

NBER WORKING PAPER SERIES

IS NEWER BETTER? PENN WORLD TABLE REVISIONS AND THEIR IMPACT
ON GROWTH ESTIMATES

Simon Johnson
William Larson
Chris Papageorgiou
Arvind Subramanian

Working Paper 15455
<http://www.nber.org/papers/w15455>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2009

We thank Daron Acemoglu, Francesco Caselli, Michael Clemens, Erwin Diewert, Robert Feenstra, Benjamin Jones, Chad Jones, Aart Kraay, Benjamin Olken, Mick Silver, Antonio Spilimbergo, Nancy Stokey, Alwyn Young, and especially Angus Deaton, Alan Heston, David Romer, and David Weil, for comments, encouragement and suggestions, 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, 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) and the World Bank. The views expressed here are those of the authors only.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Simon Johnson, William Larson, Chris Papageorgiou, and Arvind Subramanian. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates
Simon Johnson, William Larson, Chris Papageorgiou, and Arvind Subramanian
NBER Working Paper No. 15455
October 2009
JEL No. O11,O40,O47

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 this variability is systematic; and that it is intrinsic to the methodology deployed by the PWT to estimate **growth rates**. 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 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.

Simon Johnson
Sloan School of Management
MIT
E52-562
50 Memorial Drive
Cambridge, MA 02142
and NBER
sjohnson@mit.edu

William Larson
Department of Economics
The George Washington University
2115 G ST NW
Washington, DC 20052
larsonwd@gmail.com

Chris Papageorgiou
Research Department
International Monetary Fund
700 19th St. N.W.
Washington, DC 20431
cpapageorgiou@imf.org

Arvind Subramanian
Peterson Institute for International Economics,
Center for Global Development,
and Johns Hopkins University
1750 Massachusetts Ave, NW
Washington, DC 20036
asubramanian@piie.com

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 its latest available version (PWT 6.2) Equatorial Guinea is the second-fastest growing country among 40 African countries. However, according to its 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. The concerns about GDP data quality and the variability it engenders have recently led researchers to explore 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.

¹Roughly 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.

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 and Heston (2008), 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.

The focus of this paper is the time dimension of the PWT, specifically the data revisions across versions of the PWT, its 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 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. And results are 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.

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).² The latest published version is 6.2, which was released in October 2006.³ The prior version was 6.1 and that was preceded by 5.6. These are the most commonly used versions of the PWT today, and it is on these that we focus our attention. However, it is our understanding that the points made below also likely apply

²Appendix figure 1 shows the distribution of countries by the number of benchmark studies in which they participated.

³Version 7.0 will be available in 2009, but the precise release date is not yet known.

to other versions.

The core purpose of the PWT is to collect prices for the same or similar goods in different countries. This price collection operation 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 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

⁴The 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.

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

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

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.

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:

$$Y_{95} = DA_{95} + NFB_{95} \quad (1)$$

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,

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

⁷Ibid.

where the hat sign over a variable denotes 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:

$$\widehat{DA}_{95,96} = a_C \hat{C}_{95,96} + a_I \hat{I}_{95,96} + a_G \hat{G}_{95,96} , \quad (3)$$

where a_C , a_I , a_G are the 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 entirely discarded.⁸ This occurs for every new version so that potentially valuable information from previous ICPs is not used.

3 Data Variability: Patterns and Correlates

We focus on the variability of estimates between PWT version 6.1 and 6.2. These versions differ in some relatively small ways.⁹ Both versions are within the same generation of the Table, and

⁸To be precise, previous benchmark data are used in the calculation of the level of GDP in the subsequent benchmark year: for example, PWT 6.1 incorporated some of the 1985 benchmark year inputs (in PWT 5.6) into the 1996 benchmark year calculations as described on page 9 of the Data Appendix <http://pwt.econ.upenn.edu/Documentation/append61.pdf>.

⁹There 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.

therefore have almost identical methodologies and, more importantly, between these versions there is no new ICP and hence no new international price data.¹⁰ Despite these small changes, 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.¹¹

Figure 1 illustrates that the 29-year average annual growth rates differ between PWT 6.2 and 6.1.¹² 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.

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 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 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

¹⁰In addition, an important point emerges from a careful examination of the methodological history of the tables (e.g., Deaton and Heston 2008). 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.

¹¹The correlation between the 29-year annual average growth rate in the national income accounts data used in the two versions is 0.99.

¹²This 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 RGDPPL series. Appendix table 1 lists the countries in the sample. Figure 1 shows that for the sample for which there are data on real per capita PPP GDP in both PWT 6.1 and PWT 6.2, the magnitudes are broadly similar when we compare PWT 6.2 with WDI or with PWT 5.6 (see table 6) or with WEO numbers (although the sample varies depending on the comparison).

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.

Figure 4 shows some prominent cases of revisions in the GDP numbers across the three latest PWT versions. This affects many of the narratives about growth and the associated policy discussion for some low income countries stemming from the very basic question of who has done well and who has done badly, say over the past 25 years.¹³ Table 1 illustrated the severity of the data problem for Africa.¹⁴

A. Correlates of variability in GDP estimates across versions

Next, we examine systematically—based on simple regressions—the correlates 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.

We need to identify the possible influences. 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. Figure 5 shows the breakdown of

¹³This basic and important question was asked, for example, by the World Bank’s Growth Commission, headed by Michael Spence.

¹⁴More generally, do the leading alternative versions of growth data of the tables agree on the fastest growing countries around the world? For this we can compare “top 10” and “bottom 10” in terms of growth performance according to versions 6.1 and 6.2 of the tables, alongside growth rates calculated from WEO and WDI. All four sources agree on seven of the fastest growing countries: South Korea, China, Botswana, Thailand, Hong Kong, Ireland, and Malaysia. But there are also seven countries that appear on at least one list but not on all lists. And even between the IMF and the World Bank data there is disagreement on the exact ranking (and on whether Indonesia, Portugal, and Luxembourg make the cut). Again, there is more disagreement on the countries that have done worst—probably because poorer performing countries tend to have less reliable data. Only five countries appear on all four lists: Madagascar, Nicaragua, Togo, Venezuela, and Zambia.

countries by their grades in version 6.1. 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 6 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 estimates of the PPPs are inversely related to a country’s size (more specifically, its total consumption expenditure).¹⁵ 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.¹⁶ 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 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.¹⁷

¹⁵See equation 5 in Rao and Selvanathan (1992).

¹⁶We should note that it is not the case that revisions to the table change only more recent data. Appendix figure 2 shows that there are large revisions for the 1970s, the 1980s, and the 1990s looking at 10-year average annual growth rates.

¹⁷Rao and Selvanathan (1992) show that that the standard error of the Generalized 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.

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, country authorities can, and occasionally do, revise data. To capture the source of the revisions, we do the following. We take GDP growth data from the World Development Indicators (which reflect faithfully data submitted by country authorities) for the years corresponding to PWT 6.1 and 6.2, which are 2002 and 2005 respectively. We use the log of the absolute value of differences in the growth of GDP (at constant domestic prices) between these two datasets as our measure of changes in national income accounts data. 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 1.¹⁸ 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.¹⁹

We use two alternative measures for data quality. The first 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). The second measure is simply the number of ICP studies in which a country has participated. For size, we use the log of a country’s GDP. 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).

So, the basic regression we run is:

$$\begin{aligned} dif(x_{it}^{PPP}) = & \alpha_0 + \alpha_1 dif(g_{it}^{NIA}) + \alpha_2 dif(g_{it}^{PWT}) + \beta_1 grade_i^{PWT} + \\ & \beta_2 ICP_i + \beta_3 GDP_{it} + \beta_4 dist_{it} + \beta_5 (dist_{it} * ICP_i) + \varepsilon_{it}, \end{aligned} \quad (4)$$

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.

¹⁸This 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.

¹⁹In tables 3-5, we call the first source of change, “NIA-GDP growth” and the second source of change “PWT-NIA GDP growth.”

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. ICP refers to the number of ICP studies in which a country has participated. 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., $\text{abs}(t-1996)$), and is divided by 10 to express the variable in decades. $dist * ICP$ is the interaction between the distance variable and the number of ICPs variable, and 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 6.1.²⁰ Columns 1 and 2 show that the two sources of changes in the national income accounts data do affect revisions to the PPP. In column 3, it can be seen that a higher grade is negatively related to differences in the data, which confirms that the PWT team warnings on data quality were appropriate. Column 4 shows that if there have been more ICP benchmark studies for a country, this also reduces the extent of data revisions between versions 6.2 and 6.1. Log GDP (column 5) has a similar effect, although this may be a proxy for data quality or the existence of a benchmark study. These results are consistent with the quality grading as advocated for years by the PWT researchers.

Perhaps more novel is the result in column 6. 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 country size (log total GDP) and distance are graphically represented in figures 2 and 7, respectively. Figure 7 is striking in depicting a funnel 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. A key implication is that historical data in the PWT need to be viewed with particular caution. When we combine all our explanatory variables (as in column 7), all of them, except the NIA-GDP growth variable, continue to be significant. That is, the revisions to the PPP

²⁰ 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 regressors.

computed by the PWT are explained by the changes to the national income accounts data made by the PWT as well as other aspects of PWT methodology.

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 with one very important 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. But the revisions to the underlying data that seem to matter are not those generated by the country authorities but those by the PWT procedures: this is most clearly illustrated in column 8, where the variable “PWT-NIA GDP growth” trumps the variable “NIA GDP growth.” Other aspects of the PWT methodology such as data quality and country size remain significant determinants of PPP-based growth revisions.

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 6.1.²¹ 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 are more or less robust.

²¹We 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.

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). Appendix table 2 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 Appendix table 2 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 go through these in more detail below.

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), Demircuc-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 Appendix table 2. 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.²² 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 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. We replicate this result in the first column of table 7. When we re-estimated their core specification on the balanced sample (column 2), 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 (column 3). The same is true if we use the original dataset but drop countries with a data quality grade of D (column 4) or if we drop the same countries while using version 6.2 (column 5).²³

Jones and Olken (2005)

This paper, published in the *Quarterly Journal of Economics* in 2005, relates growth to leader-

²²The 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.

²³Dawson et al. (2003), and Katayama and Ponomareva (2009) also show that data quality in the PWT varies systematically across countries that have different growth rates, and that this problem appears to be serious in estimating the effect of income volatility on growth.

ship. Specifically, it estimates the effect of random leader deaths on a country's growth rate.²⁴ 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 their results are replicated in column 1 of table 8.

When we re-estimated the core specification using data from PWT 6.2 (table 8), 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); see column 3. Second, the striking differences are related to the disaggregation of the results by type of political regime. In Jones and Olken (2005), random leader deaths had a significant impact in autocracies in the year of the leader death and in the two subsequent years (column 1 in the third panel, labeled Autocrats). In the robustness check of column 3 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 (column 3 of the fourth panel, labeled Democrats). 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 (column 5 of table 8), 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 6.2.²⁵

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:

²⁴The 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 $t-5, t-4, \dots, t-1, t+1, t+2, \dots, t+5$ and equal to zero otherwise. These vectors of dummy variables are denoted PRE and $POST$.

²⁵In 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.

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 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.²⁶ Our exercise was simply to see how the list of countries changes when the data from PWT 6.2 are used; the results are in table 9. 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 4 out of 10 instances, the answer depends on which data set is used (table 9).

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 significant effect on civil conflict in Africa.²⁷ 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 (our table 10), we found that the IV

²⁶They 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.

²⁷The 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 $rainfall_{it-1}$ in the 2SLS specifications.

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 (in column 3, 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.

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 level of financial development and the initial level of income divergence with the frontier country, the United States.²⁸ One important prediction of the model for which the estimation provides confirmation is that the coefficient on this interaction is negative (see the second row in column 1 of table 11). When we re-run the Aghion, Howitt, and Mayer-Foulkes (2004) core regression using PWT 6.2 data—shown in table 11—we find that while the magnitude of this coefficient drops by about 40 percent, it remains statistically significant (second row of column 3).

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.

²⁸The 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.

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) (see column 5 of table 11).

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 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.²⁹ 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 (table 12). 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 (in column 5, the p-value for the test of the Solow restriction is 0.166). This result is interesting because it contrasts with the

²⁹The 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.

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.

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:

$$\widehat{DA}_{95,96} = a_C \hat{C}_{95,96} + a_I \hat{I}_{95,96} + a_G \hat{G}_{95,96} .$$

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). In general, this will be different from valuing these quantity changes at international prices and this difference is likely to be systematically larger for smaller countries.

This leads to the striking conclusion: Not only are growth rates of **GDP not** at international prices, but because of this fact, level GDP estimates are also not fully at international prices for nonbenchmark years.

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.³⁰

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

³⁰For example the growth rates in the World Bank's WDI are all at domestic prices.

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, yes.³¹ 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.

Here is a sketch of our proposed method. Assume that the level of GDP for a country at international prices in the benchmark year 1985 is:

$$Y_{85} = c_{85}^1 \pi_{85}^1 + c_{85}^2 \pi_{85}^2 + i_{85} \pi_{85}^i, \quad (5)$$

where the π s 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

³¹Feenstra, 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.

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:

$$Y_{96} = c_{96}^1 \pi_{96}^1 + c_{96}^2 \pi_{96}^2 + i_{96} \pi_{96}^i. \quad (6)$$

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:

$$Y_{85}^2 = c_{85}^1 \pi_{96}^1 + c_{85}^2 \pi_{96}^2 + i_{85} \pi_{96}^i, \quad (7)$$

$$Y_{96}^1 = c_{96}^1 \pi_{85}^1 + c_{96}^2 \pi_{85}^2 + i_{96} \pi_{85}^i. \quad (8)$$

The point is that the data for computing Y according to equations 7 and 8 are 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:

$$\hat{Y}_{85} = \frac{\Delta c^1 \pi_{85}^1 + \Delta c^2 \pi_{85}^2 + \Delta i \pi_{85}^i}{Y_{85}}. \quad (9)$$

Or at current year prices which is:

$$\hat{Y}_{96} = \frac{\Delta c^1 \pi_{96}^1 + \Delta c^2 \pi_{96}^2 + \Delta i \pi_{96}^i}{Y_{85}^2}. \quad (10)$$

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.

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.³²

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

³²Of 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 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.

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 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, surprisingly, 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).

Technical Appendix 1

The Penn World Table Calculation of the Level and Growth of PPP-Adjusted GDP

We outline below the construction of the purchase power parity (PPP)-adjusted level of GDP and its growth rate. In the process, we highlight some key and relatively unknown facts about PWT data construction. The description below refers to version 6.1 of PWT.³³

The procedure used by PWT to construct PPP-adjusted GDPs can be stripped down to two main steps. In the first step, international price levels and PPP-adjusted GDPs are calculated for a benchmark year (for PWT 6.1 in 1996). The second step relates to calculations for nonbenchmark years. To do this, international prices for the benchmark year are extrapolated backward and forward to produce PPP time series for each country considered. Then PPP-adjusted GDP levels and growth rates are calculated for all countries and years using the relevant PPPs from step 2. Next, we look into each of the steps, focusing on what is essential in the final construction of the PWT series.

Step 1: Obtaining PPPs and PPP-adjusted GDPs for the benchmark year

All calculations start for a benchmark year (the year for which detailed price data are collected; for PWT 6.1 in 1996) and for benchmark countries (the countries that participate in the International Comparison Program [ICP]); that is, benchmark countries are those for which actual and detailed price data are collected.³⁴

The raw data for calculating PPPs for the benchmark year are:

1. Expenditures in local currencies $(pq)_{ij}$ (obtained from the national income accounts), where i refers to categories of goods and services and j to the country; in PWT 6.1 there were 31 basic categories of goods and services.
2. The individual prices p_{ij} for these basic goods and services categories in each country covered in the ICP exercise.

Dividing these expenditures by the individual prices yield notional quantities for each of the categories (i.e., $q_{ij} = (pq)_{ij} / p_{ij}$).

These data are used to derive two sets of international prices:

1. A set of international prices for each of the 31 categories of goods and services (π_i); that is, these prices are the same for each category for all countries.³⁵
2. Purchasing power parity for each country (pl_j); PPPs for each country.

³³Special procedures are used for certain countries (i.e., China and India) at various stages. These are not described here and the interested reader is advised to consult the technical appendix to the PWT 6.1.

³⁴See table 1 for a list of benchmark years and benchmark countries used for different versions of PWT.

³⁵International prices for each category are expressed relative to the country-regionplaceUnited States; $\pi_i = 1$ for every i .

How are the (π_i) s and (pl_j) s derived? They are obtained from a system of two sets of simultaneous equations, which is also known as the Geary (1958) and Khamis (1972) (GK thereafter) aggregation procedure. These equations follow from the definition of the two sets of prices as follows:

$$\pi_i = \sum_j \left[\left(\frac{p_{ij}}{pl_j} \right) \left(\frac{q_{ij}}{\sum_j q_{ij}} \right) \right]. \quad (\text{A1})$$

Equation A1 defines the international price for each commodity. The first expression on the right-hand side (p_{ij}/pl_j) is the price of a category of goods i in country j deflated by that country's PPP, and the second $(q_{ij}/\sum_j q_{ij})$ is the share of i in country j in the world consumption of that good. The equation makes clear that the international price for each of the 31 commodities is the weighted average of their domestic prices, where the weights are the share of each country in total world consumption of that commodity.

$$pl_j = \frac{\sum_i (pq)_{ij}}{\sum_i \pi_i q_{ij}}. \quad (\text{A2})$$

Equation A2 defines the price level of a country or its PPP. These relative country prices are the heart of the PWT, subsequently used to produce cross-country comparable GDP measures. For each country, the total expenditure in domestic prices divided by the sum of expenditures on each commodity at international prices yields the PPP for that country.^{36,37}

As a result of the GK procedure, in PWT 6.1 there are international prices for 31 individual commodities, denoted as (π_i) s, and a set of country-specific PPPs, denoted as (pl_j) s.

A final point is worth making here. The 31 categories of goods and services are also placed into three more aggregate categories corresponding to the three basic categories of domestic absorption (DA)—consumption (C), investment (I), and government (G). For each country j , the country-specific PPPs for C , I , and G (plC_j, plI_j, plG_j) are also calculated. These will vary across countries simply because the composition of the 31 categories of goods and services (for which there are common international prices) will vary across countries.

These prices are used to calculate the PPP-adjusted GDPs for the benchmark year. GDP is calculated as the sum of real domestic absorption (measured at international prices) plus the net foreign balance. The constituent elements of domestic absorption— C , I , and G —are also measured

³⁶Once the PPPs are calculated for the benchmark countries for the benchmark year, the PWT proceeds to calculate the same for non-benchmark countries. This is done essentially through a two-step regression procedure, which yields coefficients that can be used to estimate the PPPs for non-benchmark countries. For the purpose of this paper, the estimation procedure is not relevant and is therefore skipped.

³⁷There are also historical data on PPPs obtained from previous ICP rounds. For example, while for PWT 6.1 the ICP was done for 115 countries for the year 1996, a similar exercise was done in 1985 for 64 countries for version PWT 5.6. These 1985 PPPs can be extrapolated using national income accounts deflators for consumption, investment, and government to yield PPPs for 1996. So, for a number of countries, there are multiple data on PPPs for the benchmark year (1996). The PWT then utilizes a weighting method for these multiple data sources to arrive at a final set of international prices or PPPs for the year 1996 for all countries (see pages 9–11 of technical appendix to the PWT 6.1 for details).

at international prices by multiplying the quantities of C , I , and G by their respective international price obtained above.

$$C_j = \pi_c q_c, \quad (\text{A3})$$

where C_j is consumption at international prices, π_c is international price of consumption and q_c is the notional quantity of consumption. By a similar procedure real I and real G are produced. Summing these three yields real domestic absorption as follows:

$$DA_j = C_j + I_j + G_j, \quad (\text{A4})$$

$$Y_j = DA_j + NFB_j. \quad (\text{A5})$$

DA_j is domestic absorption in international prices for country j , Y is PPP-adjusted GDP and NFB is net foreign balance. Note that the net foreign balance does not require the calculation of PPPs as they are values at the price level of DA (see footnote 8).

Step 2A: Calculating international prices for the non-benchmark years

Once PPPs for the benchmark year are calculated, the PWT extrapolates these PPPs going forward and backward in time. How does it do so? Take PWT 6.1 and the year 1995, which is one year before the benchmark year. For 1995, the PWT obtains nominal expenditures for the three components of domestic absorption—consumption, investment, and government—from the national income accounts. Each country j will also have a price level associated with C , I , and G (plC_j, plI_j, plG_j , respectively) for 1995. These are just the 1996 PPPs for each of these categories, calculated as shown above, deflated by the price change between 1995 and 1996 for C , I , and G , where the price changes are obtained from national income accounts.

These extrapolated domestic prices for the three components of domestic absorption obtained (plC, plI, plG) are used in place of the detailed price data collected for the benchmark year. However, unlike in the case of the benchmark year where detailed prices are collected for a large set of commodities, for the nonbenchmark years, each country has only three international prices.

With these prices and with national income data for 1995 on consumption, investment, and government expenditures, the PWT uses the GK aggregation procedure to calculate international prices (π)s for C , I , and G , as well as the PPPs (pl)s for all countries exactly as in step 1.

These international prices are then used to convert consumption, investment, and government expenditures at domestic prices into expenditures at international prices as in equation A3 in step 1.

Step 2B: Calculating the level of PPP-adjusted GDP and its growth rate for non-benchmark years

Step 2A yields real C , real I , and real G for all countries for nonbenchmark years. How is the level of PPP-adjusted GDP then calculated, say for 1995? Essentially via a circuitous procedure that first calculates the growth rate of domestic absorption between 1995 and 1996 at international prices and then applies this growth rate to the 1996 level of DA to derive the level of DA in 1995.

To this, the net foreign balance for 1995 is added to obtain the level of PPP-adjusted GDP for 1995.

$$\widehat{DA}_{95,96} = a_C \hat{C}_{95,96} + a_I \hat{I}_{95,96} + a_G \hat{G}_{95,96} , \quad (\text{A6})$$

$$DA_{95} = DA_{96} / \left(1 + \widehat{DA}_{95,96} \right) , \quad (\text{A7})$$

$$Y_{95} = DA_{95} + NFB_{95} , \quad (\text{A8})$$

where DA is domestic absorption in international prices, the hat sign over a variable denotes growth, a_C , a_I , and a_G are the shares of consumption, investment, and government spending, respectively, in domestic absorption, and NFB is net foreign balance.

Three points are worth emphasizing about equation A6. First, the growth rates of real C , I , and G , are from the national income accounts and mostly do not change across PWT versions. Because they are from national income accounts, these growth rates are at domestic, not PPP, prices. However, the weights assigned to each of these components, the shares of each component in domestic absorption, are measured at international prices in 1995 which are obtained from steps 1 and 2. For example the share of C is given by:

$$a_C = \frac{C}{DA} = \frac{\pi_c q_c}{C + I + G} \quad (\text{A9})$$

Second, the PWT computes two PPP-adjusted GDP series,³⁸ the chained series (RGDPCH), which is the most commonly used and the one recommended by the authors of the PWT, and an RGDPL (or Laspeyres series). The difference between the two is simply that in the RGDPCH, the shares change for every year the growth rate is calculated; hence note that the shares are time sensitive. In contrast, in the RGDPL series, the shares remain the same for all years and are the shares (calculated at international prices) for the benchmark year. Therefore, when the PWT is revised, the shares are revised for the benchmark year and all subsequent years. But for the RGDPL series, the changes in nonbenchmark years are not relevant.

Third, this leads to understanding why the growth rates differ between the two series in the PWT and in turn how these differ from the PPP-adjusted growth rate calculation in the WDI. In the WDI, the growth rate will typically NOT change across revisions (other than to reflect revisions of national accounts) because the growth rates are calculated from the national income accounts. In terms of equation A4, PWT and WDI use the same numbers for real consumption, investment, and government growth, all obtained from the national income accounts. It is the shares that are different: in the WDI, these shares are from the national income accounts themselves, and therefore measured at domestic prices, and change with time; for the RGDPL series, the shares are at international prices but fixed at the levels of the benchmark year; for the RGDPCH series, the shares are at international prices and change every year based on changing international prices.

Steps 2A–2B are then repeated for each of the years before and after the benchmark year to yield PPP-adjusted GDP growth rates and levels for these years.

³⁸There is a third series called RGDPTT that is not discussed here.

Technical Appendix 2

Valuation issues in the Measurement of GDP Growth in the PWT

How are growth rates computed in the PWT? To answer this question and focus on the key points, we assume that GDP comprises consumption (C) and investment (I) and is hence equal to domestic absorption (DA).

One consumption and one investment good

We start with the case where there is only one consumption and one investment good. First, some notation. The quantities of these goods are denoted by small case letters c and i , respectively. Subscripts, which apply only to value and price variables, refer to whether they are measured at domestic or international (i.e., purchasing power parity) prices. Thus P_{dom} refers to domestic price and P_{int} to international prices. Superscripts refer to the year of measurement, and can either be the current year (T) or the base year (B). For example, C^T denotes real consumption for the year T . For the price variables, there will be two superscripts, the first referring to the year of measurement and the second to the good (consumption, c , or investment, i). Thus P_{dom}^{BC} refers to the price of the consumption good measured at domestic prices for the base year. Thus,

$$GDP = DA = C + I = Y, \quad (B1)$$

$$\hat{Y}^T = \frac{C_{int}^T}{Y_{int}^T} \frac{\Delta C_{dom}^T}{C_{dom}^T} + \frac{I_{int}^T}{Y_{int}^T} \frac{\Delta I_{dom}^T}{I_{dom}^T}, \quad (B2)$$

where the hat sign over a variable denotes growth. Equation B2 simply says that GDP growth in year T is a weighted average of growth of consumption and investment. In the PWT's chain series (RGDPCH), the weights are the shares of C and I in domestic absorption measured at current international prices. So, if the growth rate is calculated for the period 1993–1994, the weights are at 1993 PPP prices obtained from the Geary-Khamis aggregation procedure described in step 1 in technical appendix 1. Note that the growth rates of C ($\frac{\Delta C_{dom}^T}{C_{dom}^T}$) and I ($\frac{\Delta I_{dom}^T}{I_{dom}^T}$) are obtained from the national income accounts and are therefore at domestic base-year prices.

Equation B2 can be rewritten as:

$$\hat{Y}^T = \frac{c^T P_{int}^{TC}}{Y_{int}^T} \frac{\Delta c^T}{c^T} \frac{P_{dom}^{BC}}{P_{dom}^{BC}} + \frac{i^T P_{int}^{Ti}}{Y_{int}^T} \frac{\Delta i^T}{i^T} \frac{P_{dom}^{Bi}}{P_{dom}^{Bi}},$$

which in turn simplifies to:

$$\hat{Y}^T = \frac{(\Delta c^T P_{int}^{TC} + \Delta i^T P_{int}^{Ti})}{Y_{int}^T}. \quad (B3)$$

Equation B3 shows that the PWT chain series for the growth rate essentially involves valuing the additions to (the quantities) of consumption and investment at current year international prices. This is a chain-weighted index because the prices used for valuing these additions change every year.

As a result, the RGDPCH series does not use benchmark year international prices; it uses current year international prices that are obtained in the PWT by extrapolating from the benchmark years

(see technical appendix 1, step 2) and applying the GK aggregation procedure. These prices change with every revision as shown below and the revisions are not random but systematically related to country attributes.

A second feature of these current year international prices is that they are calculated at a highly aggregated level, namely at the level of aggregate consumption and investment. For the benchmark year, aggregate consumption is obtained by adding up consumption of the different goods for which disaggregated price data are available. For nonbenchmark years, it is as if there is only one consumption and one investment good.

Two consumption and one investment good

Does this aggregation affect the calculation of growth rates? Suppose there are two consumption goods and one investment good. The spirit of the PWT suggests that all three goods should be valued at international prices. But how are they actually valued? If we had disaggregated international prices, we can write down how growth of DA should be computed. Essentially, equation B2 should be rewritten to take account of the extra consumption good. Thus,

$$\hat{Y}_2^T = \frac{(\Delta c^{T1} P_{int}^{T1} + \Delta c^{T2} P_{int}^{T2} + \Delta i^T P_{int}^{Ti})}{Y_{int}^T}. \quad (B4)$$

Now, the second superscript is 1 or 2 for the two consumption goods and i for the investment good. Equation B4 is just an extension of equation B3 and says that GDP growth is obtained by valuing each of the quantity changes (to the consumption goods and investment good) at their respective current international prices.

But for nonbenchmark years, we do not have disaggregated international price data. GDP growth (\hat{Y}_2^T) is measured as follows:

$$\hat{Y}_2^T = \frac{\bar{C}_{int}^T}{Y_{int}^T} \left(\frac{\Delta C_{dom}^{T1} + \Delta C_{dom}^{T2}}{\bar{C}_{dom}^T} \right) + \frac{I_{int}^T}{Y_{int}^T} \frac{\Delta I_{dom}^T}{I_{dom}^T},$$

which can be simplified to:

$$\hat{Y}_2^T = \frac{(c^{T1} + c^{T2}) \bar{P}_{int}^{TC}}{Y_{int}^T} \left\{ \frac{(\Delta c_{dom}^{T1} P_{dom}^{B1} + \Delta c_{dom}^{T2} P_{dom}^{B2})}{(c^{T1} + c^{T2}) \bar{P}_{dom}^{TC}} \right\} + \frac{\Delta i^T P_{dom}^{Bi}}{Y_{int}^T \bar{P}_{dom}^{Bi}},$$

and further to:

$$\hat{Y}_2^T = \frac{1}{Y_{int}^T} \left\{ \Delta c^{T1} \left(\frac{P_{dom}^{B1}}{\bar{P}_{dom}^{BC}} \bar{P}_{int}^{TC} \right) + \Delta c^{T2} \left(\frac{P_{dom}^{B2}}{\bar{P}_{dom}^{BC}} \bar{P}_{int}^{TC} \right) \right\} + \frac{\Delta i^T P_{int}^{Ti}}{Y_{int}^T}. \quad (B5)$$

Equation B5 shows that each of the quantity changes is valued not at current year international prices (as in equation B4) but at some hybrid of domestic and international prices. The bar sign over a variable denotes that it refers to the composite good as does the second superscript C . The price term in brackets for the two consumption goods consists of $\left\{ \left(\frac{P_{dom}^{B1}}{\bar{P}_{dom}^{BC}} \right) \right\}$ which is the domestic relative price of that good (that is, it the domestic price of good 1 relative to the average price of

the composite of goods 1 and 2); and \bar{P}_{int}^{TC} is the average international price of the composite of goods 1 and 2.

To more clearly identify the difference between how GDP growth ought to be measured and how it is, we can take the difference between equations B4–B5, which yields:

$$\hat{Y}^T - \hat{Y}_2^T = \frac{\bar{P}_{int}^{TC}}{Y_{int}^T} \left\{ \Delta c^{T1} \left(\frac{P_{int}^{T1}}{\bar{P}_{int}^{TC}} - \frac{P_{dom}^{B1}}{\bar{P}_{dom}^{BC}} \right) + \Delta c^{T2} \left(\frac{P_{int}^{T2}}{\bar{P}_{int}^{TC}} - \frac{P_{dom}^{B2}}{\bar{P}_{dom}^{BC}} \right) \right\}. \quad (B6)$$

This equation shows that the difference between the two depends on how different the relative price is of a consumption good at domestic prices ($\frac{P_{dom}^{B1}}{\bar{P}_{dom}^{BC}}$) from its relative price at international prices ($\frac{P_{int}^{T1}}{\bar{P}_{int}^{TC}}$). Note that this difference will vary across time because domestic prices are computed for a fixed base period (which can be different across countries) while the international prices are current prices.

We know that this difference in the relative prices will vary systematically across countries. It will be greater for smaller countries because under the GK 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 true growth rate that is consistent with the spirit of the PWT as represented in equation B4) for smaller countries.

A second problem is that the farther away T 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.

References

- Acemoglu, Daron, Simon Johnson, James Robinson, and Yonyong 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.
- 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. 2003. "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, and Alan Heston. "Understanding PPPs and PPP-Based National Accounts." *American Economic Journal – Macroeconomics*, forthcoming.
- 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_h.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, 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. 2009. "Does the Version of the Penn World Tables Matter? An Analysis of the Relationship between Growth and Volatility." *Canadian Journal of Economics*, forthcoming.
- 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. 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		Countries not appearing on both lists
Country	Growth	Country	Growth	
Botswana	4.50%	Botswana	5.10%	
Equ. Guinea	4.00%	Cape Verde	4.70%	
Cape Verde	3.70%	Mauritius	4.30%	
Egypt	3.70%	Egypt	3.70%	
Mauritius	3.70%	Tunisia	2.50%	
Lesotho	3.50%	Uganda	1.70%	
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		Countries not appearing on both lists
Country	Growth	Country	Growth	
Gabon	-2.60%	Equ. Guinea	-2.70%	
Zambia	-2.10%	Mozambique	-2.40%	
Madagascar	-1.90%	Zambia	-1.80%	
Togo	-1.70%	Comoros	-1.60%	
Guinea-Bissau	-1.40%	Madagascar	-1.40%	
Comoros	-1.20%	Cote d'Ivoire	-1.40%	
Niger	-0.70%	Niger	-1.30%	
Nigeria	-0.50%	Mauritania	-1.30%	
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)

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
7	2005	TBD	TBD

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 Revisions to Price Levels Across Versions

Dependent variable	Abs(difference in log P)							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Estimation	OLS							
Dataset Comparison	PWT6.2 vs. PWT6.1							
Sample	1970 to 1999							
Abs(Difference in NIA GDP Growth)	0.433** [1.98]						0.303 [1.42]	0.0383 [0.169]
Abs(Difference in PWT-NIA GDP Growth)		0.799*** [6.87]						0.486*** [7.581]
PWT Grade (D=1...A=4)			-5.657*** [-15.0]					
Total ICP Studies				-2.409*** [-13.0]			0.208 [1.36]	0.263* [1.719]
Log(GDP)					-2.091*** [-10.1]		-0.863*** [-6.19]	-0.579*** [-4.066]
Distance from Benchmark Year, in Decades (1996)						5.579*** [11.3]	6.158*** [10.8]	5.882*** [10.15]
Distance * Total ICPs							-1.086*** [-7.69]	-1.041*** [-7.254]
Constant	8.651*** [35.2]	8.073*** [22.2]	23.46*** [20.5]	16.11*** [22.6]	61.86*** [11.7]	3.925*** [9.92]	25.75*** [7.65]	17.61*** [5.092]
N	2845	3016	3016	3016	3016	3016	2845	2845
R-squared	0.001	0.057	0.085	0.062	0.058	0.062	0.117	0.147

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Robust t-statistics in brackets.

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. Total ICP studies refers to the number of ICP studies a country has participated in. GDP is total GDP and is measured at purchasing power parity. The distance variable is calculated as the absolute difference between the year of the observation and 1996 (i.e., $\text{abs}(t-1996)$), 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. The sample consists of the 104 countries in the "Long Run Sample" used in other tables and figures.

Table 4. Explaining Revisions to Levels of GDP Per Capita Across Versions

Dependent variable	Abs(difference in log per capita PPP-adjusted GDP)							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Estimation	OLS							
Dataset Comparison	PWT6.2 vs. PWT6.1							
Sample	1970 to 1999							
Abs(Difference in NIA GDP Growth)	1.144*** [3.82]						0.891*** [3.09]	0.563* [1.765]
Abs(Difference in PWT-NIA GDP Growth)		0.752*** [7.35]						0.603*** [6.017]
PWT Grade (D=1...A=4)			-4.036*** [-18.0]					
Total ICP Studies				-2.319*** [-18.1]			-0.838*** [-4.12]	-0.770*** [-3.831]
Log(GDP)					-1.758*** [-13.6]		-0.653*** [-5.06]	-0.302*** [-2.309]
Distance from Benchmark Year, in Decades (1996)						2.456*** [6.99]	3.639*** [5.62]	3.297*** [5.179]
Distance * Total ICPs							-0.741*** [-4.66]	-0.686*** [-4.379]
Constant	11.39*** [45.7]	10.21*** [33.5]	21.69*** [30.7]	17.90*** [35.4]	55.67*** [16.8]	9.495*** [23.7]	27.76*** [8.87]	17.67*** [5.485]
N	2845	3016	3016	3016	3016	3016	2845	2845
R-squared	0.006	0.081	0.07	0.091	0.066	0.019	0.112	0.155

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Robust t-statistics in brackets.

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. Total ICP studies refers to the number of ICP studies a country has participated in. GDP is total GDP and is measured at purchasing power parity. The distance variable is calculated as the absolute difference between the year of the observation and 1996 (i.e., $\text{abs}(t-1996)$), 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. The sample consists of the 104 countries in the "Long Run Sample" used in other tables and figures.

Table 5. Explaining Revisions to Growth of GDP Per Capita Across Versions

Dependent variable	Abs(difference in growth rate of per capita GDP)							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Estimation	OLS							
Dataset Comparison	PWT6.2 vs. PWT6.1							
Sample	1970 to 1999							
Abs(Difference in NIA GDP Growth)	0.694*** [4.03]						0.537*** [3.09]	0.0865 [1.171]
Abs(Difference in PWT-NIA GDP Growth)		0.842*** [26.6]						0.826*** [26.22]
PWT Grade (D=1...A=4)			-1.716*** [-17.8]					
Total ICP Studies				-0.678*** [-14.0]			-0.232*** [-4.82]	-0.0538** [-2.017]
Log(GDP)					-0.784*** [-16.1]		-0.576*** [-11.2]	-0.0967*** [-3.875]
Distance from Benchmark Year, in Decades (1996)						0.366*** [3.21]	0.263** [2.46]	-0.0130 [-0.209]
Constant	2.403*** [26.0]	0.432*** [6.32]	6.767*** [22.6]	4.409*** [23.6]	22.12*** [17.5]	2.355*** [15.4]	16.96*** [12.9]	2.984*** [4.767]
N	2845	3016	3016	3016	3016	3016	2845	2845
R-squared	0.017	0.747	0.093	0.058	0.096	0.003	0.11	0.720

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Robust t-statistics in brackets.

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. Total ICP studies refers to the number of ICP studies a country has participated in. GDP is total GDP and is measured at purchasing power parity. The distance variable is calculated as the absolute difference between the year of the observation and 1996 (i.e., $\text{abs}(t-1996)$), 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. The sample consists of the 104 countries in the "Long Run Sample" used in other tables and figures.

Table 6. Differences Between PWT Versions 6.2, 6.1, and 5.6

	<u>Annual Growth</u> <u>Rates</u> 1970 to 1990	<u>Decadal Growth Rates</u> 1970s 1980s		<u>Long-Run</u> <u>Growth Rate</u> 1970-1990
<u>GDP (RGDPCH)</u>				
<u>PWT 6.2 - PWT 6.1</u>				
Avg GR (6.2)	1.6%	2.2%	1.0%	1.6%
Mean Difference	0.0%	0.0%	0.0%	0.0%
SD Difference	5.3%	1.7%	1.4%	1.0%
<u>PWT 6.2 - PWT 5.6</u>				
Avg GR (6.2)	1.6%	2.2%	1.0%	1.6%
Mean Difference	0.1%	-0.2%	0.5%	0.1%
SD Difference	5.9%	1.9%	1.3%	1.1%
<u>PRICES (P)</u>				
<u>PWT 6.2 - PWT 6.1</u>				
Avg (6.2)	62.5	70.4	67.4	65.0
Mean Difference	2.7%	3.3%	2.4%	2.9%
SD Difference	20.7%	23.6%	14.5%	18.6%
<u>PWT 6.2 - PWT 5.6</u>				
Avg (6.2)	62.5	70.4	67.4	65.0
Mean Difference	-10.2%	-9.8%	-5.1%	-7.5%
SD Difference	27.2%	31.5%	28.9%	29.5%
Obs	2000	100	100	100

Source: Authors' calculations based on PWT 5.6, 6.1 and 6.2.

Table 7. Ramey and Ramey (1995) replication results

Dependent variable \longrightarrow	growth of per capita GDP				
	[1]	[2]	[3]	[4]	[5]
Dataset substitution	None	None	PWT6.2	None	PWT6.2
Sample	Original	Balanced	Balanced	Balanced	Balanced
PWT grades included	All	All	All	A,B,C	A,B,C
Standard deviation of growth rates	-0.177*** [2.426]	-0.151* [1.821]	-0.074 [0.94]	-0.107 [1.254]	-0.132 [1.460]
N	2208	1776	1776	1608	1608

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. t-statistics in brackets.

Notes: This table presents a replication exercise of Ramey and Ramey (1995) by updating their original PWT 5.6 GDP data with PWT 6.2 data. 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 ε_{it} in i across t , 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. All specifications replicate Table 1, equation (1), in Ramey and Ramey (1995). For presentation purposes we focus only on the key parameter - variation of growth rate - and omit other parameter estimates. Specification [1] replicates the main result in Ramey and Ramey (1995) using their original data. Specification [2] presents the same result using PWT 5.6 after dropping observations to balance the data with available observations in PWT 6.2. Specification [3] presents our main result when we replace the balanced sample using PWT 5.6 with PWT 6.2. Specification [4] replicates [2] when countries with quality grading "D" are dropped. Specification [5] replicates [3] when countries with quality grading "D" are dropped.

Source: Authors' calculations based on Ramey and Ramey (1995).

Table 8. Jones and Olken (2005) replication results

Table 3. Jones and Olshok (2008) Replication Results						
Dependent variable	→	Annual growth rate				
Hypothesis	→	Leader deaths affect growth (Wald p-value)				
		[1]	[2]	[3]	[4]	[5]
Dataset substitution		None	None	PWT6.2	None	PWT6.2
Sample		Original	Balanced	Balanced	Balanced	Balanced
PWT grades included		All	All	All	A,B,C	A,B,C
Years after leader's death		All leaders				
	t	0.057*	0.054*	0.093*	0.012**	0.110
	1	0.085*	0.122	0.218	0.028**	0.130
	2	0.067*	0.133	0.203	0.039**	0.141
	Number of leader deaths	57	52	52	45	45
Years after leader's death		All leaders, tenure >=2 Years				
	0	0.039**	0.039**	0.04**	0.011**	0.049**
	1	0.054*	0.087*	0.140	0.021**	0.068*
	2	0.031**	0.076*	0.102	0.023**	0.049**
	Number of leader deaths	47	42	42	36	36
Years after leader's death		Autocrats				
	0	0.019**	0.032**	0.199	0.012**	0.304
	1	0.016**	0.049**	0.356	0.011**	0.227
	2	0.028**	0.100*	0.314	0.026**	0.183
	Number of leader deaths	29	26	26	21	21
Years after leader's death		Democrats				
	0	0.460	0.326	0.044**	0.236	0.039**
	1	0.552	0.415	0.092*	0.291	0.069*
	2	0.432	0.370	0.134	0.359	0.113
	Number of leader deaths	22	20	20	19	19

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. t-statistics in brackets.

Notes: This table presents a replication exercise of Jones and Olken (2005) by updating their original PWT 6.1 GDP data with PWT 6.2 data. 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 fixed country and time effects. For each leader death at t , there are location-specific time dummy variables equal to 1 in one of $t-5, t-4, \dots, t-1, t+1, t+2, \dots, t+5$ and equal to 0 otherwise. These vectors of dummy variables are denoted PRE and $POST$.

Results are Wald p-values of the joint test that $PRE_z = POST_z$. All specifications replicate results from Tables III and V in Jones and Olken (2005). Specification [1] replicates the main result in Jones and Olken (2005) using their original data. Specification [2] presents the same result using PWT 6.1 after dropping observations to balance the data with available observations in PWT 6.2. Specification [3] presents our main result when we replace the balanced sample using PWT 6.1 with PWT 6.2. Specification [4] replicates [2] when countries with quality grading "D" are dropped. Specification [5] replicates [3] when countries with quality grading "D" are dropped. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Source: Authors' calculations based on Jones and Olken (2005).

Table 9. Hausmann, Pritchett and Rodrik (2005) replication results

<u>PWT 6.1</u> (Hausmann, Pritchett, and Rodrik)		<u>PWT 6.2</u>		<u>PWT 6.1</u> (Hausmann, Pritchett, and Rodrik)		<u>PWT 6.2</u>					
ARG	1963	ARG	1963	IDN	1967	IDN	1967	PAK	1979	PAK	1977
ARG	1990	ARG	1990	IDN	1987	IDN	1985	PAN	1959	PAN	1959
AUS	1961	No break		IND	1982	No break		PAN	1975	PAN	1975
BEL	1959	BEL	1959	IRL	1958	IRL	1958	PER	1959	PER	1959
BRA	1967	BRA	1968	IRL	1985	IRL	1986	No break		PHL	1970
No break		BWA	1984	No break		IRN	1966	PNG	1987	No break	
CAN	1962	CAN	1961	ISR	1957	ISR	1957	POL	1992	POL	1992
CHL	1986	CHL	1986	No break		ISR	1989	No break		PRT	1959
CHN	1978	CHN	1977	ISR	1967	ISR	1967	PRT	1985	PRT	1985
CHN	1990	CHN	1991	JOR	1973	No break		PRY	1974	PRY	1974
CMR	1972	No break		JPN	1958	JPN	1958	ROM	1979	ROM	1971
No break		CMR	1978	No break		JPN	1984	RWA	1975	RWA	1975
COG	1969	COG	1968	KOR	1962	KOR	1963	SGP	1969	SGP	1967
COG	1978	COG	1976	KOR	1984	KOR	1984	No break		SGP	1987
COL	1967	COL	1967	LKA	1979	No break		SYR	1969	SYR	1969
DNK	1957	DNK	1957	No break		LKA	1958	SYR	1974	No break	
DOM	1969	DOM	1969	LSO	1971	LSO	1971	SYR	1989	SYR	1991
DOM	1992	DOM	1991	No break		LSO	1992	TCD	1973	No break	
DZA	1975	No break		MAR	1958	MAR	1958	THA	1957	THA	1957
ECU	1970	ECU	1970	MLI	1972	MLI	1973	THA	1986	THA	1986
EGY	1976	EGY	1975	MUS	1971	MUS	1970	TTO	1975	No break	
ESP	1959	ESP	1959	MUS	1983	MUS	1983	TUN	1968	TUN	1968
ESP	1984	ESP	1984	MWI	1970	MWI	1970	TWN	1961	TWN	1961
No break		ETH	1988	MWI	1992	MWI	1990	No break		TWN	1985
FIN	1958	FIN	1958	MYS	1970	MYS	1967	No break		TZA	1992
FIN	1967	No break		MYS	1988	MYS	1988	UGA	1977	No break	
No break		GAB	1969	No break		NER	1974	UGA	1989	No break	
FIN	1992	FIN	1992	NGA	1957	No break		URY	1974	URY	1974
GBR	1982	GBR	1982	NGA	1967	NGA	1968	URY	1989	URY	1988
GHA	1965	GHA	1965	NIC	1960	NIC	1960	USA	1961	USA	1961
GNB	1969	No break		NOR	1991	NOR	1992	No break		VEN	1971
GNB	1988	No break		NZL	1957	NZL	1957	No break		ZMB	1963
No break		HND	1974	PAK	1962	PAK	1961	ZWE	1964	ZWE	1967
HTI	1990	No break									

Notes: This table compares growth breaks in per capita GDP (RGDPCH series) obtained from using the original PWT 6.1 in Hausmann, Pritchett, and Rodrik (2005) with those obtained using PWT 6.2. Growth breaks are reproduced using Hausmann, Pritchett, and Rodrik (2005) original Gauss code. **No break** indicates an inconsistency between PWT 6.1 and PWT 6.2 in obtaining growth break dates that are more than three years apart for each country considered. Hausmann, Pritchett, and Rodrik (2005) identified an additional break for BWA in 1969 using PWT 6.1, but data for Botswana does not go farther back than 1970 in PWT 6.2. There are 35 cases (out of 100; 35 percent) where such inconsistencies are detected.

Source: Authors' calculations based on Hausmann, Pritchett, and Rodrik (2005).

Table 10. Miguel, Satyanath and Sergenti (2004) replication results

Dependent variable	→ log (civil conflict)					
	[1]	[2]	[3]	[4]	[5]	[6]
Estimation	OLS	OLS	OLS	2SLS	2SLS	2SLS
Dataset substitution	None	None	PWT	None	None	PWT
Sample	Original	Balanced	Balanced	Original	Balanced	Balanced
PWT grades included	All	All	All	All	All	All
Economic Growth Rate, t	-0.145 [0.767]	-0.144 [0.727]	-0.327* [1.668]	-0.383 [0.276]	-0.284 [0.217]	-1.985 [1.351]
Economic Growth Rate, t-1	0.071 [-0.368]	0.079 [-0.395]	0.006 [-0.029]	-2.139** [2.078]	-2.078** [2.096]	-2.307** [2.327]
N	743	724	724	743	724	724
R-squared	0.52	0.50	0.49	0.39	0.38	0.29

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. t-statistics in brackets.

Notes: This table presents a replication exercise of Miguel, Satyanath and Sergenti (2004) by updating their original PWT 5.6 GDP data with PWT 6.2 data. The estimated equation is: $conflict_{it} = \alpha_i + X_{it}'\beta + \gamma_0 growth_{it} + \gamma_1 growth_{it-1} + \delta_i year_t + \varepsilon_{it}$,

with $growth_{it}$ instrumented by $rainfall_{it}$ and $rainfall_{it-1}$ in the 2SLS specifications. X contains the log of per-capita GDP in 1979, a lagged democracy index (Polity IV), ethnolinguistic fractionalization, religious fractionalization, an oil-exporter dummy variable, an index of topography (mountains), and the log of lagged population. All specifications replicate table 4, equations 3 and 5, in Miguel, Satyanath and Sergenti (2004). For presentation purposes we focus only on the key parameter - current and lagged growth rates - and omit other parameter estimates. Specification [1] replicates the main OLS result in Miguel, Satyanath and Sergenti (2004) using their original data. Specification [2] presents the same result using PWT 5.6 after dropping observations to balance the data with available observations in PWT 6.2. Specification [3] presents our main result when we replace the balanced sample using PWT 5.6 with PWT 6.2. Specifications [4], [5], [6] replicate specifications [1], [2], [3], respectively, using 2SLS, using rainfall as an instrument for current and lagged economic growth.

Source: Authors' calculations based on Miguel, Satyanath, and Sergenti (2004).

Table 11. Aghion, Howitt and Mayer-Foulkes (2005) replication results

Dependent variable	log (growth gap)				
	[1]	[2]	[3]	[4]	[5]
Dataset substitution	None	None	PWT6.2	None	PWT6.2
Sample	Original	Balanced	Balanced	Balanced	Balanced
PWT grades included	All	All	All	A,B,C	A,B,C
Financial development	-0.015 [0.93]	-0.015 [0.94]	0.011 [0.93]	-0.009 [0.60]	0.013 [1.21]
Financial development * Initial GDP gap (1960)	-0.061*** [5.35]	-0.061*** [4.84]	-0.041*** [4.16]	-0.048*** [3.87]	-0.031** [3.25]
Initial GDP gap (1960)	1.507*** [3.14]	1.505*** [2.83]	0.402 [1.02]	1.031* [1.95]	0.090 [0.25]
N	71	60	60	57	57
R-squared	0.51	0.34	0.37	0.31	0.37
Implied convergence threshold, $-\beta_y / \beta_{fy}$	24.70	24.83	9.76	21.61	2.92
Number of countries above threshold	37 out of 71	37 out of 71	65 out of 71	48 out of 71	71 out of 71
Instruments: legal origins, legal origins * initial GDP gap					
Conditioning set: EMPTY					

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. t-statistics in brackets.

Notes: This table presents a replication exercise of Aghion, Howitt, and Mayer-Foulkes (2005) by updating their original GDP data obtained from Levine, Loayza and Beck (2000) dataset with PWT 6.2 data. 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. All

specifications replicate table 1, equation (1), in Aghion, Howitt, and Mayer-Foulkes (2005). Specification [1] replicates the main result in Aghion, Howitt, and Mayer-Foulkes (2005) using their original data. Specification [2] presents the same result using Levine, Loayza and Beck's data after dropping observations to balance the data with available observations in PWT 6.2. Specification [3] presents our new result when we replace the balanced sample using Levine, Loayza and Beck's data with PWT 6.2 data. Specification [4] replicates [2] when countries with quality grading "D" are dropped. Specification [5] replicates [3] when countries with quality grading "D" are dropped.

Source: Authors' calculations based on Aghion, Howitt, and Mayer-Foulkes (2005).

Table 12. Mankiw, Romer, and Weil (1992) replication results

Dependent variable	Log of GDP per capita in 1985				
	[1]	[2]	[3]	[4]	[5]
Dataset substitution	None	None	PWT6.2	None	PWT6.2
MRW sample	Non-Oil	Non-Oil	Non-Oil	Non-Oil	Non-Oil
Sample	Original	Balanced	Balanced	Balanced	Balanced
PWT grades included	All	All	All	A,B,C	A,B,C
log(I/GDP)-log(n+g+d)	0.738*** [5.96]	0.713*** [5.21]	0.431*** [3.22]	0.754*** [5.28]	0.430*** [2.99]
log(school) –log(n+g+d)	0.657*** [9.07]	0.622*** [7.92]	0.711*** [6.98]	0.639*** [7.59]	0.761*** [6.86]
N	98	83	83	77	77
R-squared	0.78	0.79	0.73	0.78	0.71
Solow restriction p-value	0.654	0.653	0.210	0.584	0.166

***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. t-statistics are in brackets.

Notes: This table presents a replication exercise of Mankiw, Romer, and Weil (1992) by updating their original PWT 4.0 GDP and I/GDP data with PWT 6.2 data. 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 from 1960-1985. All specifications replicate the Table 2 restricted regressions in Mankiw, Romer, and Weil (1992). The "Solow Restriction" is the restriction that $\pi_1 = \pi_2$. Specification [1] replicates the main result in Mankiw, Romer, and Weil (1995) using their original data. Specification [2] presents the same result using PWT 4.0 for the sample of non-oil countries after dropping observations to balance the data with available observations in PWT 6.2. Specification [3] presents our main result when we replace the balanced sample using PWT 4.0 with PWT 6.2. Specification [4] replicates [2] when countries with quality grading "D" are dropped. Specification [5] replicates [3] when countries with quality grading "D" are dropped.

Source: Authors' calculations based on Mankiw, Romer, and Weil (1992).

Appendix Table 1. Countries in the long-run sample

ISO	Country	GDP	Differences between PWT 6.2 and PWT 6.1				Benchmarks	Grade
			Prices	GDP Growth	I/GDP	C/GDP		
ARG	Argentina	26.69	2%	-0.08%	-9%	-1%	2	B
AUS	Australia	26.77	8%	0.06%	-1%	-5%	3	A
AUT	Austria	25.99	-6%	-0.02%	-7%	-3%	5	A
BDI	Burundi	22.14	5%	0.84%	-30%	2%	0	C
BEL	Belgium	26.14	2%	0.02%	-5%	2%	6	A
BEN	Benin	22.61	8%	0.10%	16%	-9%	2	C
BFA	Burkina Faso	22.92	-2%	-0.32%	3%	13%	0	C
BOL	Bolivia	23.77	4%	0.17%	6%	0%	2	C
BRA	Brazil	27.77	30%	-0.30%	-7%	3%	3	C
BRB	Barbados	22.07	10%	-2.29%	-68%	-15%	2	C
BWA	Botswana	22.94	14%	-0.54%	2%	-7%	3	C
CAN	Canada	27.24	6%	-0.03%	0%	1%	4	A
CHE	Switzerland	25.96	-1%	0.20%	4%	2%	2	A
CHL	Chile	25.74	-5%	-0.34%	11%	-5%	2	B
CHN	China	28.9	25%	1.95%	49%	-16%	0	C
CIV	Cote d'Ivoire	24.12	12%	1.11%	5%	-4%	3	C
CMR	Cameroon	24.09	19%	-0.03%	-33%	-7%	3	C
COG	Congo, Republic of	22.46	-8%	-1.48%	16%	9%	2	C
COL	Colombia	26.13	3%	0.07%	7%	1%	3	C
COM	Comoros	20.55	14%	0.79%	56%	-9%	0	D
CPV	Cape Verde	21.02	-11%	-0.73%	-11%	-7%	0	D
CRI	Costa Rica	23.94	0%	0.17%	-42%	17%	1	C
DNK	Denmark	25.6	4%	0.03%	-6%	-8%	5	A
DOM	Dominican Republic	24.45	-2%	-0.31%	-18%	-3%	1	C
DZA	Algeria	25.8	4%	-0.30%	-9%	-12%	0	D
ECU	Ecuador	24.71	-14%	0.21%	17%	6%	2	C
EGY	Egypt	26.3	4%	0.41%	17%	2%	2	C
ESP	Spain	27.23	1%	0.13%	-7%	-11%	5	B
ETH	Ethiopia	24.33	12%	1.55%	-16%	5%	2	C
FIN	Finland	25.28	4%	-0.32%	10%	-21%	4	A
FJI	Fiji	22.09	3%	-0.22%	-6%	10%	1	C
FRA	France	27.89	0%	0.06%	-7%	-12%	6	A
GAB	Gabon	23.4	4%	-1.65%	-61%	-43%	1	C
GBR	United Kingdom	27.86	0%	0.12%	-5%	-15%	6	A
GER	Germany	28.27	-4%	-0.02%	4%	-1%	1	B
GHA	Ghana	23.82	-3%	0.61%	-4%	17%	0	C
GIN	Guinea	23.64	15%	-0.42%	-28%	-7%	1	C
GMB	Gambia, The	20.66	48%	-0.03%	26%	12%	0	C
GNB	Guinea-Bissau	20.48	23%	-3.30%	-25%	-36%	0	D
GNQ	Equatorial Guinea	20.36	78%	7.18%	-30%	-28%	0	D
GRC	Greece	25.61	5%	-0.14%	-4%	-1%	4	B
GTM	Guatemala	24.46	0%	-0.23%	-7%	-1%	1	C
HKG	Hong Kong	25.88	0%	-0.14%	-2%	1%	3	A
HND	Honduras	23.27	0%	0.23%	1%	1%	1	C
HUN	Hungary	25.32	-1%	0.08%	-3%	-20%	5	C
IDN	Indonesia	27.43	-5%	-0.48%	9%	-9%	2	C
IND	India	28.35	0%	-0.01%	-10%	-9%	4	C
IRL	Ireland	24.89	7%	-0.33%	13%	-8%	5	A
IRN	Iran	26.54	-17%	-0.24%	56%	-19%	4	C

Appendix Table 1 (contd.) . Countries in the long-run sample

ISO	Country	GDP	Differences between PWT 6.2 and PWT 6.1				Benchmarks	Grade
			Prices	GDP Growth	I/GDP	C/GDP		
ISL	Iceland	22.5	2%	0.03%	-5%	-7%	2	B
ISR	Israel	25.43	-2%	-0.06%	-11%	-9%	2	B
ITA	Italy	27.8	1%	-0.09%	-1%	-5%	6	A
JAM	Jamaica	23.2	-18%	0.17%	-8%	23%	3	C
JOR	Jordan	23.52	-24%	-2.21%	8%	11%	1	C
JPN	Japan	28.72	5%	-0.07%	-1%	-8%	6	A
KEN	Kenya	24.32	0%	-1.15%	10%	-1%	5	C
KOR	Korea, Republic of	27.19	9%	0.16%	2%	-3%	5	B
LKA	Sri Lanka	24.89	-2%	1.49%	27%	-10%	4	C
LSO	Lesotho	21.78	6%	1.39%	13%	15%	0	D
LUX	Luxembourg	23.44	-5%	-0.09%	8%	-7%	5	A
MAR	Morocco	25.4	0%	-0.08%	-9%	1%	3	C
MDG	Madagascar	23.18	-3%	-0.38%	33%	-9%	3	C
MEX	Mexico	27.21	11%	0.13%	-2%	-6%	3	C
MLI	Mali	22.96	-4%	1.05%	14%	-7%	3	C
MOZ	Mozambique	23.44	89%	1.76%	17%	-8%	0	D
MRT	Mauritania	21.93	16%	1.73%	110%	-2%	0	C
MUS	Mauritius	23.41	-4%	0.09%	-8%	-4%	2	C
MWI	Malawi	22.8	2%	-0.06%	-23%	7%	4	C
MYS	Malaysia	26.03	2%	0.94%	-2%	-2%	2	C
NAM	Namibia	22.81	-26%	0.00%	-36%	1%	0	D
NER	Niger	22.71	-3%	1.25%	-10%	-14%	0	D
NGA	Nigeria	25.47	6%	0.87%	-16%	-3%	3	C
NIC	Nicaragua	23.39	18%	0.92%	-21%	-11%	0	C
NLD	Netherlands	26.6	-1%	-0.20%	-1%	-6%	6	A
NOR	Norway	25.59	-1%	0.06%	-11%	-27%	4	A
NPL	Nepal	24.1	1%	-0.30%	4%	-1%	2	C
NZL	New Zealand	24.97	-1%	0.16%	-3%	-15%	1	B
PAK	Pakistan	26.46	3%	-0.13%	6%	-7%	4	C
PAN	Panama	23.64	2%	0.92%	-9%	14%	2	C
PER	Peru	25.35	-9%	-0.22%	-7%	-1%	2	C
PHL	Philippines	26.23	-2%	0.08%	-7%	-2%	5	C
PNG	Papua New Guinea	23.68	28%	2.69%	-38%	-5%	0	D
PRT	Portugal	25.75	-5%	-0.07%	-4%	-5%	4	B
PRY	Paraguay	23.99	-10%	-0.02%	1%	18%	1	C
ROM	Romania	25.55	22%	0.05%	-28%	-15%	2	C
RWA	Rwanda	22.44	3%	-0.22%	-31%	13%	1	C
SEN	Senegal	23.19	-3%	0.10%	-20%	13%	3	C
SLV	El Salvador	23.97	0%	0.60%	9%	17%	1	C
SWE	Sweden	25.99	1%	-0.05%	0%	-7%	3	A
SYR	Syria	24.09	58%	-0.91%	-34%	-8%	2	C
TCO	Chad	22.56	22%	0.38%	-8%	6%	0	D
TGO	Togo	22.08	10%	0.04%	44%	-7%	0	D
THA	Thailand	26.75	1%	0.10%	1%	6%	3	C
TTO	Trinidad & Tobago	23.24	-13%	-0.65%	91%	-1%	2	C
TUN	Tunisia	24.73	-1%	-0.09%	21%	2%	3	C
TUR	Turkey	26.56	27%	0.02%	-14%	1%	3	C
TZA	Tanzania	23.44	-10%	1.83%	-77%	24%	3	C
UGA	Uganda	23.67	-46%	-1.56%	15%	-6%	0	D
URY	Uruguay	24.24	-3%	0.28%	10%	-1%	3	B
USA	USA	29.74	0%	-0.06%	-2%	0%	6	A
VEN	Venezuela	25.83	1%	1.35%	1%	1%	2	C
ZAF	South Africa	26.45	7%	0.42%	-21%	5%	0	C
ZMB	Zambia	22.74	3%	-0.07%	62%	25%	4	C
ZWE	Zimbabwe	24.28	-25%	-0.34%	-36%	6%	3	C

Appendix Table 2. Replication exercise results summary

Author	Journal	Year	Tables Replicated	Replication Exercise Results Summary
Results not robust to dataset				
Aghion, Howitt and Mayer-Foulkes	QJE	2005	Table 1, Cols 1-4	Initial GDP gap becomes insignificant in many exercises. Implied convergence thresholds become implausibly low.
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.
Jones and Olken	QJE	2005	Tables 3, 5	Many differences across exercises. Most pronounced are changes in significance of Democrats' and Autocrats' deaths.
Ramey and Ramey	AER	1995	Table 1	Standard deviation of growth rates coefficient becomes weakly significant in one exercise and insignificant in three.
Results robust to dataset changes:				
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.
Mankiw, Romer and Weil	QJE	1992	Table 2, Cols 1, 2	I/GDP coefficient halves in magnitude, nearly violating CRS Solow restriction in some exercises. Otherwise estimates are robust.
Miguel, Satyanath and Sergenti	JPE	2004	Table 4, Cols 1-7	Instruments seem to reduce or eliminate parameter estimate differences when using PWT 6.2 data vs. their data.
Sachs and Warner		1995	Table 11, Col 7	Some changes in significance levels, but parameter magnitudes remain largely unchanged.

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

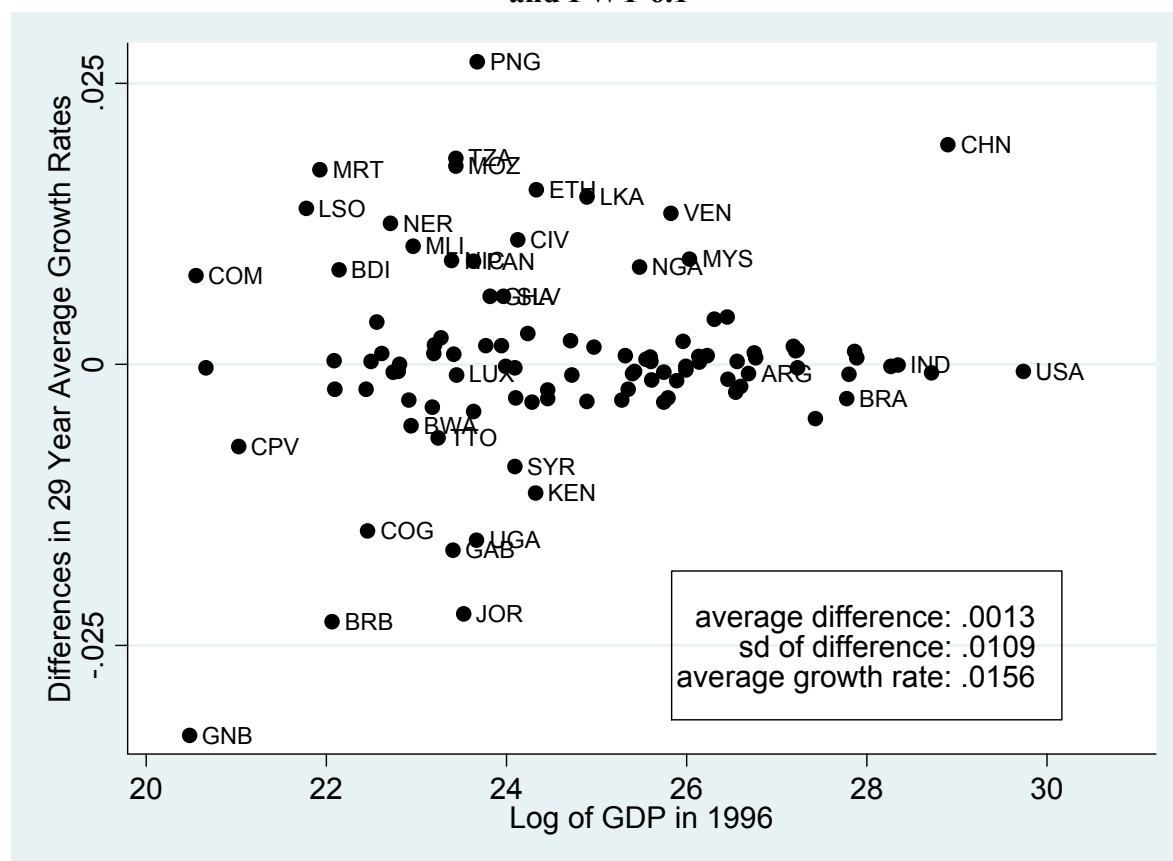
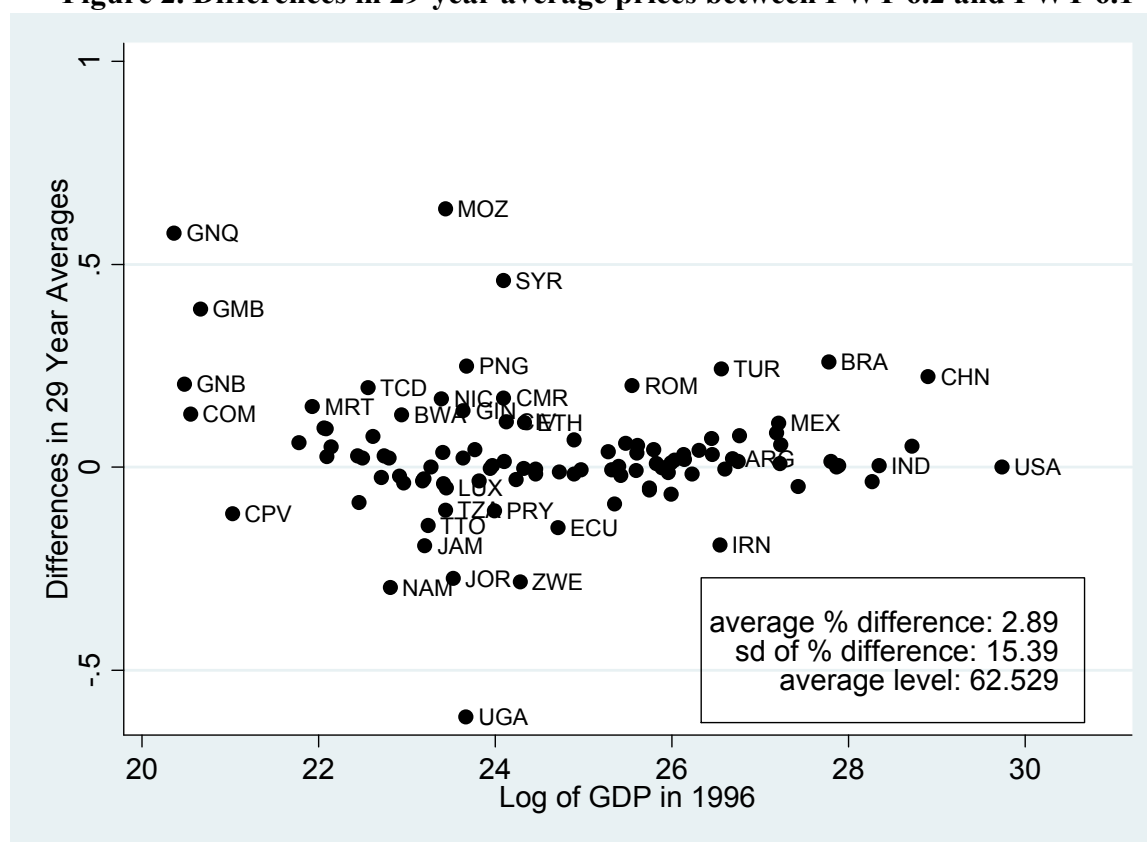


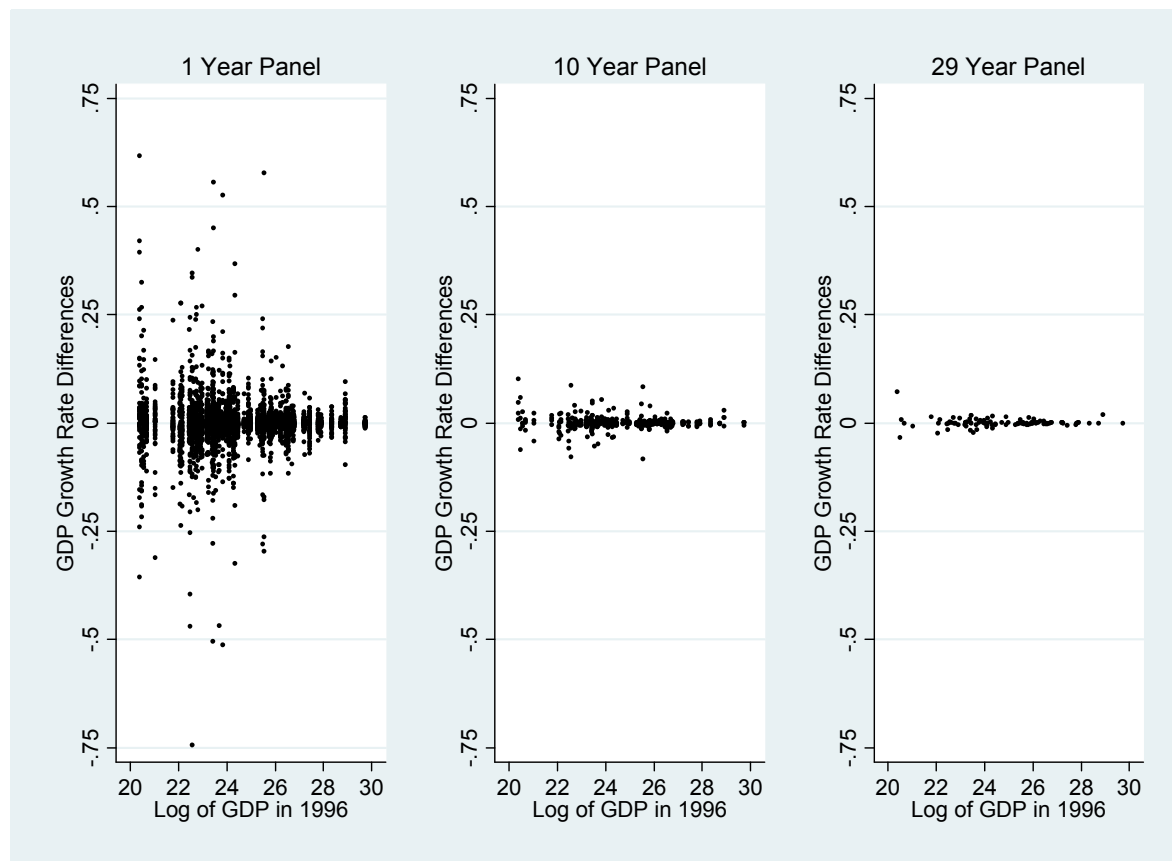
Figure 2. Differences in 29-year average prices between PWT 6.2 and PWT 6.1



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 the 104 countries in the “long-run sample” used in other tables and figures.

Source: Authors’ calculations.

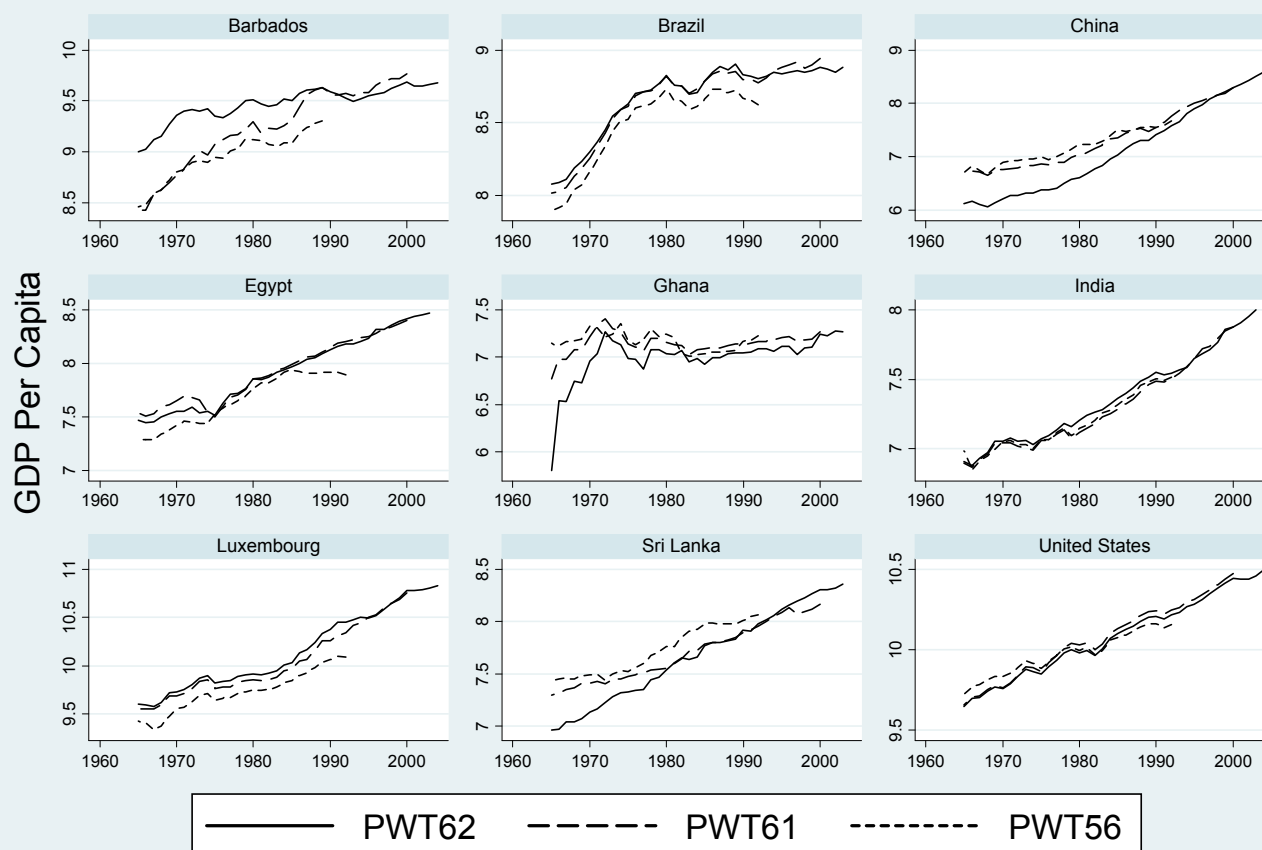
Figure 3. Differences in annual per capita GDP growth between PWT 6.2 and PWT 6.1



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: 3840; avg. difference = .0008; sd of difference = .0539; avg. growth rate = .0186; 10 Yr Panel: # of obs: 384; difference = .0008; sd of difference = .0163; avg. growth rate = .0186; 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” used in other tables and figures. Eight countries were dropped because they did not have data for all four decades. This sample is called the “decades sample.”

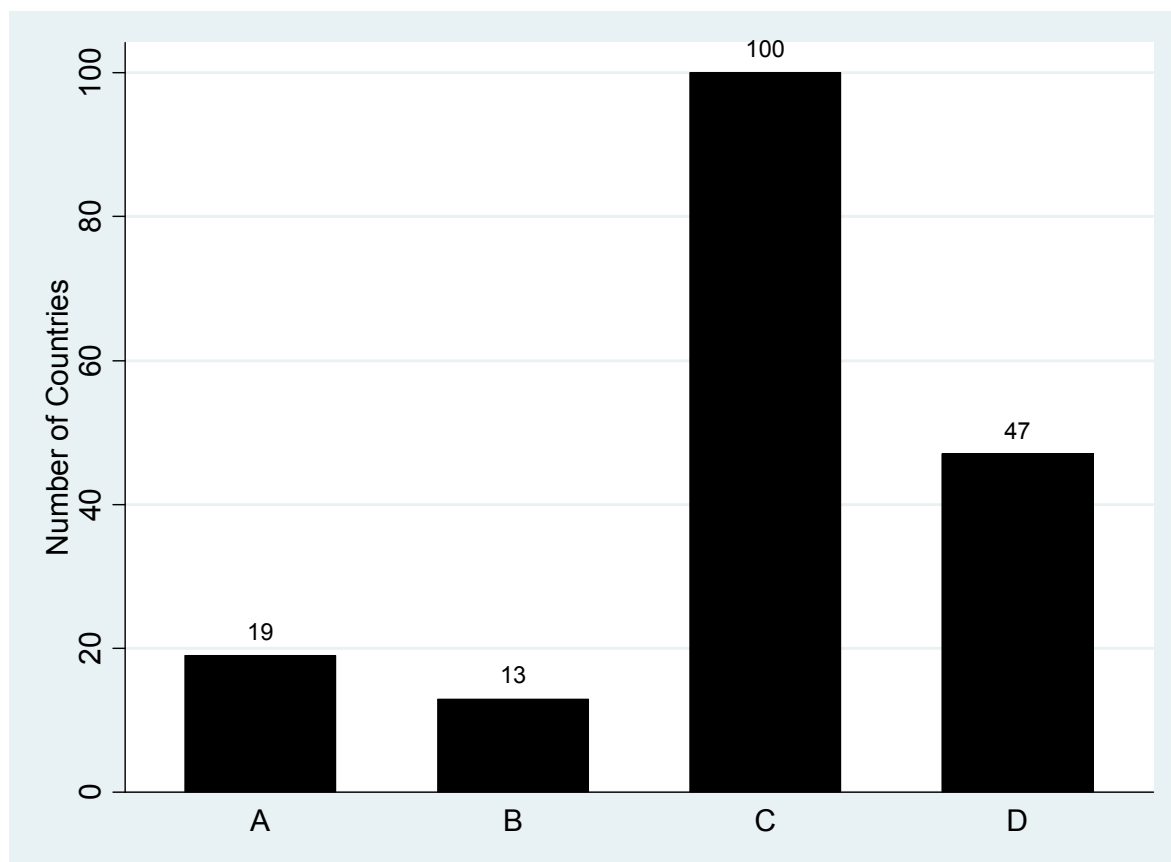
Source: Authors’ calculations.

Figure 4. GDP per capita: PWT 6.2, PWT 6.1, and PWT 5.6



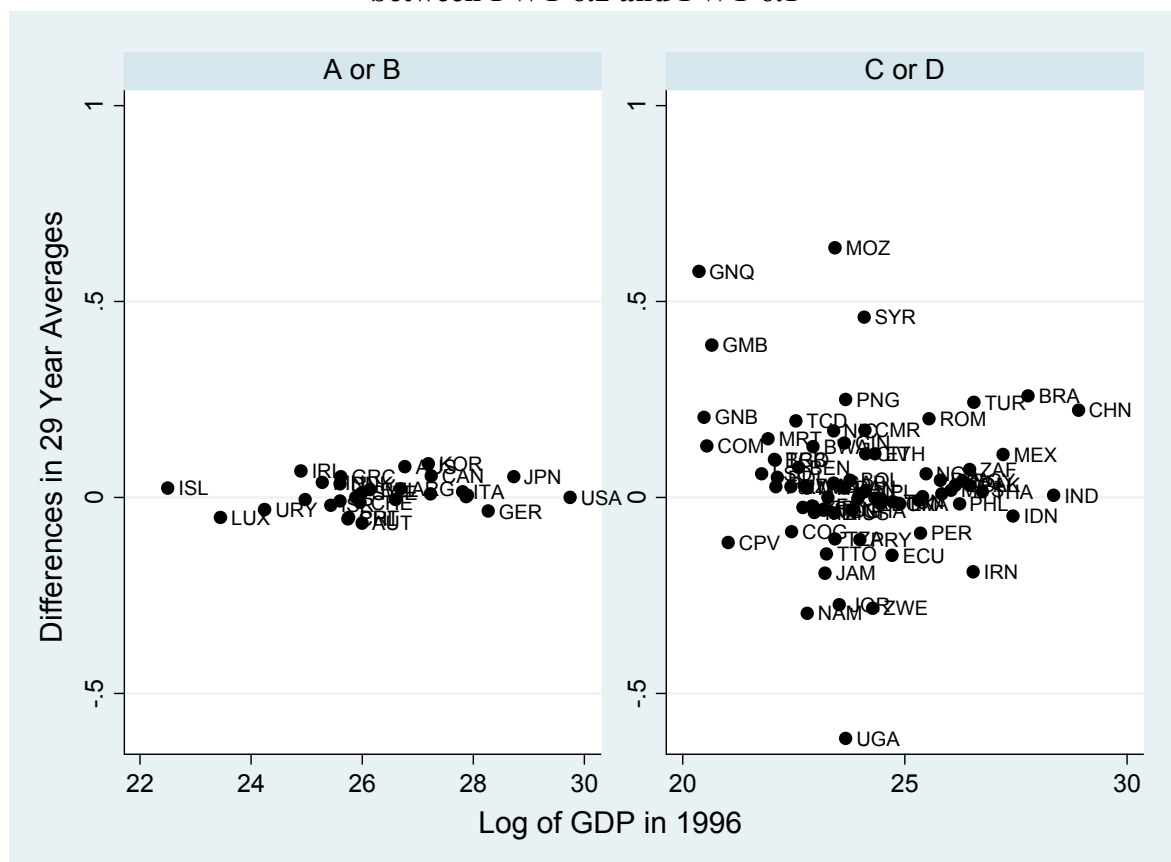
Notes: Per-capita GDP is measured in 2000 international dollars. Because PPPs in each version of the Penn World Table are relative to USA, and base years in each version of the PWT are different, re-basing requires USA GDP deflators for the base years in each of the PWT versions.

Source: Authors' calculations.

Figure 5. Countries with respective PWT data quality grades

Source: ICP benchmark studies and quality grades used in PWT 6.1.

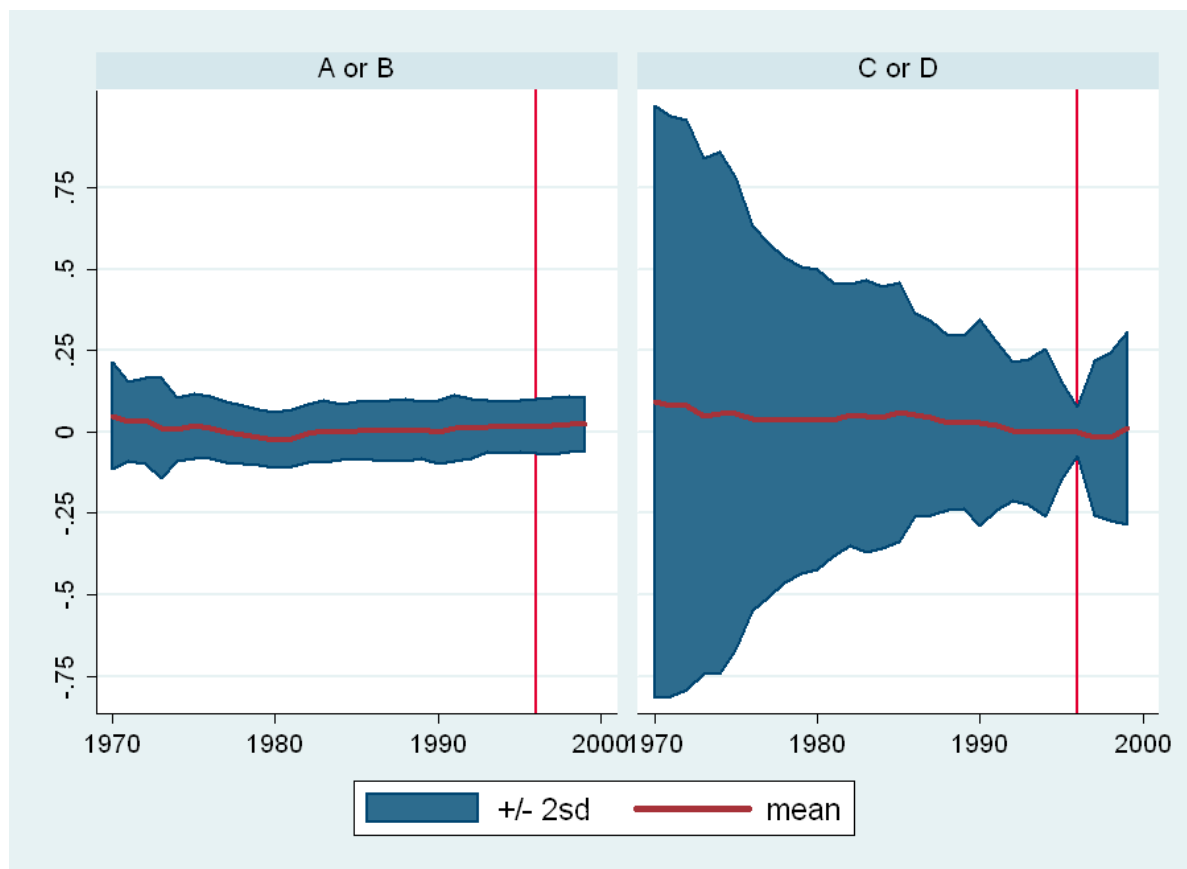
Figure 6. Differences in 29-year average GDP growth by PWT quality grade between PWT 6.2 and PWT 6.1



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 the 104 countries in the “long-run sample” used in other tables and figures. 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.

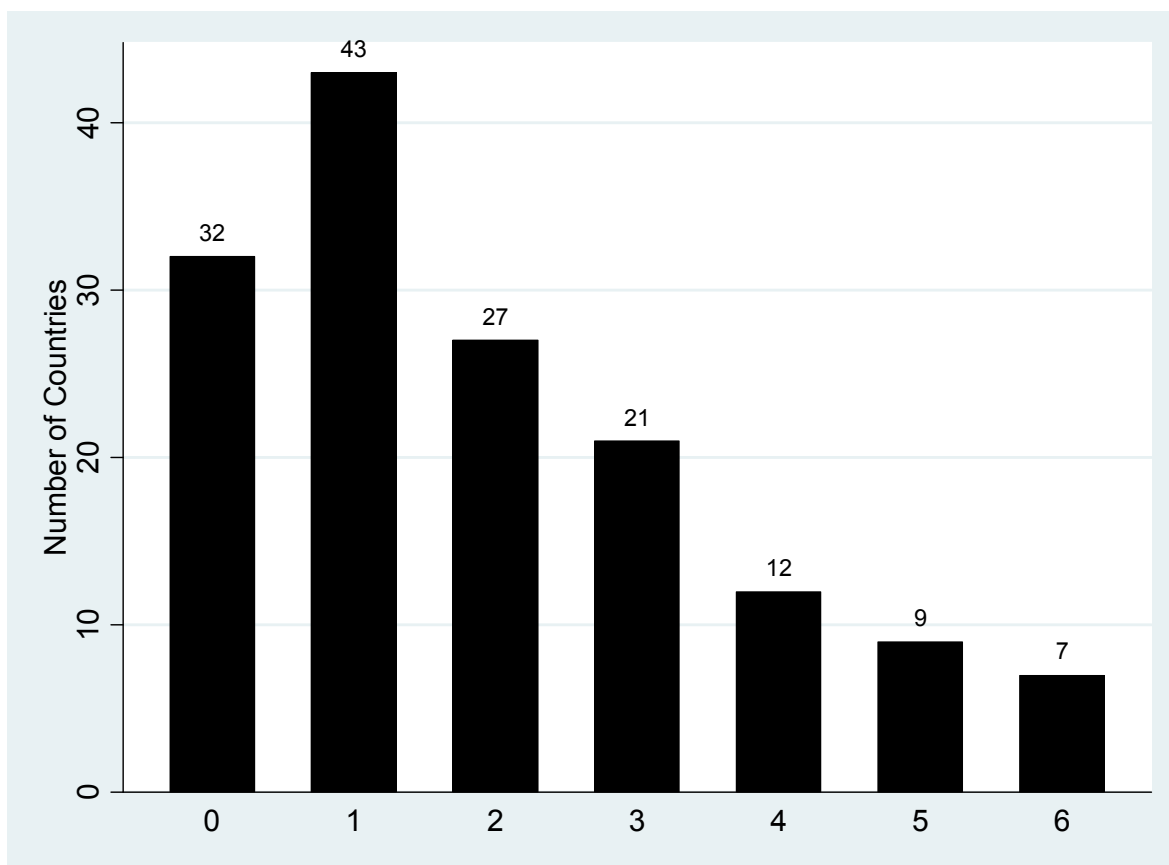
Source: Authors’ calculations.

Figure 7. Evolution of differences in levels of prices across time between PWT 6.2 and PWT 6.1



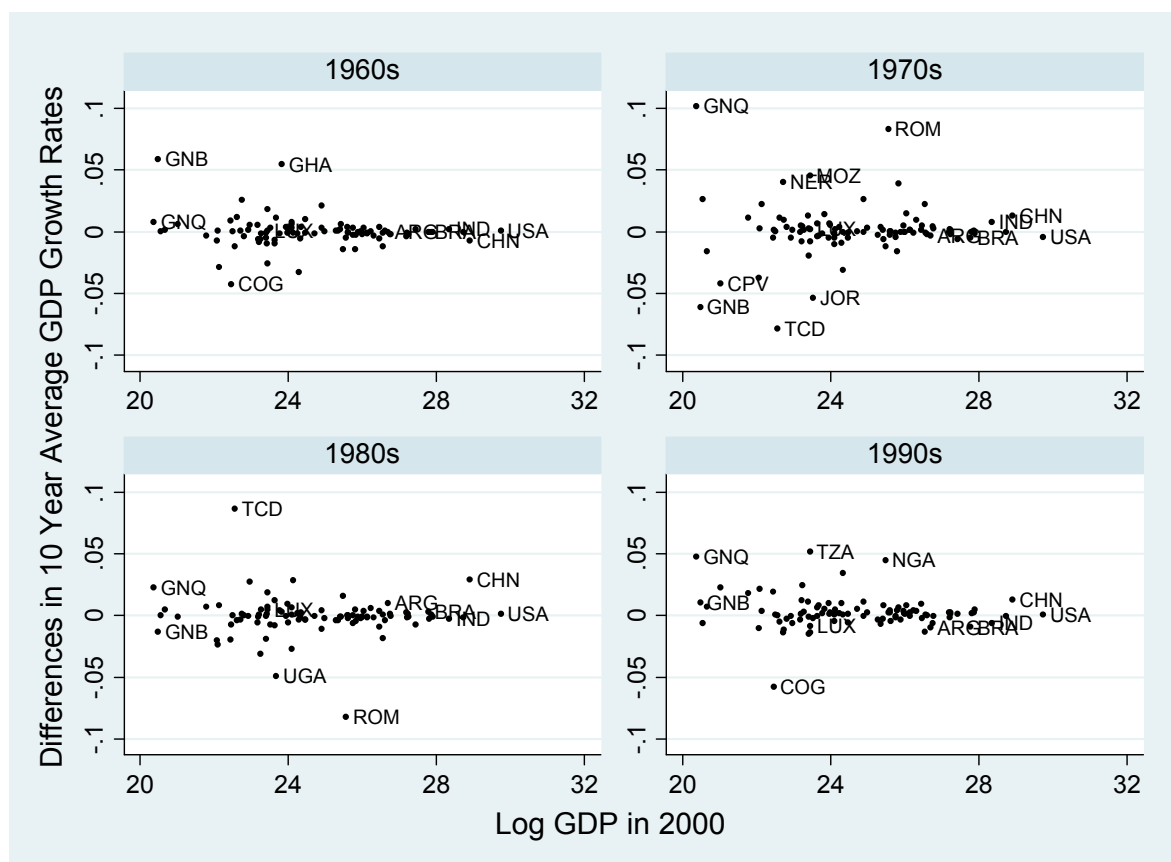
Notes: Prices is the Price (P) series. Differences in levels between the two versions of PWT are calculated as the log level of P from PWT 6.2 minus the log level of P from PWT 6.1. Each mean and standard deviation is computed across countries for a given year. Sample of countries for each year includes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

Source: Authors' calculations.

Appendix Figure 1. Countries with respective number of ICP benchmark studies

Source: ICP benchmark studies used in PWT 6.1.

Appendix Figure 2. 10-year average per capita GDP growth by decade between PWT 6.2 and PWT 6.1



Notes: Ten-year average annual per capita GDP growth rates are computed using the RGDPC series for the period 1960–2000. Average annual growth rates are calculated as $[log(RGDPC_{t+10}) - log(RGDPC_t)]/10$. 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. 1960s: avg. difference = .0001; sd of difference = .0123; avg. growth rate = .0279. 1970s: avg. difference = .0012; sd of difference = .0217; avg. growth rate = .0212. 1980s: avg. difference = -.0008; sd of difference = .0163; avg. growth rate = .0084. 1990s: avg. difference = .0028; sd of difference = .0130; avg. growth rate = .0168. The sample of countries is constant across decades, and consists of 96 of the 104 countries in the “long-run sample” used in other tables and figures. Eight countries were dropped because they did not have data for all four decades. This sample is called the “decades sample.”

Source: Authors’ calculations.