly related to PPP-adjusted GDP (i.e., initial GDP), or components thereof (such as consumption, investment, or government spending—directly measured by PWT as well). This introduces measurement error in the dependent variable, which although challenging could potentially be corrected econometrically. That an error-in-variable is a probable cause is supported by the fact that the 2SLS results of Miguel, Satyanath, Sergenti (2003) survive the robustness checks—despite using annual panel data. In Miguel, Satyanath, Sergenti (2003), the core result is not robust to data revision in the OLS version but it is robust in the IV version. Having a good instrument evidently helps overcome measurement error.

5. Valuation: Are PWT GDP Estimates Really PPP-Based?

This is an odd question to ask. The raison d'être of the PWT is to come up with PPP prices and use them for computing GDP estimates. For the benchmark year, PPPs are calculated from the price data that are collected and these PPPs are used to compute GDP. So, in the benchmark year, the GDP estimates are at international prices. But matters turn out to be much murkier for years other than the benchmark year.

Recall equation 3:

$$\widehat{DA}_{95,96} = a_{95}^{C}\hat{C}_{95,96} + a_{95}^{I}\hat{I}_{95,96} + a_{95}^{G}\hat{G}_{95,96} \ .$$

In this equation, the shares of the different components in DA are at international prices and computed by the PWT. However, the growth rates of the components (C, I, and G) are obtained from the national income accounts. As such, these growth rates are at domestic, not international, prices.

It is possible to further decompose equation 3 to see how the additions to domestic absorption are being valued. In technical appendix 2, we illustrate with a simple example that when there is more than one good, equation 3 results in valuing the quantity additions to each of the components of domestic absorption at some hybrid of domestic and international prices (see equation B5 in technical appendix 2).

The difference between the actual growth measurement and the "ideal" measurement where all goods are valued at international prices depends on the difference between domestic prices of a good and its international price (equation 6 in technical appendix 2). We know that this difference in prices will vary systematically across countries. It will be greater for smaller countries because under the Geary-Khamis aggregation procedure, domestic prices of larger countries have a greater weight when computing international prices. This is the Gerschenkron effect. Thus, GDP growth rates are likely to be measured with greater error (relative to the ideal growth rate) for smaller countries. A second problem is that the farther away the measurement period is from the base year, the greater the discrepancy. Hence growth rate calculations for years farther away from the benchmark year are likely to have greater measurement error.

More generally, the fact that growth rates of GDP are not at international prices, implies that level GDP estimates are also not fully at international prices for nonbenchmark years. Conceptually, the farther away the estimate is from the benchmark year, the less it is at international prices.

Given the documented problems with generating a reasonable PPP-adjusted growth GDP series, one possibility is for researchers to use the PWT estimates for the levels and national income estimates for GDP growth rates. In this view, the way forward is to use PPP prices for valuing the level of GDP and to use domestic prices for valuing growth rates.[31]

However, there are two problems with this approach. First, if the PWT allows cross-sectional comparisons at two different points in time based on PPP prices, consistency requires that the change in these levels across time (the growth rates) should also be at PPP prices. Current estimates in the PWT do not allow for such consistent intertemporal comparisons because of the valuation problems we identified.

Second, there is an even deeper problem in the current methodology. Level GDP estimates in the nonbenchmark years are derived from growth rates that are not at PPP prices. Therefore, level GDP estimates in such years are themselves not at PPP prices. So, under current PWT methodology, even pure cross-sectional comparisons are not entirely valid, undermining the basic rationale of the PWT approach. Note that since GDP levels for nonbenchmark years are calculated by applying growth rates to benchmark year level GDP estimates, the farther away we move from the benchmark year the more the level GDP estimates will be composed of quantities that are not valued at international prices.

6. Thoughts on an Alternative PPP-GDP Chained Growth Series

The existing PWT PPP-GDP series suffer from three shortcomings. First, as documented previously, the methodology leads to large and systematically biased variations across versions of the Table. Second, estimates of PPP-adjusted GDP growth rates and PPP-adjusted level of GDP for the nonbenchmark years are not at international but rather some "hybrid" prices, which goes against the raison d'être of PWT. Third, each new generation of PWT leads to discarding useful information from the disaggregated price data in all previous ICP benchmark studies.

Is there a way of calculating GDP estimates that overcome these problems? In principle,

[31] For example the growth rates in the World Bank's WDI are all at domestic prices.

yes.[32] First and foremost, the alternative approach could use the data and the estimates of the level of GDP compiled in the benchmark years (i.e., 1980, 1985, 1996, and 2005).

The current methodology starts with the level of GDP in the benchmark year, then calculates the growth rates, which are in turn used to calculate the level of GDP in the benchmark years. We would propose doing it the other way around: to use the level estimates for the benchmark years and calculate the growth rates from these level estimates. Take the PPP-adjusted GDPs calculated for 1985 and 1996 from disaggregated ICP data. While cross-sectional comparisons can be made for each of the two years, there cannot be intertemporal comparisons based on these estimates because they are not at common prices. The aim, therefore, is to come up with GDP level estimates that are at common prices.

In what follows we offer some preliminary thoughts on one possible alternative method in producing estimates of PPP-adjusted GDP growth rates. We qualify that this is only an attempt to provide a possible road map rather than to proposed a complete alternative methodology, certainly beyond the scope of this paper.

Assume that the level of GDP for a country at international prices in the benchmark year 1985 is:

$$Y_{85} = c^1_{85}\pi^1_{85} + c^2_{85}\pi^2_{85} + i_{85}\pi^i_{85}, \tag{5}$$

where the πs are all international prices calculated from the Geary-Khamis aggregation, and superscripts refer to the type of consumption good (1 or 2) or to the investment good i (for expositional simplicity, we assume that GDP comprises only C and I, and that C comprises only c^1, c^2).

The level of PPP-adjusted GDP in the second benchmark year is:

$$Y_{96} = c^1_{96}\pi^1_{96} + c^2_{96}\pi^2_{96} + i_{96}\pi^i_{96}. \tag{6}$$

To calculate growth rates, we would ideally need to value either period 1985 quantities at 1996 prices or 1996 quantities at 1985 prices. Thus we can derive two new GDP estimates that are, respectively:

$$Y^2_{85} = c^1_{85}\pi^1_{96} + c^2_{85}\pi^2_{96} + i_{85}\pi^i_{96}, \tag{7}$$

[32] Feenstra, Ma, and Rao (2009) have proposed a theoretically sophisticated measure for intertemporal comparisons but it only applies to consumption and not to the other components of domestic absorption.

$$Y^1_{96} = c^1_{96}\pi^1_{85} + c^2_{96}\pi^2_{85} + i_{96}\pi^i_{85}. \tag{8}$$

The point is that the data for computing Y according to equations 7 and 8 generated by the PWT. For each benchmark year, the quantities of the different consumption and investment goods are available at the disaggregated level (recall that these are the inputs for the Geary-Khamis aggregation procedure, and these quantities are derived simply by deflating expenditures on each good at domestic prices by domestic prices). And for each benchmark year, the Geary-Khamis procedure yields disaggregated international prices.

Equations 5 and 8 will yield a GDP series that is intertemporally comparable (at period 1985 prices); equations 6 and 7 will also yield a GDP series that is internationally comparable but at period 1996 prices. It is now simple to calculate growth rates.

We can then derive the growth rates of GDP from 1985 to 1996 either at base-year prices, which is:

$$\hat{Y}_{85} = \frac{\Delta c^1\pi^1_{85} + \Delta c^2\pi^2_{85} + \Delta i\pi^i_{85}}{Y_{85}}. \tag{9}$$

Or at current year prices which is:

$$\hat{Y}_{96} = \frac{\Delta c^1\pi^1_{96} + \Delta c^2\pi^2_{96} + \Delta i\pi^i_{96}}{Y^2_{85}}. \tag{10}$$

Our preferred approach, however, would be to calculate a chained series that can be used for international comparisons. This would involve averaging the two growth rates emerging from equations 9–10 to yield a chained growth estimate between these two years. The way we would envisage is that the chaining would be based on successive ICPs. Thus, there would be a growth estimate between 1985 and 1996 based on the ICPs in 1985 and 1996. Similarly, there would be an estimate between 1996 and 2005 based on the ICPs in these two years. And so on. The advantage of this approach would be that every time a new ICP is conducted, it would not change historical growth rates of GDP. Only growth rates going forward until the next ICP would be estimated.

Of course, a number of practical complications will have to be addressed. First, the basket of disaggregated goods varies across benchmark years, so some way of deriving a common basket will have to be found (i.e., equations 5–8 will need to have a common basket of goods). In other words, we would need the detailed price data for the composites of C, I, and G aggregates between the two benchmark years; although such data are available for

OECD countries, they are not readily available for non-OECD countries.

A second problem with our proposal is that it would only provide a growth rate for the horizon between successive ICPs (either 10 years or over 5 years, if in the future ICPs are conducted every 5 years). There will not be a high frequency (i.e., annual) growth rate. But in assessing whether this is a major loss, two points should be kept in mind: annual data seem particularly problematic because of the current extrapolation methodology. Moreover, the motivation of the PWT is to make meaningful comparisons of standards of living across countries at a given point in time. Taking long averages seems to preserve this scope but high frequency data (especially annual frequency comparisons) become unusable.

A third problem relates to samples. For 1985 and 1996, but especially for previous benchmark years, the sample of countries for which ICPs have been done remains limited. How to derive the chained growth numbers for a large enough sample of countries going back before 1985 or 1980 needs further thought. This would not be a problem going forward because many more countries are now participating in the ICP studies.

Once we have a chained growth estimate of the type we have proposed, this is how we would envisage the use of different PWT GDP series. First, for pure cross-country comparisons of the level of per capita income, researchers should use the estimates from the different benchmark years. Thus, if a researcher wants to compare incomes across countries in 2005, the benchmark estimate for 2005 should be used. If the time period is 1996, the benchmark estimate from the ICP estimate of PWT 6.1 (not 6.2) should be used. And if the time period is 1985, the estimate from PWT 5.6 for 1985 should be used. The key message here is that researchers should use that version of the PWT that is closest to the timing of inquiry and should *not* use the most recent version. That is, if the year of inquiry is 1985, the most recent PWT series (PWT 6.1 or 6.2) should NOT be used because the level estimate for 1985 in PWT 6.1 and 6.2 is NOT at international prices. Of course, if the year of inquiry is say 1990, researchers could use either the PWT 6.1 estimate for 1996 or the PWT 5.6 estimate for 1985.

Second, if the object is intertemporal growth comparisons, especially growth over the medium and long run, researchers should use the new chained growth estimates that we have proposed.[33]

[33] Of course, these chained growth estimates could, in principle, lead to a level GDP series (at least for the benchmark years) that is at a common international price. Thus, based on a chained growth estimate based on 1985 and 1996 ICPs, we could have a chained level estimate for 1985 that applies the new growth estimate

Our third suggestion relates to the use of annual data. Given that researchers will continue to use annual growth data, should they use the PWT series or some alternative series such as that from the WDI? It must be clearly understood that annual growth series in both PWT and WDI are not at PPP prices. In the WDI, all growth (even the PPP series) is measured at domestic prices and not at PPP prices. In the PWT, annual growth is measured at hybrid (domestic and PPP) prices. But the deficiencies we have identified in the PWT estimates are not overcome by using WDI data. Given the sensitivity of the growth results to data revisions at annual frequency, we would emphasize that research using annual data should demonstrate robustness to alternative data series. That would be our preferred research strategy. Alternatively, if we had to choose one series for research based on annual data, we would probably favor GDP growth data (constant price series either in local currencies or dollars) from the World Development Indicators for two reasons: first, at annual frequency, PPP effects are less important so the costs of foregoing the use of PPP-adjusted data are smaller. Second, WDI growth rates are subject to only one layer of revision, namely by the country authorities, whereas, the PWT further revises these national growth rates before they become inputs for the PPP-based growth estimates. If the PWT growth estimates were truly at PPP prices the benefits may outweigh the costs of additional layers of revisions but that is not the case as we have shown. But the use of WDI data must come with the clear recognition that the data are not strictly speaking comparable across countries because they are not at common international prices.

7. Conclusion

The PWT suffers from problems of variability and valuation. There is considerable variability in the level and growth of PPP-adjusted GDP estimates and in the estimates of the PPPs across alternative versions of the PWT. This variability stems in part from changes to underlying national income accounts data but is also systematically related (inversely) to the size of a country and to the distance of the data from the benchmark data. Because this variability is intrinsic to the PWT methodology there is little basis for knowing whether

to the level estimate in 1996, and extrapolates backwards. Or, we could have a new chained level estimate for 1996 that applies the new growth estimate to the level GDP numbers for 1985. But for pure cross-country comparisons, we would urge that researchers not use the level estimates derived from this chained growth procedure. Rather, they should follow the suggestion earlier about using that version of the PWT that is closest to the period of interest.

newer versions are better than older versions. In fact, the distance from benchmark finding suggests that in every new version, historic data tends to become more variable.

This variability is such that some standard results in the growth literature are not robust across alternative versions of the PWT. It does appear safe to use PPP GDP level data, when looking at cross-sections. Long-run changes, e.g., over 30 years, also appear to be robust. Medium-run growth rates, such as 10-year panels, may also be safe, although we flag this as a point for further investigation.

In terms of robustness to data revisions across PWT version, it is generally not safe to use annual data. The exception would be for countries with quality grades of A and B or if there is a good instrumentation strategy. In general, annual data from non-OECD countries should be treated with caution.

Thus, growth studies should demonstrate robustness to different versions of the PWT, especially where high frequency data are used. Where possible, robustness should also be demonstrated for samples of countries for which benchmark data are available and for samples excluding the small countries or those with a quality grading of C or D. If results do not survive one or both of these checks, a much bigger health warning should be attached to any policy implications.

On valuation, we also found that for years other than the benchmark year, GDP growth and level estimates from the PWT are not at PPP prices. To overcome this problem as well as the systematic variability (with respect to country size and distance from the benchmark year) of the GDP numbers, we suggested an alternative approach to calculating a chained growth estimate.

It remains to be seen how version 7.0 of the PWT will confirm or change any of this assessment. It is also unclear whether that version will definitively supersede all previous versions (and all alternatives, including the WDI).

References

Acemoglu, Daron, Simon Johnson, James Robinson, and Yunyong Thaicharoen. 2003. "Institutional Causes, Macroeconomic Symptoms: Volatility, Crises and Growth." *Journal of Monetary Economics* 50: 49–123.

Aghion, Philippe, Peter Howitt, and David Mayer-Foulkes. 2005. "The Effect of Financial Development on Convergence: Theory and Evidence." *Quarterly Journal of Economics* 120: 173–222.

Almas, Ingvild. 2012. "International Income Inequality : Measuring PPP Bias by Estimating Engel Curves for Food." *American Economic Review*, forthcoming.

Balassa, Béla. 1964. "The Purchasing Power Parity Doctrine: A Reappraisal." *Journal of Political Economy* 72: 584–596.

Barro, Robert J. 1999. "Determinants of Democracy." *Journal of Political Economy* 107: S158–S183.

Burnside, Craig, and David Dollar. 2000. "Aid, Policies, and Growth." *American Economic Review* 90: 847–868.

Ciccone, Antonio, and Marek Jarocinski. 2008. "Determinants of Economic Growth: Will Data Tell?" European Central Bank Working Paper 852. Frankfurt: European Central Bank.

Dawson, John W., Joseph P. DeJuan, John J. Seater, and E. Frank Stephenson. 2001. "Economic Information Versus Quality Variation in Cross-Country Data." *Canadian Journal of Economics* 34: 988–1009.

Deaton, Angus. 2006. "Purchasing Power Parity Exchange Rates for the Poor: Using Household Surveys to Construct PPPs." Princeton University Research Program in Developing Studies working paper. Princeton, NJ: Princeton University.

Deaton, Angus. 2010. "Price Indexes, Inequality, and the Measurement of World Poverty." *American Economic Review* 100: 5–34.

Deaton, Angus, and Alan Heston. 2010. "Understanding PPPs and PPP-Based National Accounts." *American Economic Journal – Macroeconomics*, 2: 1–35.

De Long, J. Bradford, and Lawrence H. Summers. 1991. "Equipment Investment and Economic Growth." *Quarterly Journal of Economics* 106: 445–502.

Demirguc-Kunt, Asli, Luc Laeven, and Ross Levine. 2004. "Market Structure, Institutions, and the Cost of Financial Intermediation." *Journal of Money Credit and Banking* 36: 593–622.

Dowrick, Steve. 2005. The Penn World Table: A Review. *Australian Economic Review* 38: 223–228.

Dowrick, Steve, and John Quiggin. 1997. "True Measures of GDP and Convergence." *American Economic Review* 87: 41–64.

Easterly, William, Michael Kremer, Lant Pritchett, and Lawrence H. Summers. 1993. "Good Policy or Good Luck?: Country Growth Performance and Temporary Shocks." *Journal of Monetary Economics* 32: 459–483.

Feenstra, Robert, Alan Heston, Marcel P. Timmer, and Haiyan Deng. 2009. "Estimating Real Production and Expenditures Across Nations: A Proposal for Improving the Penn World Tables." *Review of Economics and Statistics* 91: 201–212.

Feenstra, Robert, Hong Ma and D.S. Prasada Rao. 2009. "Consistent Comparisons of Real Incomes across Space and Time." UC Davis Department of Economics working paper. Davis, CA: UC Davis.

Geary, Roy C. 1958. "A Note On the Comparison of Exchange Rates and Purchasing Power Between Countries." *Journal of the Royal Statistical Society, Series A (General)* 121: 97–99.

Gershenkron, Alexander. 1947. "The Soviet Indices of Industrial Production." *Review of Economic Studies* 29: 217–26.

Handbook of the International Comparison Program, http://unstats.un.org/unsd/methods/icp/ipc6_htm.

Hausmann, Ricardo, Lant Pritchett, and Dani Rodrik. 2005. "Growth Accelerations." *Journal of Economic Growth* 10: 303–29.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2009. "Measuring Economic Growth from Outer Space." NBER Working Papers No. 15199. Cambridge, MA: National Bureau of Economic Research.

Heston, Alan. 1994. "National Accounts: A Brief Review of Some Problems in Using National Accounts Data in Level of Output Comparisons and Growth Studies." *Journal of Development Economics* 44: 29–52.

Heston, Alan. 2001. "Treatment of China in PWT 6." University of Pennsylvania, http://pwt.econ.upenn.edu/Documentation/China.pdf.

Heston, Alan, Robert Summers, and Bettina Aten. 2002. Penn World Table Version 6.1. Center for International Comparisons at the University of Pennsylvania (CICUP). October.

Heston, Alan, Robert Summers, and Bettina Aten. 2006. Penn World Table Version 6.2. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. September.

IMF (International Monetary Fund). 2008. *World Economic Outlook*. Washington.

Jones, Ben, and Benjamin Olken. 2005. "Do Leaders Matter? National Leadership and Growth Since World War II." *Quarterly Journal of Economics* 120: 835–864.

Katayama, Hajime, and Natalia Ponomareva. 2010. "Does the Version of the Penn World Tables Matter? An Analysis of the Relationship between Growth and Volatility." *Canadian Journal of Economics* 43: 152–179.

Khamis, Salem H. 1972. "A New System of Index Numbers for National and International Purposes." *Journal of the Royal Statistical Society, Series A (General)* 135: 96–121.

Kravis, Irving B., Alan Heston, and Robert Summers. 1978. Real GDP Per Capita for More Than One Hundred Countries. *Economic Journal* 88: 215–242.

Levine, Ross, Norman Loayza, and Thorsten Beck. 2000. "Financial intermediation and growth: Causality and Causes." *Journal of Monetary Economics* 46: 31–77.

Mankiw, N. Gregory, David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107: 407–437.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* 112: 725–753.

Neary, J. Peter. 2004. "Rationalizing the Penn World Table: True Multilateral Indices for International Comparisons of Real Income." *American Economic Review* 94: 1411–1428.

Nuxoll, Daniel A. 1994. "Differences in Relative Prices and International Differences in Growth Rates." *American Economic Review* 84: 1423–1436.

Ramey, Garey, and Valerie A. Ramey. 1995. "Cross-Country Evidence on the Link between Volatility and Growth." *American Economic Review* 85: 1138–1151.

Rao, Prasada D.S., Alicia N. Rambaldi, and Howard E. Doran. 2009. "An Econometric Approach to Construct World Tables of Purchasing Power Parities and Real Incomes: Analytical Properties and Tables for 1970–2005," working paper, University of Queensland.

Rao, Prasada D.S. , and E.A. Selvanathan. 1992. "Computation of Standard Errors for Geary-Khamis Parities and International Prices: A Stochastic Approach." *Journal of Business and Economics Statistics* 10: 109–115.

Sachs, Jeffrey D., and Andrew M. Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* 1995: 1–118.

Samuelson, Paul A. 1994. "Facets of Balassa-Samuelson Thirty Years Later." *Review of International Economics* 3: 201–226.

Srinivasan, T. N. 1994. "Data Base for Development Analysis: An Overview." *Journal of Development Economics* 44: 3–27.

Summers, Robert, and Alan Heston. 1980. "A New Set of International Comparisons of Real Product and Price Levels Estimates For 130 Countries, 1950–1985." *Review of Income and Wealth* 26: 19–66.

Summers, Robert, and Alan Heston (1984) "Improved International Comparisons of Real Product and its Composition, 1950-80," *Review of Income and Wealth 30*: 207–261.

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

Summers, Robert, and Alan Heston. 1996. "International Price and Quantity Comparisons: Potentials and Pitfalls." *American Economic Review* 86: 20–24.

Samuelson Paul. 1964. "Theoretical Notes on Trade Problems." *Review of Economics and Statistics* 23: 145–154.

van Veelen, Matthijs. 2002. "An Impossibility Theorem Concerning Multilateral

International Comparisons of Volumes." *Econometrica* 70: 369–75.

World Bank. 2007. *World Development Indicators.* Available at the World Bank's website, www.worldbank.org.

World Bank. 2008a. *Global Purchasing Power Parities and Real Expenditures: 2005 International Comparison Program.* Available at the World Bank's website, www.worldbank.org.

World Bank. 2008b. *ICP 2003-2006 Handbook.* Available at www.worldbank.org.

Young, Alwyn. 2009. "Real Consumption Measures for the Poorer Regions of the World." London School of Economics working paper. London: London School of Economics.

Table 1. Average Growth Rates in African Countries (1975-1999)

Top 10 Countries				
PWT 6.2		PWT 6.1		
Country	Growth	Country	Growth	Countries Not Appearing on Both Lists
Botswana	4.50%	Botswana	5.10%	Congo, Republic of
Equ.Guinea	4.00%	Cape Verde	4.70%	Equatorial Guinea
Cape Verde	3.70%	Mauritius	4.30%	Ethiopia
Egypt	3.70%	Egypt	3.70%	Malawi
Mauritius	3.70%	Tunisia	2.50%	Mali
Lesotho	3.50%	Uganda	1.70%	Uganda
Tunisia	2.70%	Morocco	1.70%	
Mali	2.00%	Lesotho	1.50%	
Ethiopia	1.60%	Congo, Rep of	1.50%	
Morocco	1.60%	Malawi	1.20%	
Bottom 10 Countries				
PWT 6.2		PWT 6.1		
Country	Growth	Country	Growth	Countries Not Appearing on Both Lists
Gabon	-2.60%	Equ. Guinea	-2.70%	Chad
Zambia	-2.10%	Mozambique	-2.40%	Cote d'Ivoire
Madagascar	-1.90%	Zambia	-1.80%	Gabon
Togo	-1.70%	Comoros	-1.60%	Guinea-Bissau
Guinea-Bissau	-1.40%	Madagascar	-1.40%	Mauritania
Comoros	-1.20%	Cote d'Ivoire	-1.40%	Namibia
Niger	-0.70%	Niger	-1.30%	Nigeria
Nigeria	-0.50%	Mauritania	-1.30%	Togo
Chad	-0.50%	Togo	-1.00%	
Mozambique	-0.40%	Namibia	-0.90%	
				Country Switching Lists: Equ. Guinea

Notes:
This table presents the countries with the top 10 and bottom 10 average growth rates between 1975 and 1999 calculated using PWT 6.2 and PWT 6.1 GDP data (RGDPCH). 40 countries have complete data from 1975-1999. Six countries appear on the top 10 list for one dataset but not the other. Eight countries appear on the Bottom 10 list for one dataset but not the other. One country, Equatorial Guinea, switches lists. According to PWT 6.1, it is the worst-performing country; according to PWT 6.2, it is the second highest performing country.

Table 2. The Evolution of the Penn World Table

PWT Version	Benchmark Year	Year Released	Countries in the PWT
1	1970	1980	119
2	-	-	-
3	1975	1984	115
4	1980	1988	128
5	1985	1991	134
5.6	1985	1995	151
6.1	1996	2002	168
6.2	1996	2006	188
6.3	2005	2009	190
7	2005	2011	190

ICP Phases	Year(s)	Number of Countries
Phase 1	1970	10
Phase 2	1973	16
Phase 3	1975	35
Phase 4	1980	61
Phase 5	1985	62
Phase 6	1996	115
Phase 7	2005	146

Notes:

Penn World Table, Version 2, was never released. There are some discrepancies in these numbers across different PWT sources. This table uses information from the datasets themselves and the PWT 6.1 appendix. Other information can be found in the various papers published along with releases of versions of the Penn World Table.

Table 3. Explaining Differences in Price Levels Across Versions

Dependent variable →	**Abs(difference in log P)**					
	[1]	[2]	[3]	[4]	[5]	[6]
Estimation	**OLS**					
Dataset Comparison →	**PWT6.2 vs. PWT6.1**					
Sample	**1970 to 1999**					
Abs(Difference in NIA GDP Growth)	0.920***					0.00207
	[0.258]					[0.243]
Abs(Difference in PWT-NIA GDP Growth)		0.799***				0.415***
		[0.116]				[0.0656]
PWT Grade (D=1…A=4)			-5.644***		0.258	-0.470
			[0.319]		[0.545]	[0.352]
1996 ICP Participant			0.0121		0.0121	2.275***
			[0.800]		[0.766]	[0.581]
Log(GDP)				-2.247***		-0.713***
				[0.211]		[0.158]
Distance from Benchmark Year, in Decades (1996)					17.61***	8.337***
					[1.819]	[0.875]
Distance * PWT Grade					-5.171***	-2.116***
					[0.608]	[0.264]
China	7.520***	5.865**	4.113*	15.58***	4.113***	10.74***
	[2.309]	[2.495]	[2.466]	[2.468]	[1.530]	[1.995]
Constant	8.268***	8.018***	23.38***	65.57***	3.276**	21.12***
	[0.263]	[0.363]	[1.177]	[5.381]	[1.280]	[3.720]
N	2,845	3,016	3,016	3,016	3,016	2,845
R-squared	0.009	0.058	0.086	0.065	0.191	0.169

Notes:

The dependent variable is the absolute value of the difference in the log-level of prices (P) across PWT 6.2 and PWT 6.1. NIA GDP is the GDP per capita, constant price, series from the World Development Indicators dataset (WDI), as reported by national authorities. WDI 2002 data corresponds to PWT 6.1 and WDI 2005 corresponds to PWT 6.2. PWT-NIA GDP is a GDP per-capita, constant price, series constructed using the Penn World Table's national income accounts datafiles. The variable PWT grade can take on four values, 1-4, with 1 representing the worst PWT rating of D, and 4 representing the best PWT rating of A. The distance variable is calculated as the absolute difference between the year of the observation and 1996 (i.e., abs(t-1996)), and is divided by 10 to express the variable in decades. 1996 is the "Benchmark Year" used in constructing both PWT 6.2 and PWT 6.1. "China" is a fixed effect due to the substantial revisions to the treatment of China between PWT 6.1 and PWT 6.2. The sample consists of the 104 countries in the “Long Run Sample” used in other tables and figures.
***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Robust t-statistics in brackets.

Table 4. Explaining Differences in Levels of GDP Per Capita Across Versions

Dependent variable →	Abs(difference in log per capita PPP-adjusted GDP)					
	[1]	[2]	[3]	[4]	[5]	[6]
Estimation	**OLS**					
Dataset Comparison →	**PWT6.2 vs. PWT6.1**					
Sample	**1970 to 1999**					
Abs(Difference in NIA GDP Growth)	1.714***					0.543
	[0.373]					[0.365]
Abs(Difference in PWT-NIA GDP Growth)		0.752***				0.567***
		[0.102]				[0.104]
PWT Grade (D=1…A=4)			-3.409***		-0.675*	0.443
			[0.254]		[0.374]	[0.342]
1996 ICP Participant			-2.136***		-2.136***	-0.917
			[0.693]		[0.687]	[0.661]
Log(GDP)				-1.965***		-0.991***
				[0.124]		[0.120]
Distance from Benchmark Year, in Decades (1996)					8.028***	4.562***
					[0.991]	[0.810]
Distance * PWT Grade					-2.395***	-1.268***
					[0.317]	[0.242]
China	12.80***	12.08***	9.558***	20.58***	9.558***	15.67***
	[2.701]	[2.929]	[2.843]	[2.825]	[2.394]	[2.667]
Constant	10.78***	10.09***	21.62***	60.57***	12.45***	32.24***
	[0.261]	[0.303]	[0.718]	[3.211]	[1.033]	[2.983]
N	2,845	3,016	3,016	3,016	3,016	2,845
R-squared	0.026	0.088	0.079	0.086	0.113	0.149

Notes:

The dependent variable is the absolute value of the difference in the log-level of GDP across PWT 6.2 and PWT 6.1. NIA GDP is the GDP per capita, constant price, series from the World Development Indicators dataset (WDI), as reported by national authorities. WDI 2002 data corresponds to PWT 6.1 and WDI 2005 corresponds to PWT 6.2. PWT-NIA GDP is a GDP per-capita, constant price, series constructed using the Penn World Table's national income accounts datafiles. The variable PWT grade can take on four values, 1-4, with 1 representing the worst PWT rating of D, and 4 representing the best PWT rating of A. The distance variable is calculated as the absolute difference between the year of the observation and 1996 (i.e., abs(t-1996)), and is divided by 10 to express the variable in decades. 1996 is the "Benchmark Year" used in constructing both PWT 6.2 and PWT 6.1. "China" is a fixed effect due to the substantial revisions to the treatment of China between PWT 6.1 and PWT 6.2. The sample consists of the 104 countries in the “Long Run Sample” used in other tables and figures. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Robust t-statistics in brackets.

Table 5. Explaining Differences in Growth of GDP Per Capita Across Versions

Dependent variable →	**Abs(difference in growth rate of per capita GDP)**					
	[1]	[2]	[3]	[4]	[5]	[6]
Estimation			**OLS**			
Dataset Comparison →			**PWT6.2 vs. PWT6.1**			
Sample			**1970 to 1999**			
Abs(Difference in NIA GDP Growth)	1.080***					0.0484
	[0.157]					[0.0996]
Abs(Difference in PWT-NIA GDP Growth)		0.842***				0.823***
		[0.0316]				[0.0323]
PWT Grade (D=1…A=4)			-1.741***		-1.322***	-0.161**
			[0.0930]		[0.149]	[0.0672]
1996 ICP Participant			0.0843		0.0843	0.109
			[0.239]		[0.239]	[0.132]
Log(GDP)				-0.820***		-0.106***
				[0.0498]		[0.0255]
Distance from Benchmark Year, in Decades (1996)					1.221***	-0.0762
					[0.364]	[0.185]
Distance * PWT Grade					-0.367***	0.0246
					[0.118]	[0.0554]
China	0.276	-0.0270	-0.434	3.589***	-0.434	0.465**
	[0.430]	[0.123]	[0.454]	[0.440]	[0.475]	[0.184]
Constant	2.091***	0.436***	6.771***	22.97***	5.377***	3.395***
	[0.0960]	[0.0681]	[0.306]	[1.287]	[0.457]	[0.663]
N	2,845	3,016	3,016	3,016	3,016	2,845
R-squared	0.048	0.747	0.093	0.101	0.099	0.720

Notes:

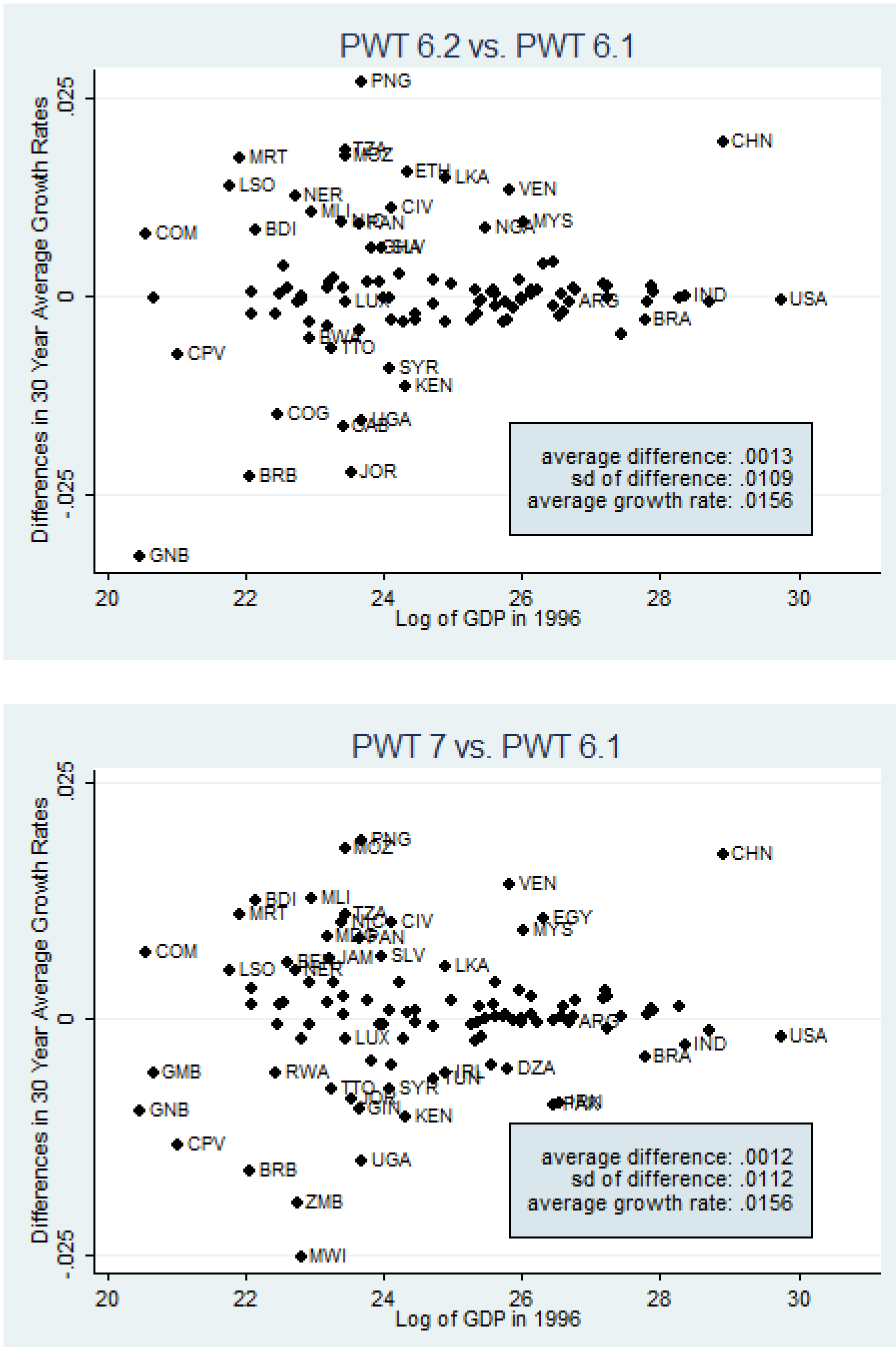
The dependent variable is the absolute value of the difference in the growth rate of GDP across PWT 6.2 and PWT 6.1. NIA GDP is the GDP per capita, constant price, series from the World Development Indicators dataset (WDI), as reported by national authorities. WDI 2002 data corresponds to PWT 6.1 and WDI 2005 corresponds to PWT 6.2. PWT-NIA GDP is a GDP per-capita, constant price, series constructed using the Penn World Table's national income accounts datafiles. The variable PWT grade can take on four values, 1-4, with 1 representing the worst PWT rating of D, and 4 representing the best PWT rating of A. The distance variable is calculated as the absolute difference between the year of the observation and 1996 (i.e., abs(t-1996)), and is divided by 10 to express the variable in decades. 1996 is the "Benchmark Year" used in constructing both PWT 6.2 and PWT 6.1. "China" is a fixed effect due to the substantial revisions to the treatment of China between PWT 6.1 and PWT 6.2. The sample consists of the 104 countries in the “Long Run Sample” used in other tables and figures.
***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Robust t-statistics in brackets.

Table 6. Replication Exercise Results Summary

Author	Journal	Year	Tables Replicated	Replication Exercise Results Summary
Results not robust to dataset changes:				
Aghion, Howitt and Mayer-Foulkes	QJE	2005	Table 1, Cols 1-4	Using PWT 6.2, initial GDP gap becomes insignificant in many exercises. Implied convergence thresholds become implausibly low. Significance is restored using PWT 7.
Hausmann, Rodrik, and Pritchett	JEG	2005	Table 1	Discrepancies identified in 35 out of 100 growth acceleration cases between PWT 6.2 and PWT 6.1. Discrepancies identified in 40 out of 110 cases between PWT 6.1 and 7.
Jones and Olken	QJE	2005	Tables 3, 5	Many differences across exercises using PWT 6.2. Most pronounced are changes in significance of Democrats' and Autocrats' deaths. Results change further using PWT 7.
Mankiw, Romer and Weil	QJE	1992	Table 2, Cols 1, 2	I/GDP coefficient halves in magnitude using PWT 6.2 and falls to nearly zero with PWT 7 in the "I" sample. CRS Solow restriction is rejected in the latter.
Ramey and Ramey	AER	1995	Table 1	Standard deviation of growth rates coefficient becomes weakly significant in one exercise and insignificant in three. The statistical insignificance of the volatility term, holds when using PWT 7.
Results robust to dataset				
Acemoglu, Johnson, Robinson and Thaicharoen	JME	2003	Table 2, Rows 1-4	Minor changes to parameter magnitudes and significance levels.
Barro	JPE	1999	Table 1, Col 2	Minor changes to parameter magnitudes and significance levels.
Burnside and Dollar	AER	2000	Tables 1-3	Some changes in significance levels, but parameter magnitudes remain largely unchanged.
DeLong and Summers	QJE	1991	Table 1	Some changes in parameter magnitudes, but significance levels remain largely unchanged.
Demirguc-Kunt, Laeven and Levine	JMCB	2004	Tables 5, 7	Minor changes to parameter magnitudes and significance levels.
Easterly, Kremer Pritchett, and Summers	JME	1993	Table 5, Cols 2, 4	Minor changes to parameter magnitudes and significance levels.
Miguel, Satyanath and Sergenti	JPE	2004	Table 4, Cols 1-7	Instrumentation required to establish main findings using both PWT 6.2 data and native data. Findings established without instrumentation using PWT 7.
Sachs and Warner		1995	Table 11, Col 7	Some changes in significance levels, but parameter magnitudes remain largely unchanged.

Notes:
This table presents a summary of replication exercises. Detailed results for the top five papers that are not robust to dataset changes are reported in *Supplementary Materials* available online.

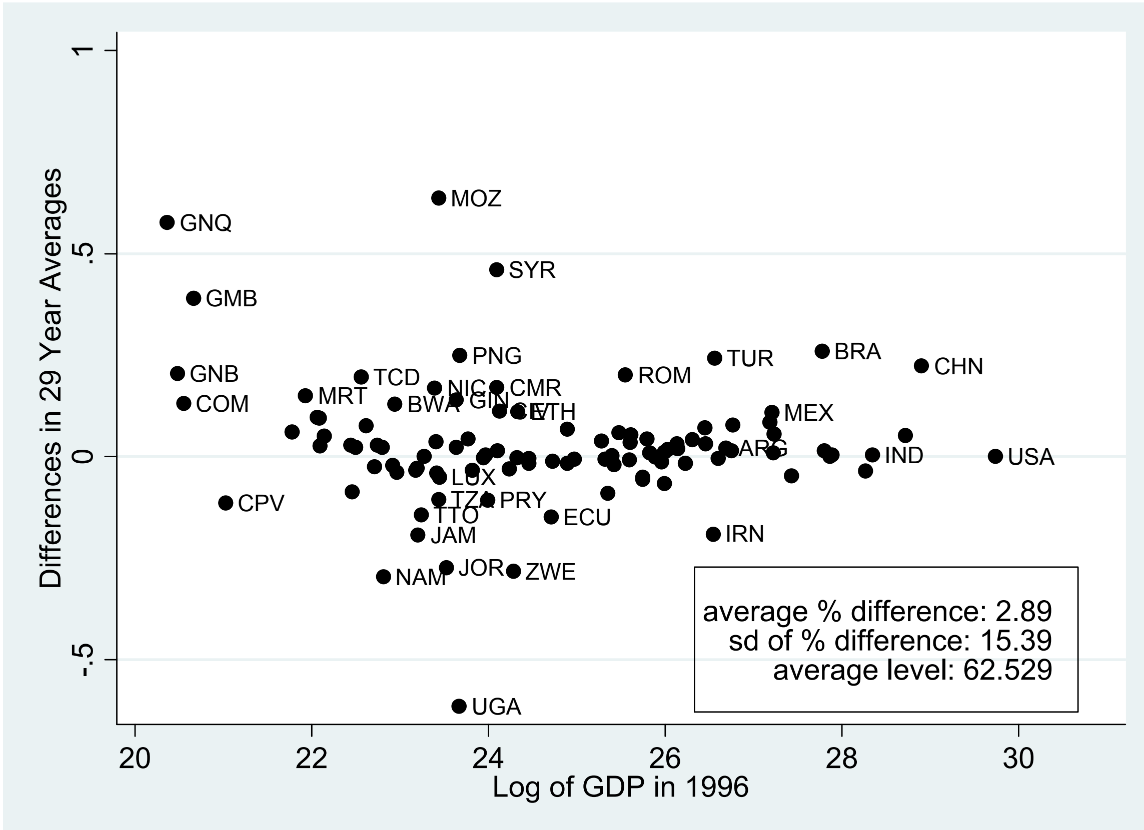
Figure 1. Differences in 29-year average per capita GDP growth rates between PWT 6.2 and PWT 6.1, and PWT 7 and PWT 6.1



Notes: Top panel presents differences between PWT 6.2 and PWT 6.1, and bottom panel differences between PWT 7 and PWT 6.1. Twenty-nine-year average annual per capita GDP growth rates are computed using the RGDPCH series for the period 1970–1999. Differences in growth rates between the two versions of PWT are calculated as GDP growth from PWT 6.2 or PWT 7 minus GDP growth from PWT 6.1. In the top panel, average differences for 29-year average annual per capita GDP growth rates are very close to zero (.0013), whereas the standard deviation is .0109. Average growth rate is about 1.56 percent. These statistics are very similar in the bottom panel. The sample in both panels consists of 104 countries.

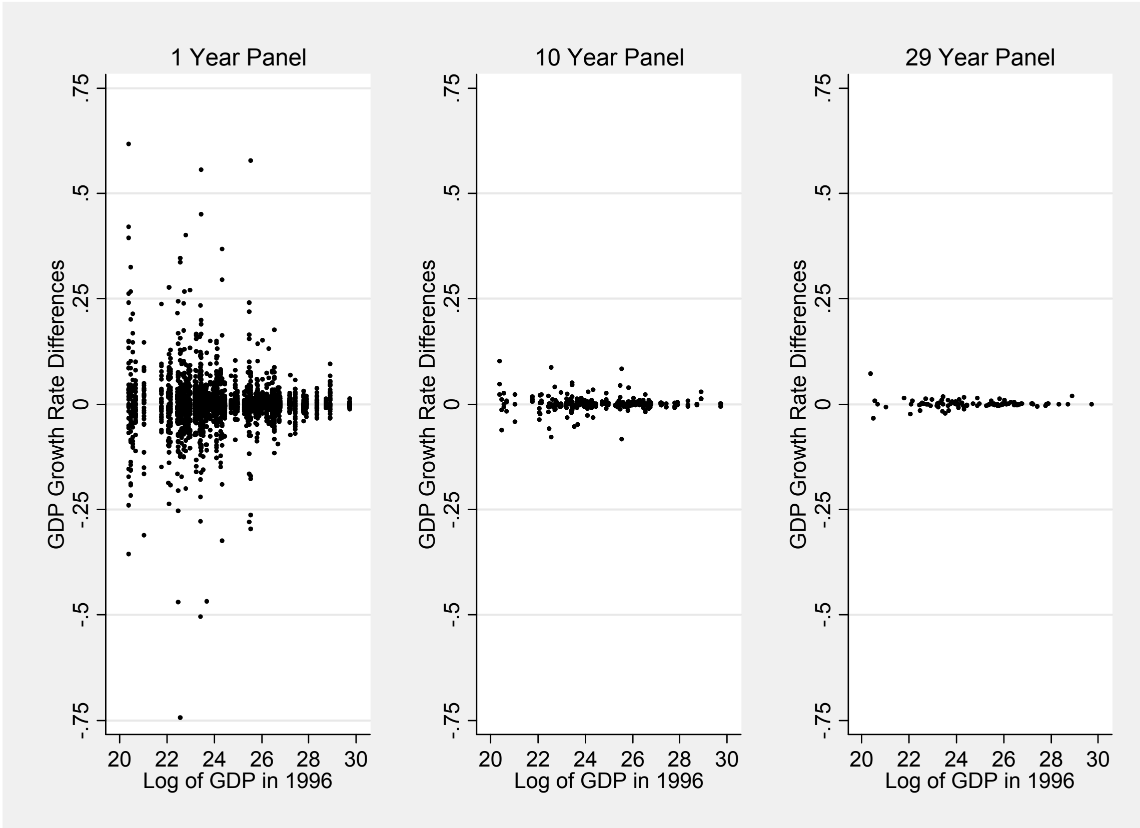
Source: Authors' calculations.

Figure 2. Differences in 29-year average prices between PWT 6.2 and PWT 6.1



Notes: Twenty-nine-year average prices are computed using the average of the P series for the period 1970–1999. Differences in averages between the two versions of PWT are calculated as the log of average prices from PWT 6.2 minus the log of average prices from PWT 6.1. Average differences for 29-year average prices are very close to zero (2.89) whereas the standard deviation is 15.39. The average price level is 62.93, relative to a price level of 100 for the United States. The sample consists of 104 countries.
Source: Authors’ calculations.

Figure 3. Differences in annual per capita GDP growth between PWT 6.2 and PWT 6.1



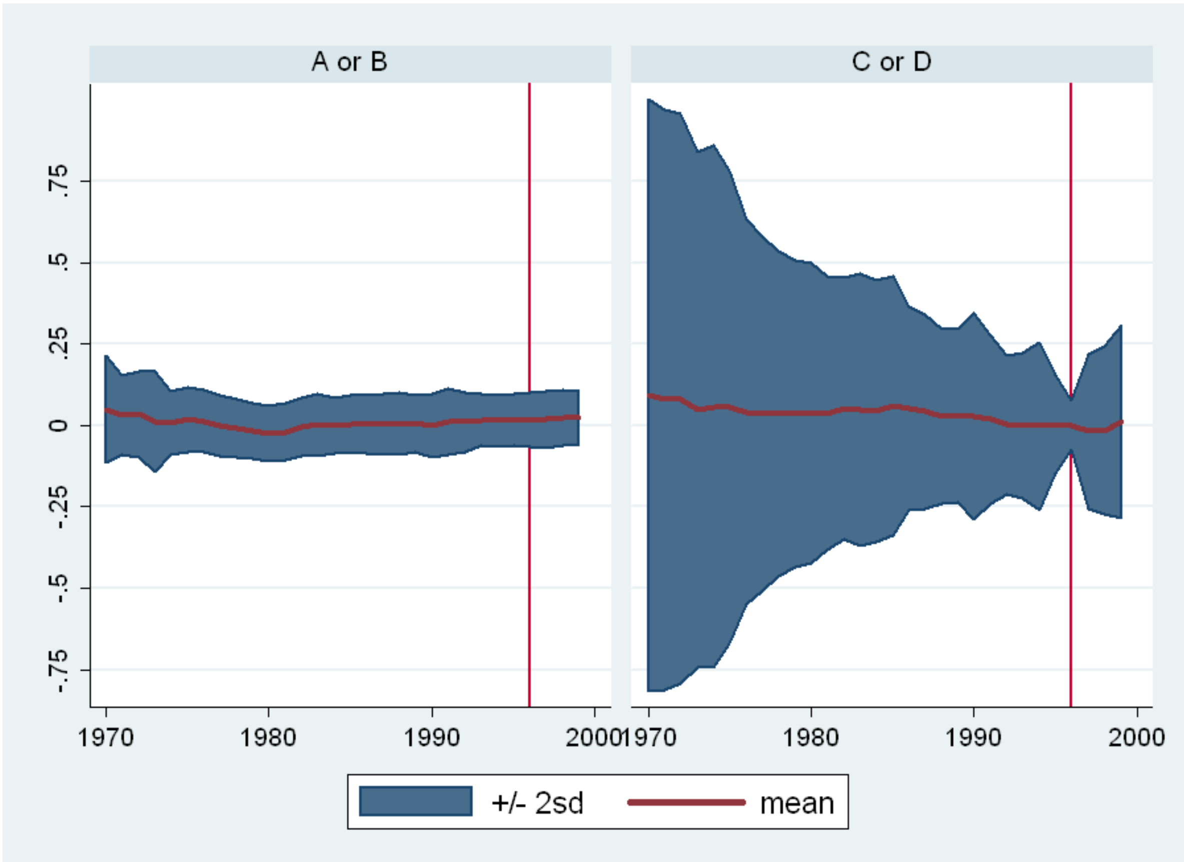
Notes: Differences in growth rates between the two versions of PWT are calculated as GDP growth from PWT 6.2 minus GDP growth from PWT 6.1. 1 Yr Panel: # of obs: 2880; avg. difference = .0008; sd of difference = .0572; avg. growth rate = .0157; 10 Yr Panel: # of obs: 288; difference = .0010; sd of difference = .0174; avg. growth rate = .0155; 29 Yr Panel: # of obs: 96; difference = .0010; sd of difference = .0108; avg. growth rate =.0151. The sample of countries is constant across the three figures above, and consists of 96 of the 104 countries in the "long-run sample". Eight countries were dropped because they did not have data for all three decades. This sample is called the "decades sample."

Source: Authors' calculations.

Figure 4. Differences in 29-year average GDP growth by PWT quality grade between PWT 6.2 and PWT 6.1

Notes: Twenty-nine-year average annual per capita GDP growth rates are computed using the RGDPCH series for the period 1970–1999. Differences in growth rates between the two versions of PWT are calculated as GDP growth from PWT 6.2 minus GDP growth from PWT 6.1. The sample consists of 104 countries. A or B: average difference: -.0002; sd of difference: .0015; average growth rate: .0233; countries: 29. C or D: average difference: .0019; sd of difference: .0127; average growth rate: .0126; countries: 75.
Source: Authors' calculations.

Figure 5. Evolution of differences in levels of prices across time between PWT 6.2 and PWT 6.1



Notes: Prices is the Price (P) series. Differences in levels between the two versions of PWT are calculated as the log level of P from PWT 6.2 minus the log level of P from PWT 6.1. Each mean and standard deviation is computed across countries for a given year. Sample of countries for each year includes countries for which there is data for every year between 1970 and 1999. Vertical line denotes benchmark year, 1996. The sample consists of 104 countries.
Source: Authors' calculations.