Price Levels and Economic Growth
Making Sense of the PPP Changes between ICP Rounds

Martin Ravallion

The World Bank
Development Research Group
Director's office
March 2010
Abstract

To the surprise of many observers, the 2005 International Comparison Program (ICP) found substantially higher purchasing power parity (PPP) rates, relative to market exchange rates, in most developing countries. For example, China’s price level index—the ratio of its PPP to its exchange rate—doubled between the 1993 and 2005 rounds of the ICP. The paper tries to explain the observed changes in PPPs. Consistently with the Balassa-Samuelson model, evidence is found of a “dynamic Penn effect,” whereby more rapidly growing economies experience steeper increases in their price level index.

This effect has been even stronger for initially poorer countries. Thus the widely-observed static (cross-sectional) Penn effect has been attenuated over time. On also taking account of exchange rate changes and prior participation in the ICP’s price surveys, 99 percent of the variance in the observed changes in PPPs is explicable. Using a nested test, the World Bank’s longstanding method of extrapolating PPPs between ICP rounds using inflation rates alone is out performed by the model proposed in this paper.

This paper—a product of the Director’s office, Development Research Group—is part of a larger effort in the department to understand and improve current macroeconomic data. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at mrvallion@worldbank.org.
Price Levels and Economic Growth:
Making Sense of the PPP Changes between ICP Rounds

Martin Ravallion¹

Development Research Group,
World Bank, 1818 H Street NW, Washington DC
mravallion@worldbank.org

Keywords: Purchasing power parity; Balassa-Samuelson; economic growth, China

JEL: E31, O47

¹ These are the author’s views and should not be attributed to the World Bank or any affiliated organization. The author is grateful to Eric Swanson for comments on an earlier draft.
1. Introduction

Purchasing Power Parity exchange rates (“PPPs” for short) have been mainly derived from the International Comparison Program (ICP), which collects data on prices across countries. The ICP started in 1968 with surveys for just 10 countries. The latest ICP for 2005—thought to be the largest international statistical operation ever—collected prices for a great many goods and services, grouped under 155 “basic headings” (corresponding to the expenditure categories in the national accounts) for each of 146 countries in six regions (Africa, Asia-Pacific, Commonwealth of Independent States, South America, Western Asia and Eurosat-OECD). Region-specific product lists were developed and the regional PPPs were linked through a common set of global prices. The 2005 ICP’s governance structure entailed that each of the six regional ICP offices worked closely with government statistics offices in each country, while the World Bank provided global management and estimated the final PPPs. World Bank (2008a) provides estimates of the PPP for GDP and its main components for 2005. World Bank (2008b) compares the results to those based on the main prior ICP round, for 1993.2

Dramatic revisions to past PPPs are implied by the results of the 2005 ICP. The estimates of GDP (at PPP) were revised downwards for most developing countries (World Bank, 2008b). The new PPP for China attracted much attention, given that it implies that the country’s GDP per capita at PPP for 2005 is 40% lower than we thought, at $4,091 rather than the prior estimate for the same year of $6,760 (World Bank, 2008b). China’s price level index—the PPP for GDP divided by exchange rate (US=100%)—went from 19% in 1993 to 42% in 2005.3 However, this was also the first time that China had officially participated in the ICP; priors had been based on an estimate of the country’s PPP for 1993 that was not based on a 1993 price survey, but rather was an updated version of an older PPP for China from non-ICP price data.4

Some observers have questioned whether China’s new PPP is credible. Bhalla (2008) argues that, when combined with the official growth rates, the new PPP implies that China was too poor to be believed in (say) 1950; in Bhalla’s words, the World Bank’s numbers imply that

---

2 The 1993 ICP rounds was also the first time the Bank had estimated global PPPs, though Penn World Tables had done so (using somewhat different methods) for each of the prior ICP rounds.

3 While the term “price level index” is widely used in the literature, it is potentially confusing, since the index is a relative price—the ratio of two nominal prices. But I will stick to common usage.

4 More precisely, the previous PPPs were derived using a bilateral comparison of 1986 prices between the United States and China as documented in Ruoen and Kai (1995).
“most Asians (were) dead in 1950.” Maddison and Wu (2008) and Deaton and Heston (2010) raise similar objections, leading Maddison and Wu to claim that the new PPP for China is “weird” and “implausible.”

It is far from clear whether these extrapolations back in time constitute a valid basis for validating the new PPP for China. Ravallion (2010) points out that the new PPP only implies that China was as poor (in terms of GDP per capita) in 1950 as the Democratic Republic of the Congo is today, and that about 400 million people in the world (40 million of them in China) currently live below that income level; they survive, albeit at very meager levels of consumption. Nonetheless, when one looks closely at how China’s PPP was estimated there are reasons to suspect that it is too high. Chen and Ravallion (2010) point to likely sampling biases in the ICP’s price surveys for China that would lead to an overestimation of the level of prices. However, their proposed correction still implies a large increase in China’s price level between 1993 and 2005.5

The interpretation of these large data revisions is clouded by the fact that enumerable changes in data and methods were introduced in the 2005 ICP round, as described in World Bank (2008a,b) and Deaton and Heston (2010). A potentially important difference is that (compared to the 1993 and prior ICP rounds) stricter quality standards were used in the 2005 price surveys, to assure that the ICP was obtaining prices for internationally comparable commodities. This is important given that one expects that lower quality goods are consumed in poorer countries, creating a risk that (without strict standards in defining the products to be priced) one will underestimate the cost of living in poor countries by confusing quality differences with price differences. With better funding of the ICP in 2005, clearer product descriptions were developed.

Since the ICP began, these changes between ICP rounds have cast doubt on the inter-temporal comparability of the resulting PPPs. Indeed, a common practice in data compilations—both by Penn World Tables (PWT) and the World Bank’s World Development Indicators (World Bank, 2009, for example)—has been to let the national price data largely or fully override the ICP data for inter-temporal comparisons. In other words, the PPP conversion is only done at the relevant ICP benchmark year, with national price data used for inter-temporal comparisons. It has been argued that this is the most reasonable position to take, given the changes in

5 The ICP aims to collect prices from a representative sample of outlets in each country. However, this was not possible in China and the ICP only covered 11 cities.
methodology between ICP rounds (as argued by Dalgaard and Sørensen, 2002, and World Bank, 2008b).

Theoretical arguments have been made both for and against this practice. Nuxoll (1994) argues that the real growth rates measured in domestic prices better reflect the trade-offs facing decision makers at the country level, and thus have a firmer foundation in the economic theory of index numbers. Of course, this means that the economic aggregates may well lose purchasing power comparability as one goes further back in time from the ICP benchmark year. In the context of studies of economic growth using PWT, Johnson et al. (2009) argue instead that comparisons should only be made between ICP rounds, since only then can one be sure that the economic aggregates are consistently evaluated purchasing power parity.

All this begs a neglected question in the literature: are the changes in PPPs between ICP rounds explicable in terms of observable economic and statistical factors? Is China’s new PPP (for example) really so “weird”?

This paper tries to answer these questions by comparing the latest (2005) PPPs with those for 1993 and 1985. Unlike most past empirical studies of PPPs, which have focused on the cross-sectional differences in price levels, this paper is concerned with explaining the observed changes over time in PPPs. Changes in exchange rates are highly correlated with the observed changes in PPPs (though one should be cautious in giving that correlation a casual interpretation). However, the focus here is on the price level index. For the same reason that one uses PPPs rather than exchange rates for international comparisons, it can be hypothesized here that the PPP will tend to rise relative to the exchange rates in a growing economy. In the models of Balassa (1964) and Samuelson (1964) this happens if economic growth comes with higher labor productivity in the traded-goods sector (relative to non-traded goods). This can be thought of as a dynamic Penn effect (DPE), corresponding to the widely-observed static Penn effect in which the price level index tends to be higher in richer countries. 6 Whether one would see a DPE in a growing developing country is a moot point. It can be argued that such economies are characterized by factor-market imperfections and surplus labor, dulling the Balassa-Samuelson mechanism that generates the Penn effect. In addition to the possibility of a DPE, one can expect that there will be measurement errors confounding the PPP comparisons.

6 The term “Penn effect” appears to be due to Samuelson (1994) and stems from Penn World Tables (PWT) (Summers and Heston, 1991), which provided the price level indices across countries that were most widely used to establish this effect empirically.
A better understanding of how PPPs evolve over time might also throw useful light on how best to update PPPs between the ICP’s benchmark years. There are two distinct sources of missing data in estimating PPPs; the first is that some countries chose not to participate in the ICP’s price surveys for a given benchmark year, while the second is that there are long gaps between those benchmark years. Existing data compilations (such as the World Bank’s *World Development Indicators*) use extrapolations based on non-ICP data to fill in these missing data. However, very different types of data are used for the two sources of missing data. The extrapolations to deal with the first source are based on GDP per capita at market exchange rates, exploiting the static Penn effect. By contrast, in addressing the second problem, the PPPs for non-benchmark years are estimated by re-scaling the PPP from the most recent ICP round according to the inflation rate (GDP deflator for the GDP PPP, and Consumer Price Index for the consumption PPP) in the country in question relative to the US inflation rate. The reliability of this method is unclear. While in theory, a suitable inter-temporal price index could deliver reliable extrapolations, it is far from obvious how well the methods used in practice perform, notably in reflecting the changes in the relative price of non-traded goods in growing economies. However, as this paper shows, there is another possible method for the inter-temporal extrapolations, consistent with how the first problem of missing data is dealt with. If the relationship between price levels and mean income—the static Penn effect that is used to deal with the cross-sectional missing data—also holds over time then this could be exploited in the dynamic extrapolation methods between ICP rounds, thus offering scope for reducing the need for the large data revisions often implied by each new round of ICP data.

To see how well the measured changes over time in price levels can be explained, I have assembled a data file of the price level index (PPP for GDP over exchange rate) and GDP per capita at PPP for the 2005 and 1993 ICP rounds for all countries (developed or developing). These are the World Bank’s estimates, rather than Penn World Tables (PWT). I will also study the changes in PPPs between 1985 (using PWT) and 2005.

The following section summarizes the arguments as to why one might expect the PPP changes to be predictable. Section 3 describes the empirical models to be estimated while section

---

7 Neither the underlying prices nor the aggregation methods are typically the same between the national deflators and the PPPs constructed by the ICP.

8 There are methodological differences between the World Bank’s PPPs and those in PWT; see Deaton and Heston (2010) for a useful overview of those differences.
4 presents the results. Section 5 tests for whether the 2005 PPP for China—the country that has clearly received the most attention in the debates surrounding the release of the results of the 2005 ICP—is consistent with the pattern seen across countries. Section 6 concludes by summarizing the paper’s findings and drawing some implications for future analyses of ICP data.

2. What might explain the price level changes?

The focus here is on the changes in PPPs at the country level. Some of the country-specific factors that one might expect to influence price levels can be treated as largely time-invariant between ICP rounds. For example, Clague (1985) shows that natural resource endowments will influence the price level index at a given level of GDP per capita. However, such endowments can be treated as country-level fixed effects for the present purposes. By focusing instead on the changes in PPPs, the following analysis will eliminate the influence of all additively-separable error components stemming from such country-specific factors.

An important clue to why PPPs change over time can be found in the very same reason PPPs were developed. It has long been recognized that international comparisons of GDP at exchange rates are deceptive about the differences in purchasing power, given that some commodities are not internationally traded, notably most services. Without trade, there is no mechanism for assuring price parity across borders. The most common economic explanation of why the PPP would differ systematically from the nominal exchange rate is the Balassa-Samuelson model (outlined independently by Balassa, 1964, and Samuelson, 1964).\(^9\) This assumes a competitive market economy in which all factors of production are fully employed and are freely mobile between the traded and non-traded-goods sectors. The relative price of non-traded goods is then given by the labor productivity differential between traded and non-traded goods. To see this more formally, let \(MP_T\) denote the marginal (physical) product of labor in the traded goods sector and let \(MP_N\) denote the corresponding marginal product in the production of non-traded goods. Also let \(P_T\) and \(P_N\) denote the prices of traded and non-traded goods while \(W_T\) and \(W_N\) are the corresponding wage rates. Under standard assumptions (including competitive, profit-maximizing, producers) we have \(W_T=P_T MP_T\) and \(W_N=P_N MP_N\).

With perfect labor mobility \((W_T=W_N)\), the key relationship generating the Balassa-Samuelson effect is then immediate, namely that \(P_N/P_T=MP_T/MP_N\).

---

\(^9\) An alternative explanation was proposed by Bhagwati (1984) based on factor endowments, leading (labor-intensive) services to be cheaper in poor countries.
In using the Balassa-Samuelson model to explain why PPPs tend to be lower (relative to market exchange rates) in poorer countries, it is assumed that the more developed the country the higher its labor productivity in traded goods, but that productivity for non-traded goods does not vary systematically with level of development. A higher marginal product of labor in traded goods production comes with a higher wage rate, which is also binding on the non-traded goods sector (given that labor is freely mobile), implying a higher price of non-traded goods in more developed countries and thus a higher overall price level. By the same reasoning, low real wages in poor countries entail that non-traded goods tend to be cheaper. The ratio of the purchasing power parity rate to the market exchange rate will thus be an increasing function of income.

This argument helped reinforce the (considerable) international statistical effort that has gone into the development of purchasing power parity exchange rates, led by the *International Comparison Program* (although PPPs existed before Balassa-Samuelson). The PPP rate expresses a currency’s value in terms of its purchasing power over commodities, both traded and non-traded, relative to the numeraire currency (almost invariably the S$US). The PPP is based on the prices actually paid for goods; the exchange rate does not directly enter into its calculation (though of course it will matter indirectly, via domestic prices).

The Balassa-Samuelson model offers a theoretical explanation for the Penn effect. Balassa (1964) found evidence that richer countries tend to have higher price levels in data for 12 countries. Since then most empirical tests of the Penn effect have used cross-sectional data from the ICP. Every round of the ICP appears to have confirmed the Penn effect (Summers and Heston, 1991; Heston, Nuxoll and Summers, 1994; Rogoff, 1996; World Bank, 2008a; Deaton and Heston, 2010). Based on such evidence, the Wikipedia entry on the Penn effect describes it as “a consistent econometric result for at least fifty years” ([http://en.wikipedia.org/wiki/Penn_effect](http://en.wikipedia.org/wiki/Penn_effect)).

The Balassa-Samuelson effect should hold over time as well as across countries, as long as the data are in reasonable accord with the assumptions of the model. As a poor country develops, its productivity in the traded goods sector will rise, as will the real wage rate, and so its PPP will move closer to its exchange rate. There is some supportive evidence in time series data for specific developed countries, notably Japan, but not others (Rogoff, 1996). Past tests of
whether the implications of the Balassa-Samuelson model (including the Penn effect) hold over time have been largely confined to developed countries.\textsuperscript{10}

Arguments can be made for and against the Balassa-Samuelson assumptions. A key assumption is that richer countries have higher relative productivity in traded goods. Balassa (1964) presented (influential) evidence supporting that assumption. But technology has changed considerably since 1964, entailing greater potential for productivity growth in the services sector. Take, for example, India’s booming business services sector. This sector has seen very high growth since the early 1990s, facilitated by the availability of skilled labor and changes in information technology.\textsuperscript{11} Superficially this does not sound much like the Balassa-Samuelson model. However, it should be noted that this change has also come with a transformation of many business services into internationally traded commodities. So it can be argued that India’s rising productivity in services is in fact consistent with Balassa-Samuelson. By this view, it is the presumption that services are non-traded that is now questionable, given technological change.

However, even if growth does come with rising productivity for traded goods, the way labor markets work in reality may not pass this effect onto wage rates in the non-traded goods sector. This could happen if there are impediments to labor mobility—“labor-market segmentation.” For example, labor hiring in the traded-goods sector may be subject to (explicit or implicit) contracts favoring incumbents, leaving the services sector as a residual employer. Or there may be specialized skill requirements, which effectively restrict entry to the traded-goods sector in poor countries with limited human capital. A wedge between wage rates in the two sectors could also arise if the traded goods sector is the “formal” sector, which is taxed, while services are informal, and un-taxed. In these circumstances we may find a persistent wage gap (with $W_T > W_N$), creating a potential disconnect between relative prices and relative labor productivities between the two sectors, thus breaking the Balassa-Samuelson effect. Whether that actually happens depends on how the relative wage rate ($W_T/W_N$) is in fact determined. However, the key point is that with market imperfections it is an open question whether the PPP will start to approach the exchange rate in initially poor but now growing developing economies, or whether it continues to lag.

\textsuperscript{10} The only exception I know of is Choudhri and Khan (2005) who find evidence consistent with Balassa-Samuelson effects in panel data for 16 developing countries.
\textsuperscript{11} See Kotwal et al. (2009) for evidence on this point.
Clearly there are also measurement errors in the PPPs. Much of this is essentially unobservable. However, one observable source of error in the comparison of PPPs over time is that not all of the countries that participated in the 2005 round had participated in 1993. (The introduction noted the example of China.) Most of the PPPs for these non-benchmark countries in 1993 were estimated econometrically by the ICP team using regressors observed for both sets of countries. Any bias in those estimates will be reflected in the subsequent changes observed when the country participates properly in the ICP’s price surveys. This too could dull the effect of economic growth on price levels.

3. **Modeling changes in price levels**

Some notation will be helpful. Let $PPP_{ri}$ denote the PPP rate for country $i$ in year $r$ using ICP round $r$ and let $E_{ri}$ be the corresponding exchange rate. By definition, the price level index is $P_{ri} = PPP_{ri} / E_{ri}$. Also let $Y_{ri}$ denote GDP per capita at the market exchange rate, while $Y'^{PPP}_{ri}$ is GDP per capita at PPP. Thus $Y_{ri} = GDP_{ri} / E_{ri}$ where $GDP_{ri}$ is GDP in local currency units and $Y'^{PPP}_{ri} = GDP_{ri} / PPP_{ri}$.

The basic model for changes in the price level index is as follows:

$$\ln(P_{05r} / P_{93r}) = \alpha + \beta \ln(Y_{05r} / Y_{93r}) + \epsilon_{ri}$$

(1)

Five remarks will help motivate and interpret this model. First, this equation can be interpreted as the time-differenced version of the widely used double-log model found in the cross-country literature on the Penn effect, incorporating a year effect but common slope (though this will be relaxed later). Second, unlike the cross-sectional specification, the parameter estimates in equation (1) will be robust to any (time-invariant) country characteristics that jointly influence the level of prices and GDP. Third, if $\beta > 0$ then there is evidence of a DPE. Fourth, if $\alpha > 0$ ($<0$) then the 2005 ICP schedule of price levels is higher (lower) than that for 1993 at any given

---

12 See World Bank (2008a). The log GDP per capita in 1993 PPP was regressed on log GDP per capita at exchange rates and the log of the secondary school enrollment rate This is equivalent to regressing the log of the price level index on these same two variables (though with different parameters of course). Note that China’s PPP for 1993 was estimated by a different method.

13 The most common specification for the Penn effect in the literature expresses the log of the price level index as a linear function of the log of GDP per capita at exchange rates. Note that this is equivalent to the alternative specification sometimes found in which the log of GDP at PPP is a linear function of the log of GDP at exchange rates; the slope parameters in these two specifications sum to unity.
level of GDP per capita. Finally, it can be noted that equation (1) can be written equivalently in terms of GDP at PPP, giving:

$$\ln(P_{05i} / P_{93i}) = \left( \frac{\alpha}{1 - \beta} \right) + \left( \frac{\beta}{1 - \beta} \right) \ln(Y_{05i}^{PPP} / Y_{93i}^{PPP}) + \frac{e_i}{1 - \beta}$$

(2)

The basic specification in (1) will be augmented in three ways. First, one can think of equations (1) and (2) as restricted forms of the following equation:

$$\ln(P_{05i} / P_{93i}) = \alpha + \beta \ln(Y_{05i} / Y_{93i}) + \gamma \ln(E_{05i} / E_{93i}) + e_i$$

(3)

It is of interest to see if the restriction that $\gamma = 1$ cannot be rejected. One can, however, question a causal interpretation of equation (3); indeed, the “law of one price” implies the market exchange rate is determined by the PPP, although the fact that price level indices are less than unity for many countries discredits this as a model of exchange rate determination, at least in the short term.\footnote{Rogoff (1996) reviews the literature.} Nonetheless, it would clearly be worrying if one could reject the null that $\gamma = 1$.

Second, I will allow the parameters to vary according to differences in initial GDP per capita and whether the 1993 PPP was derived from actual price surveys—in which case we refer to the country as a 1993 “benchmark country” (following common practice).

The third variation entails testing a nested model encompassing the above specification and the inflation-adjustment method used by the World Bank’s World Development Indicators to update PPPs over time between ICP rounds. By this method, the extrapolated PPP for GDP for date $t$ ($>1993$), using the 1993 ICP round (say) as the benchmark, is given by

$$\hat{PPP}_t = PPP_{93i} (DEF_t / DEF_{93i}) / (DEF_{tUS} / DEF_{93US})$$

where $DEF_i$ is the GDP deflator (or CPI when updating the consumption PPP) for country $i$ at date $t$ (where $i=US$ denotes the deflator for the USA). The encompassing test entails adding a term in $\ln(DEF_{05i} / DEF_{93i})$ to equation (3). If one cannot reject the joint null that the coefficient on this extra variable is unity, while $\beta = \gamma = 0$, then the inflation-adjustment method is supported by the data.

4. **Results**

Figure 1 gives the empirical density functions, indicating that the price level index is below 100% for the bulk of the data in both years. Nonetheless, the index rose between 1993 and 2005 for 74% of countries. And it rose by 10 percentage points or more in almost half the
countries (44% to be precise). For 2005, the (unweighted) mean price index across the 133 countries that participated in the 1993 ICP was 59% (median 49%) as compared to 53% in 1993 (median 41%). The mean $\ln(P_{05i}/P_{93i})$ is 0.162, with a standard error of 0.026 (n=133).

So an upward revision to the price level index is generally indicated. However, the bulk of this increase was for countries with initially low price levels. Indeed, the cumulative distribution functions implied by the density functions in Figure 1 are virtually indistinguishable for price level indices above 60%.

Regressing $\ln(P_{ri})$ on $\ln(Y_{ri})$, the data confirm the static (cross-sectional) Penn effect within each round, as widely reported in the literature; the regression coefficient of $\ln(P_{05i})$ on $\ln(Y_{05i})$ is 0.216, with a White standard error of 0.013 (n=144);\footnote{Using GDP at PPP instead, the regression coefficient of $\ln(P_{05i})$ on $\ln(Y_{05i}^{PPP})$ is 0.237 (s.e.=0.020), while for the 1993 ICP it is 0.343 (s.e.=0.027). Note that White standard errors are used when relevant in this paper.} Figure 2 plots the data from the 2005 ICP. For the 1993 round, the regression coefficient is 0.293 (s.e.=0.012; n=134). So there is evidence that the Penn effect has become weaker over time, consistent with the fact that the proportionate increases tended to be larger in initially poorer countries, as can be seen in Figure 3. (The slope of the regression line in Figure 3 is -0.059 with a standard error of 0.014.)

Turning to the changes over time, let us begin by testing the homogeneity restriction that $\gamma = 1$. The estimate of the unrestricted model (equation 3) is:

$$\ln(PPP_{05i} / PPP_{93i}) = -0.029 + 0.290 \ln(Y_{05i} / Y_{93i}) + 1.011 (E_{05i} / E_{93i}) + \hat{\epsilon}_i$$

$$R^2 = 0.987; n = 125$$

The restriction clearly performs well. Given that this is a regression for changes rather than levels, it is also notable that almost 99% of the variance is accounted for.

Relative to the above model, the data do not support the inflation-adjustment method for updating PPPs. When added to equation (4), the coefficient on $\ln(DEF_{05i} / DEF_{93i})$ is 0.094, with a standard error of 0.102; other coefficients and their standard errors change little. The inflation-adjustment method does not perform badly on its own; the regression coefficient of $\ln(PPP_{05i} / PPP_{93i})$ on $\ln(DEF_{05i} / DEF_{93i})$ is 0.981 (s.e.=0.044), with $R^2=0.958$. However, the specification in (4) clearly outperforms this method in the nested test.
Imposing the restriction that $\gamma = 1$, the regression coefficient of $\ln(P_{05} / P_{93})$ on $\ln(Y_{05} / Y_{93})$ is $\hat{\beta} = 0.283$ (s.e. = 0.054; n = 132) with $\hat{\alpha} = -0.019$ (s.e. = 0.043) and $R^2 = 0.212$. This matches quite well the cross-sectional estimate of the coefficient for the Penn effect, suggesting that latent country characteristics are not an important source of bias in past tests for the Penn effect using cross-sectional data.

Figure 4 plots the relationship between changes in price levels and growth rates. As is clear from Figure 4, the price level does not start to rise until there is sufficient growth. The expected change in $\ln(P_{05} / P_{93})$ is zero when $\ln(Y_{05} / Y_{93}) = 0.066$, although the latter number is not significantly different from zero (s.e. = 0.140). Thus one cannot reject the null hypothesis that the price level index conditional on GDP per capita is unchanged between the 1993 and 2005 ICP rounds.

Recall that this sample includes both developed and developing countries. The data suggest that the DPE is stronger in poorer countries, as can be seen from the following regression indicating a significant negative interaction effect:

$$\ln(P_{05} / P_{93}) = 0.008 + (0.604 - 0.049 \ln Y_{93}) \ln(Y_{05} / Y_{93}) + \hat{\epsilon}_i \quad R^2 = 0.246; n = 132 \quad (5)$$

I will return to this negative interaction effect.

Strikingly, the DPE is not in evidence if one uses growth rates in GDP at PPP instead (equation 2). Indeed, the regression coefficient of $\ln(P_{05} / P_{93})$ on $\ln(Y_{05}^{PPP} / Y_{93}^{PPP})$ is negative, with $\hat{\beta} = -0.117$, but this is not significantly different from zero (s.e. = 0.073). Yet, the corresponding estimate for $\beta$ of 0.283 using growth rates at exchange rates implies $\hat{\beta} / (1 - \hat{\beta}) = 0.394$. The most likely explanation appears to be that a large endogeneity bias emerges when one tests for the DPE using growth is measured in PPP $s$. Latent factors influencing the PPP will jointly affect both the price level index and the GDP growth rate, though in opposite directions (given that the PPP appears in the numerator of the price index, but the denominator of deflated GDP). Estimating the test equation in the form of (1) avoids this problem. Alternatively, one can use a 2SLS estimate of equation (1), using $\ln(Y_{05} / Y_{93})$ as the instrument for $\ln(Y_{05}^{PPP} / Y_{93}^{PPP})$. Then one exactly retrieves the same estimate of $\beta / (1 - \beta)$, namely 0.394 (s.e. = 0.105), as implied by the OLS estimate of equation (1).
There is also evidence of what can be termed an “ICP participation effect,” whereby the relationship between the price level changes and growth rates differs significantly between the 1993 benchmark \((D_{93i} = 1)\) and non-benchmark \((D_{93i} = 0)\) countries, as is evident in the following regression:

\[
\ln(P_{05i} / P_{93i}) = D_{93i}(-0.192 + 0.433 \ln(Y_{05i} / Y_{93i}))
\]

\[
+ (1 - D_{93i})(0.247 + 0.193 \ln(Y_{05i} / Y_{93i})) + \hat{\epsilon}_i
\]

\[
R^2 = 0.428; n = 132
\]

An F-test rejects the restricted form in which the model is the same when \(D_{93i} = 1\) as \(D_{93i} = 0\); \(F(2,128)=23.516; \text{prob.}<0.001\). Differentiating the DPE between the 1993 benchmark and non-benchmark countries doubles the share of the variance in \(\ln(P_{05i} / P_{93i})\) that is explained.

The DPE is clearly stronger for benchmark countries; Figure 5 plots the relationship for this sub-sample. We now find that the estimated expected value of \(\ln(P_{05i} / P_{93i})\) is positive when \(\ln(Y_{05i} / Y_{93i}) \geq 0.443\), and this switch point is significantly different from zero (s.e.=0.058). Only when the growth rate (annualized log difference) exceeds 3.7% do we find upward pressure on the price level.

So, amongst the 1993 ICP participants one finds that the 2005 price levels are actually lower at given GDP per capita than those of 1993. To put the point another way, these results suggest that it is economic growth in developing countries that explains the upward shift in price levels implied by the 2005 ICP, rather than statistical factors such as the stricter quality standards in the 2005 ICP’s price surveys. The statistical-comparability problem appears to stem largely from the subset of 2005 ICP countries that had not participated in the 1993 ICP round.

The negative interaction effect between the DPE and initial GDP per capita evident in equation (5) turns out to be largely attributable to this ICP participation effect, as can be seen from the following encompassing specification:

\[
\ln(P_{05i} / P_{93i}) = D_{93i}[-0.188 + (0.677 - 0.032 \ln Y_{93i}) \ln(Y_{05i} / Y_{93i})]
\]

\[
+ (1 - D_{93i})(0.266 + (0.549 - 0.063 \ln Y_{93i}) \ln(Y_{05i} / Y_{93i})) + \hat{\epsilon}_i
\]

\[
R^2 = 0.444; n = 132
\]

---

\(^{16}\) As a further check on the restriction that \(\gamma = 1\), on re-estimating in the form of (3), the augmented specification corresponding to (6), gave \(\hat{\gamma} = 1.003\) with s.e.=0.011 with \(R^2=0.991\).
The interaction effect between GDP growth and the initial level of GDP is still evident but it is not significantly different from zero in either segment of the data. The ICP participation effect dominates.

Why is the Penn effect so much weaker for non-benchmark countries? Some explanations can be suggested. First, it might be conjectured that the ICP participation effect reflects the fact that the above model does not include all regressors used to predict the 1993 PPP for the non-benchmark countries. The simplest way to check this is to see if the results change when one controls fully for the econometrically-estimated price level indices for the non-benchmark countries in 1993, by re-estimating (6) in the following form:

\[
\ln\left(\frac{P_{05}}{P_{93}}\right) = D_{93}(-0.192 + 0.433\ln(Y_{05}/Y_{93})) + (1 - D_{93})(2.766 + 0.154\ln Y_{05} - \ln P_{93}) + \hat{\epsilon}_i
\]

(8)

\[R^2 = 0.484; n = 132\]

It remains the case that the Penn effect is weaker for the non-benchmark countries. So this is not the explanation.

Second, it might be conjectured that the ICP participation effect stems from non-linearities in the Penn effect, given that the non-benchmark countries tended to be poorer. However, a significant difference between the benchmark and non-benchmark models was also evident when I interacted the growth rate with the initial level of GDP per capita (and the interaction effects were individually insignificant, as long as one allowed the model to differ between benchmark and non-benchmark countries). The difference also persisted when I allowed for non-linearity in the double log model for the Penn effect, by adding terms in \(\ln(Y_{05})^2 - \ln(Y_{93})^2\).

The mystery remains. Aside from a low-average income, the non-benchmark countries are quite heterogeneous. It appears that all countries were invited to participate in the 1993

---

17 The missing variable is the secondary school enrollment rate.
18 Note that when \(D_{93} = 0, P_{93}\) is predicted based on the observed covariates used by the ICP.
19 The non-benchmark countries in the 1993 ICP had an (unweighted) mean GDP per capita of $2,268, as compared to $6,923 for benchmark countries.
20 The specific countries are Albania, Angola, Bhutan, Bosnia and Herzegovina, Brunei, Burkina Faso, Burundi, Cambodia, Cape Verde, Central African Republic, Chad, China, Colombia, Comoros, Democratic Republic of the Congo, Cote d'Ivoire, Cyprus, Djibouti, Equatorial Guinea, Ethiopia, Gambia,
ICP, but the (unknown) process determining why some countries chose not too was based on variables that are correlated with the strength of DPE. When interpreted in terms of the Balassa-Samuelson model, it appears that the countries that chose not to participate had latent characteristics that made their productivity differences for tradable goods less responsive to the differences in their GDP per capita.

It is unclear what those characteristics might be, but one can speculate on one possible explanation. Some countries undoubtedly chose not to participate in the 1993 ICP for idiosyncratic, possibly political, reasons. But it can be expected that many non-participants lacked the public-institutional capacity for implementing the ICP’s surveys. Furthermore, it can be conjectured that weak statistical capacity is probably correlated with weak institutions more generally, including weak states. Suppose now that institutional capacity is cooperant with labor in the production of traded goods—such that the marginal product of that labor is lower when institutions are weaker. Then we can see that non-participation in the ICP could signify a weaker relationship between GDP and the relative productivity of labor in the traded-goods sector, and (hence) a weaker Penn effect in the data.

The ICP participation effect goes some way toward explaining why the proportionate increases in the price level index tended to be higher in countries with lower GDP per capita in 1993, as evident in Figure 2. Simply adding a control for benchmark countries brings the slope of the regression line in Figure 2 down from -0.059 (s.e.=0.014) to -0.037 (s.e.=0.014). Adding the 1993 GDP to equation (6) gives:

$$\hat{Y}_i = \beta_0 + \beta_1 D_{ipi} + \beta_2 \ln(Y_{ipi}) - \beta_3 \ln(Y_{ipi}) + \epsilon_i$$

(Estimating this regression in the form of equation (3), the coefficient on the log difference in the exchange rate is again not significantly different from unity; $\hat{\gamma} = 0.995$ with s.e.=0.010 and the $R^2$ rises to 0.991.)

Ghana, Guinea-Bissau, India, Israel, Lesotho, Macao China, Macedonia, Malta, Mauritania, Mozambique, Niger, Paraguay, Rwanda, South Africa, Sudan, Togo, and Uganda.

21 The weighted mean coefficient on GDP in 1993 is -0.039 (s.e.=0.002), as compared to a regression coefficient without controls of -0.059 (s.e.=0.014).
It is also of interest to see how well one can explain the changes between the 1985 and 2005 PPPs. Here there are further comparability problems, stemming from the differences between the World Bank’s methods and those of PWT. Even so, the corresponding estimate of equation (3) has similar explanatory power (in obvious notation):

\[
\ln(PPP_{05}/PPP_{85}) = -0.168 \pm 0.417 \ln(Y_{05}/Y_{85}) + 0.991(MER_{05}/MER_{85}) + \hat{\epsilon},
\]

\[R^2 = 0.981; n = 54\]

Imposing the restriction that \(\gamma = 1\) the regression coefficient of \(\ln(P_{05}/P_{85})\) on \(\ln(Y_{05}/Y_{85})\) is \(\hat{\beta} = 0.431\) (s.e.=0.075; \(n=55\); \(R^2=0.488\)) and \(\hat{\alpha} = -0.194\) (s.e.=0.104). So it appears that the DPE is even stronger over 1985-2005, despite the methodological differences between PWT and the World Bank’s methods. Figure 6 plots the data for this longer period (with fewer observations, and confined to 1985 benchmark countries). The switch point (at which \(E\ln(P_{05}/P_{85}) = 0\)) is at \(\ln(P_{05}/P_{85}) = 0.451\) (s.e.=0.168), corresponding to an annual growth rate of 2.3%.

5. China’s controversial PPP revisited

How much of the observed change in China’s price level index implied by the 2005 ICP is accountable to the dynamic Penn effect? If one adds a dummy variable for China to equation (6) or (7) the coefficient is 0.278 (s.e.=0.077), while it is 0.239 (s.e.=0.083) when added to (8). Yet the observed change in the log of China’s price level index is 0.795. So the model can account for 65-70% of the change in China’s price level.

This assessment is affected little by allowing for over-estimation of China’s growth rate. Maddison (2007) has claimed that China’s long-run growth rate is over-estimated by possibly two percentage points per year (though also see Holz’s, 2006, comments on Maddison’s assumptions). However, even cutting two percentage points off China’s annual growth rate and re-estimating the regressions, the China coefficient is still 0.320 (s.e.=0.066) in equation (6) and 0.282 (s.e.=0.071) in equation (8).

The rest of the change in China’s price level could well stem from the sampling bias in the 2005 ICP’s price surveys for China, as noted in the introduction. The correction for that bias

\[22\text{ Given that the 2\% is annual, the term in } \ln(Y_{05}/Y_{85}) \text{ for non-benchmark countries was replaced by } \ln(Y_{05}/Y_{85}) - 0.24 \text{China where } \text{China is a dummy variable for China, which also appears as a separate regressor for the purpose of this test.}\]
proposed by Chen and Ravallion (2009), using non-ICP data on rural prices, brought China’s (expenditure-weighted) price index for consumption (rather than GDP as a whole) down from 52% to about 45%, though still considerably higher than the prior estimates of around 25% based on the 1993 ICP (19% for GDP). Assuming a similar correction for the GDP price index, the combined effect with the DPE (given China’s high growth rate), leaves the doubling of China’s price index almost fully explained.

6. Conclusions

The paper finds evidence consistent with the existence of a Balassa-Samuelson effect over time such that the PPP exchange rate starts to rise relative to the market rate in growing economies. The paper finds no evidence that this effect is any weaker in poorer countries—indeed there is evidence of an even stronger effect of economic growth on price levels in initially poorer countries. Thus the widely-observed static “Penn effect” (whereby the price level index is lower in poorer countries) has been attenuated over time.

The higher price levels for developing countries implied by the 2005 ICP are accountable in part to their economic growth. On its own, the dynamic Penn effect accounts for about one fifth of the variance in the proportionate changes in the price levels over 1993-2005, rising to one half over 1985-2005. An augmented version of the basic model allowing for measurement error in the PPP estimates not based on price surveys can explain almost half the variance in the proportionate changes in price levels.

This degree of explanatory power for price level changes certainly does not eliminate concerns about the comparability of PPPs between ICP rounds. The dynamic Penn effect alone still leaves almost 80% of the variance in the proportionate changes in the price level index unexplained, though this drops to 60% or less once one allows for the measurement errors associated with the need to estimate PPPs econometrically for the countries that did not participate in the 1993 ICP. Nor has the paper got far in explaining the systematic effect of such measurement errors; participation in the 1993 ICP is clearly correlated with some omitted variables that matter to the size of the Penn effect. However, the results of this paper do cast doubt on the extreme view of “PPP non-comparability” that has motivated data construction and analysis, in which past ICP rounds are essentially ignored at each update.
The results of this paper do not suggest that China’s new PPP is as “weird” or “implausible” as some observers have claimed. Given China’s high growth rate, it is not too surprising that the country’s price level index rose appreciably between the 1993 and 2005 ICPs. This paper’s calculations suggest that about two thirds of that increase is accountable to the dynamic Penn effect. The bulk of the remainder may well reflect an upward bias in China’s PPP due to the ICP’s weak coverage of China’s rural areas.

A potentially important implication for future data compilations concerns how PPPs are updated between ICP rounds. The World Bank’s current methods for such updating do not allow explicitly for the dynamic Penn effect identified in this paper (yet the extrapolations used to fill in missing PPPs in a given benchmark year are explicitly based on the static Penn effect). The results of this paper suggest an alternative approach in which the dynamic Penn effect would be brought explicitly into the inter-temporal extrapolations for the price-level index, using market exchange rates to back out the implied PPPs for non-benchmark years. For that purpose, it is encouraging that this paper finds that 99% of the variance in PPP changes between ICP rounds can be explained by just two variables: GDP growth rates and changes in nominal exchange rates. This model yields more reliable estimates than the standard inflation-adjustment method used for updating PPPs between ICP rounds.
References


____________, 2009, World Development Indicators, World Bank, Washington DC.
Figure 1: Kernel densities for price level indices in 1993 and 2005

Figure 2: Static Penn Effect, 2005
Figure 3: Larger upward revisions to price levels in initially poorer countries

Figure 4: Dynamic Penn Effect, 1993-2005: full sample
Figure 5: Dynamic Penn Effect, 1993-2005: 1993 benchmark countries only

Figure 6: Dynamic Penn Effect, 1985-2005