International Comparison Program

[01.03]

Calibrating measurement uncertainty in purchasing power parity exchange rates

Angus Deaton

7th Technical Advisory Group Meeting
September 17-18, 2012

Washington DC

Calibrating measurement uncertainty in purchasing power parity exchange rates

Angus Deaton

Research Program in Development Studies

Princeton University

Draft version, April 2012

1. Introduction

As is the case for all price indexes, the raw material of PPP exchange rates is a set of prices and expenditure shares. Relative prices are not the same in all countries so that some averaging is required and even if we know just what sort of average we want, its estimation will be subject to uncertainty that is greater the greater is the variability of relative prices. There is also a choice over aggregation procedures, and there are many different price index formulas that can be used to turn the price and share information into a set of PPPs, one for each country. The multiplicity of indexes is a familiar issue in price indexes—the choice of Paasche, Laspeyres, Fisher, Törnqvist, or some other—but in the PPP case, there is also the choice of the multilateral aggregation formula—such as Gini EKS, Geary-Khamis, weighted CPD, or Iklé. Typically, a particular set of PPPs, from the ICP tables, from the Penn World Table, or from Eurostat, selects an aggregation formula and presents a set of estimates based on that aggregation formula. However, the choice of aggregation formula has some degree of arbitrariness, and when the expenditure patterns and/or the relative prices differ from one country to another, the choice may matter. While there are certainly arguments for and against each aggregation formula, these are not usually decisive, so that any specific choice resolves the inherent uncertainty in a more or less arbitrary way. For example, if we need to calculate a PPP for Canada relative to the US, or a set of PPPs for the countries of the EEC, cases where expenditure patterns are relative prices are quite similar, the choice of aggregation formula matters much less to the results than would be the case if we are calculating

PPPs for the whole world, or even for Africa, or for the countries of the OECD, which include places as diverse as Mexico and Japan.

As will be shown, the choice of aggregation formula among those considered here is not as important in generating uncertainty as the variation that comes from the sampling of goods. Within each country, prices of different commodities vary more or less around the around the PPP average. In comparisons of countries with similar structure and similar relative prices, such as the US and Canada, or France and Belgium, the price relatives of individual goods are relatively narrowly distributed around the PPP. In such cases, the sampling of goods makes little difference, and the average price relative, weighted appropriately to yield the PPP, is relatively precisely determined. But for countries across which relative prices are very different, the sampling of goods matters, and the PPP itself is much more uncertain. This source of variation induces substantial uncertainty into the price indexes.

In this paper, I propose a method for calculating standard errors for PPPs that takes into account the uncertainty that comes from the fact that relative prices and expenditure patterns differ from country to country and I do so for a number of PPP formulas. The basic ideas were first laid out in a working paper by the author and Olivier Dupriez, and in Deaton and Dupriez (2011), they were applied to the calculation of poverty-weighted purchasing power parity exchange rates. The ideas are closely related to, but different from and differently motivated than the procedures in Hill and Timmer (2006).

The paper is laid out as follows. I start, in Section 2, with a simple example that is intended to establish notation and clarify the basic ideas. Section 3 lays out

2

the formulas for bivariate index numbers where each country is compared with the US as base. Section 4 extends the analysis to multilateral indexes. All sections are illustrated using data from ICP 2005.

2. Preliminaries

Consider a "star" world in which we are making direct comparisons between the US and other countries, one at a time. For countries that are "similar" to the US, like Canada or Ireland, there is less uncertainty about relative prices than for countries that are "different," like Tajikistan and Chad. For example, over the 128 basic headings in the 2005 ICP, the standard deviation of the log of the price relatives for Canada relative to the US is 0.25. For China, the standard deviation is 0.77, for India, it is 0.81, and for Tajikistan—which is typically the most extreme case—it is 1.35. Relative prices are much more dissimilar between Tajikistan and the US, or China or India and the US, than between Canada and the US. In the two country context, differences in price relatives will typically show up in differences between the Paasche and Laspeyres price indexes and, indeed, the absolute value of the log of the ratio of the Laspeyres to the Paasche (the "Paasche-Laspeyres spread") has often been used precisely as a measure of dissimilarity in relative prices, see for example Hill (1999).

For reasons that will become apparent, I shall work with the logarithmic form of the ratio so that, for the record and to establish notation, I write for country c and base country 1, here the US,

$$\ln \rho_{1c} = \ln\left[\sum_{i=1}^{n} s_{i1}\left(\frac{p_{ic}}{p_{i1}}\right)\right] + \ln\left[\sum_{i=1}^{n} s_{ic}\left(\frac{p_{i1}}{p_{ic}}\right)\right] \qquad (1)$$

where the first subscript is always the base country for any comparison, ρ is the

Paasche Laspeyres spread, s_{ic} is the share of good i in total expenditure of country c,

p_{ic} is the price of good i in country c, and there are n goods. For total domestic

absorption in ICP 2005, the ratio ρ runs from 1.048 for Ireland and 1.057 for Canada

through 1.61 and 1.66 for India and China to 5.11 and 9.62 for Tajikistan.

Another way approaching price differences is via what is known as the country-product-dummy (CPD) specification. According to this, first introduced by Summers (1973) as an approximation useful for imputing missing values, the logarithm of each price is written as the sum of a commodity term, a country term, and a residual. Hence, we have

$$\ln p_{ic} = a_c + \beta_i + \varepsilon_{ic} \qquad (2)$$

where a_c is a country effect and β_i is a commodity effect and ε_{ic} is a residual. I take

(2) to be exact, and to *define* the parameters a and β as projections of the log prices

on a set of country and commodity dummies. If there were no trade costs and all

goods were tradable and freely traded between countries, so that the law of one

price held, the residual terms in (2) would be zero. The quantities a_c would be the

logarithm of the exchange rate for c relative to the base country 1 (which is the

omitted country in the projection regression) and the quantities β_i the logarithm of

the price of good i relative to a numeraire good. These relative prices are identical in

all countries so that, for example, the price of cows' milk cheese is the same in Paris

and Cambodia once we have deflated by the Cambodian to French exchange rate. When this is true, all reasonable price index formulas will give the same result, and all reasonable PPP formulas will deliver the market exchange rates as PPPs.

When the law of one price fails, the residual terms ε_{ic} measure deviations from the law of one price, or the deviations of actual prices from the prices that would obtain under perfect arbitrage with exchange rates given by PPPs. Given this interpretation, one way of assessing the uncertainty associated with various PPPs is to examine the variation in calculated PPPs that are driven by random draws from the distribution of these deviations. The randomness here is perhaps best thought of as driven by the choice of different goods by the ICP. In what follows, I make these calculations conditional on the expenditure patterns s_{ic}. An obvious alternative would be to think about a model of expenditure determination in which demand patterns are thought of as responding to the deviations in prices, though that would require a commitment to substantive modeling beyond the statistical approach of this paper.

To see how this stochastic approach works in a simple example, consider equation (2) for the US and Canada, and subtract the US equation from the Canadian equation to get

$$\ln\left(\frac{p_{id}}{p_{ic}}\right) = \ln p_{id} - \ln p_{ic} = (a_d - a_c) + (\varepsilon_{id} - \varepsilon_{ic}) \tag{3}$$

where I have used d for Canada and c for the US to emphasize that this works for any pair of countries. The quantity $(a_d - a_c)$ is the logarithm of the Canadian to US price index, and we can estimate it by an OLS regression, or more simply, just by taking the

average of the log price ratios over all *n* goods. This gives us the log of the geomean PPP exchange rate

$$\ln P_{cd}^G = \frac{1}{n}\sum_{i=1}^{n}\ln\frac{p_{id}}{p_{ic}} \tag{4}$$

The standard deviation of the log price ratios, by (3), is

$$s.d.\ln\left(\frac{p_{id}}{p_{ic}}\right) = \sqrt{\sigma_d^2 + \sigma_c^2} \tag{5}$$

where I have assumed for simplicity that there are no covariances across goods or across countries and that the variances are the same for all goods. Equations (4) and (5) imply that the standard error of the log of the geomean PPP rate for the two countries is

$$s.e.\ln P_{cd}^G = \sqrt{\frac{\sigma_d^2 + \sigma_c^2}{n}} \tag{6}$$

Equations (4) and (5) provide a direct measure of relative price variability that can be regarded as an alternative to the Paasche-Laspeyres spread as a measure of price dissimilarity. In fact, the two are quite closely related, as we shall see below.

Figure 1 shows a plot of the log Laspeyres to Paasche ratio on the vertical axis against the variance of the price ratios. In each case, the base country is the US, and the comparisons are with the other 145 countries in ICP 2005. As we can see, the two measures are close to one another, so that in this simple two country case, the new measure gives similar results to the standard one. Or perhaps more simply, the variance of the log price ratios is approximately equal to the log of the ratio of the Laspeyres and Paasche indexes. This also means that the standard error of the log

geomean can be approximated by the square root of the log Laspeyres-Paasche ratio divided by the number of goods.

Why should this be so? At an intuitive level, the Paasche Laspeyres spread is one measure of the range of possibilities in the comparison of prices between two countries. The failure of arbitrage standard error measures the degree of uncertainty associated with deviations from the law of one price. If the law of one price were to hold, the Laspeyres and Paasche indexes would be identical, and there would be no
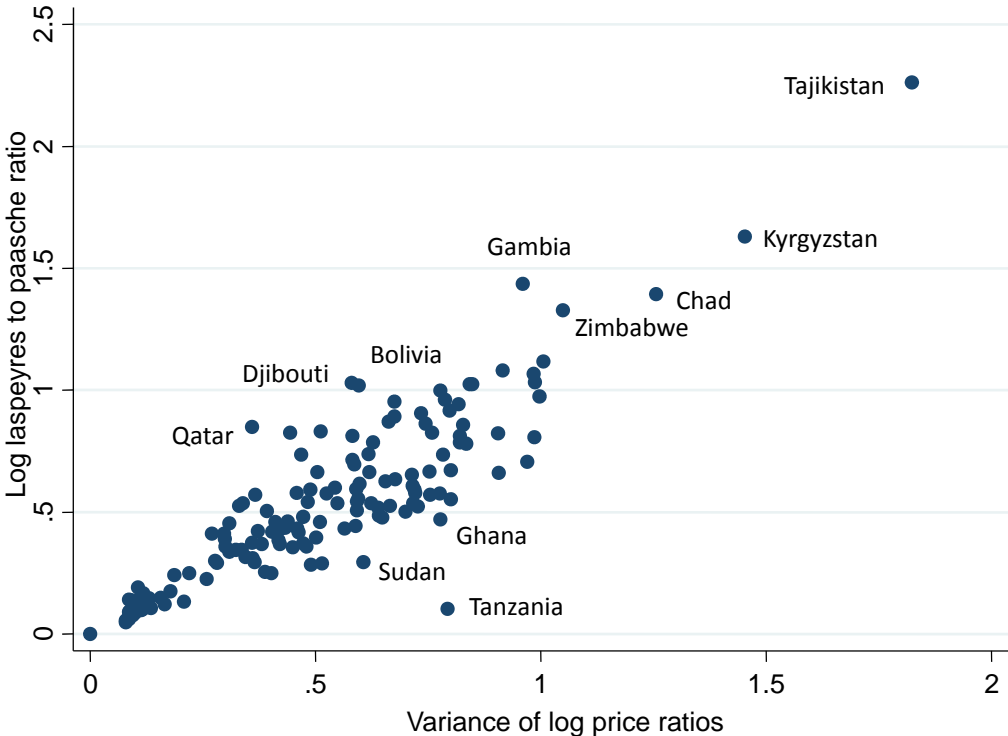

Figure 1: Variance of log price ratios and Laspeyres-Paasche ratio: US base

spread. More formally, we can rewrite the Laspeyres Paasche ratio (1) using the price formula (2) to give

$$\ln \rho_{dc} = \ln\left[\sum_{i=1}^{n} s_{ic} \exp(\varepsilon_{id} - \varepsilon_{ic})\right] + \ln\left[\sum_{i=1}^{n} s_{id} \exp(\varepsilon_{ic} - \varepsilon_{id})\right] \tag{7}$$

Equation (7) is exact, but is more helpful if we take a second order expansion in the ε_{ic}. This yields

$$\ln \rho_{dc} \simeq \sum_{i=1}^{n}(s_{ic} - s_{id})(\varepsilon_{id} - \varepsilon_{ic}) + \frac{1}{2}\sum_{i=1}^{n}(s_{ic} + s_{id})(\varepsilon_{id} - \varepsilon_{ic})^2 \tag{8}$$

Taking expectations, and once again assuming that ε_{ic} are orthogonal both over commodities and countries, and that the variances are identical over commodities, we have

$$E(\ln \rho_{cd}) = \sigma_c^2 + \sigma_d^2 \tag{9}$$

To this degree of approximation then, we have the result that the variance of the log price ratios is the log Laspeyres to Paasche ratio. Note that the absence of the shares in (9) depends on the assumption that the variances are the same for all goods; if the variances are different, the expectation of the log ratio will differ from the variance of the log price ratios. Even so, Figure 1 shows that the approximation is close in the ICP 2005 data.

We can also use Figure 1 to read off some of the relevant magnitudes. For the two "closest" countries to the US, Ireland and Canada, the standard error of the geomean PPP is 2.5 percent. For Gambia it is 8.7 percent, Kyrgyzstan 10.7 percent, and Tajikistan 11.9 percent. For two important countries in the middle, India and China, the standard errors are 7.1 percent and 6.8 percent respectively.

I have so far established that there is a close link between the Laspeyres to Paasche ratio and the variance of log price relatives in the two country case. This

result also implies that there is a similar relationship between the Laspeyres to Paasche ratio and the "failure of arbitrage" standard error of the log of the geomean PPP between two countries. In the next section, I extend this analysis to more relevant price indexes.

3. Törnqvist and Fisher indexes in a bilateral star system

The geomean index is not a practical contender for a PPP if only because it ignores the expenditure weights, the very fact that makes it useful at the finest level of detail where no weights exist. In this section, I consider more realistic formulas, and derive the failure of arbitrage standard errors.

The two-country Törnqvist index is written

$$\ln P_{cd}^{T} = \frac{1}{2}\sum_{i=1}^{n}(s_{ic} + s_{id})\ln\frac{p_{id}}{p_{ic}} \tag{10}$$

which fits with (2) to give, without approximation

$$\ln P_{cd}^{T} = (a_{d} - a_{c}) + \frac{1}{2}\sum_{i=1}^{n}(s_{ic} + s_{id})(\varepsilon_{id} - \varepsilon_{ic}) \tag{11}$$

The log of the Fisher index is half the sum of the log Laspeyres and log Paasche, so that, following (7), we have

$$\ln P_{cd}^{F} = a_{d} - a_{c} + \frac{1}{2}\ln\left[\sum_{i=1}^{n}s_{ic}\exp(\varepsilon_{id} - \varepsilon_{ic})\right] - \frac{1}{2}\ln\left[\sum_{i=1}^{n}s_{id}\exp(\varepsilon_{ic} - \varepsilon_{id})\right] \tag{12}$$

If we make a first-order approximation to (12), we get back to the Törnqvist (11), so that to this order, the Fisher and Törnqvist indexes are identical. Equation (11) gives us the variance of the two-country Törnqvist as well as an approximation to the two-country Fisher

$$V(\ln P_{cd}^T) = \overline{s}^{cd} \, {}^{\prime}V_{\varepsilon\varepsilon}^{cd} \overline{s}^{cd} \qquad (13)$$

where the overbar denotes the mean of the two country expenditure shares and $V_{\varepsilon\varepsilon}^{cd}$ is

the variance covariance matrix of $\varepsilon_d - \varepsilon_c$. In the examples in the previous section V_{cd}^{ε}

was taken to be a diagonal matrix with each element equal to $\sigma_c^2 + \sigma_d^2$. In this case,

and if, in addition, the expenditure shares are identical for all goods, (13) reduces to

the (square of) (6), so that the variance of the log Törnqvist (and approximate

variance of the log Fisher) is the same as the variance of the log geomean; indeed in

this case, the Törnqvist is identical to the geomean. More generally, with unequal

budget shares, the variance of the log Törnqvist will be at least as large as the

variance of the log geomean.

We have already seen the close relationship between the log of the Laspeyres-

Paasche ratio and the variance of the log geomean and this carries through to the

variance of the log Törnqvist. From (8) and (11) in the case where the covariances are

zero

$$E\ln\rho_{dc} = \sum_{i=1}^{n} \overline{s}_i (\sigma_{di}^2 + \sigma_{ci}^2) \qquad (14)$$

$$V(\ln P_{cd}^T) = \sum_{i=1}^{n} \overline{s}_i^2 (\sigma_{di}^2 + \sigma_{ci}^2) \qquad (15)$$

When the budget shares are identical, $V(\ln P_{cd}^T) = E(\ln\rho_{cd}) / n$. Figure 2 shows that a

plot of the square root of the log of the bilateral Laspeyres to bilateral Paasche

divided by the square root of n, the number of goods, against the standard error of

the log Törnqvist from (8). As before, the US is always the base country and these are

pairwise indexes, so that I am working with a star system with the US as the star. The

10

two colored points are India and China. The cross-country correlation in the figure is ρ=0.89. In these 2005 data, the log Fishers and the log Törnqvists are very close to the log geomean index, so that Figures 1 and 2 are closely related.
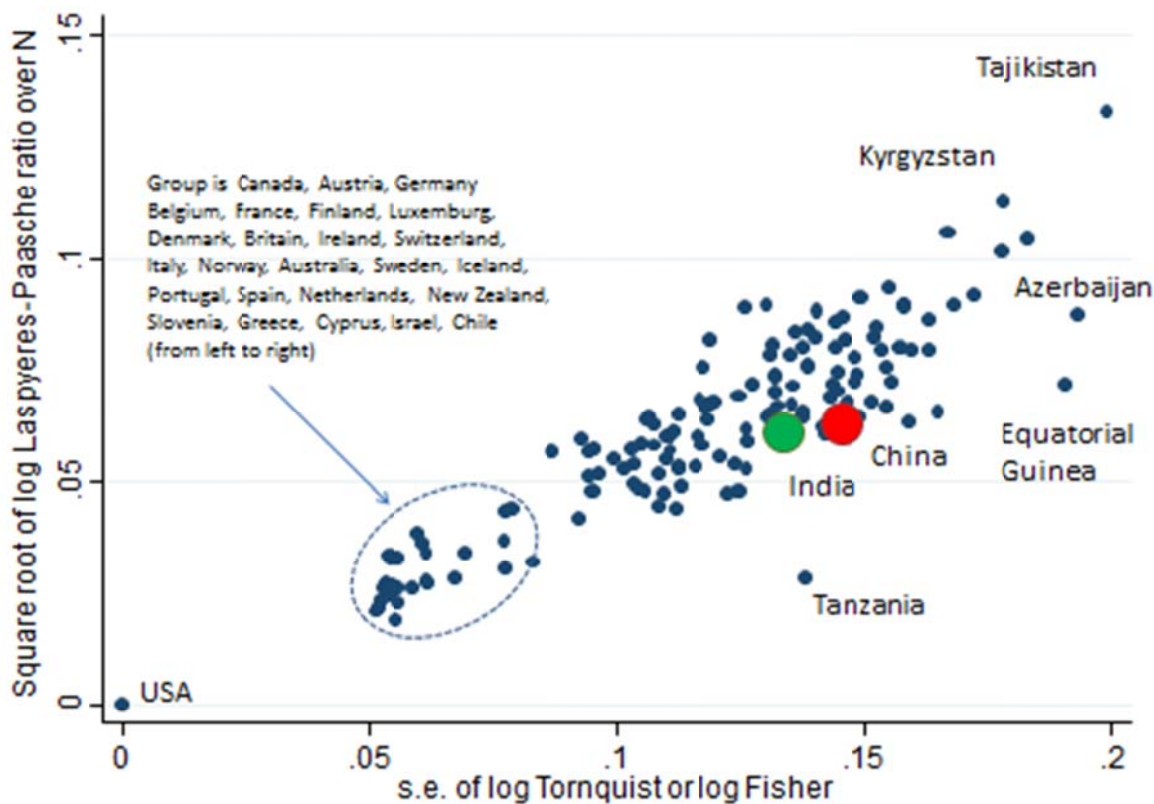

Figure 2: Paasche-Laspeyres spread and standard error of bilateral indexes

A number of points are worth noting here before going on to the multilateral indexes. First, the standard errors of the logarithms of the bilateral indexes are large, with many countries clustering around values of 0.10 to 0.15. For India and China, for example, if we were to take two standard errors on either side of the mean as a confidence interval, the ratio of Chinese or Indian to American prices could be 30 percent larger or smaller than measured. Given that the ICP takes GDP in local

currencies as given, this carries through to the same margin of uncertainty for GDP (or at least domestic absorption) in international prices. Second, for the group of countries at the bottom left of Figure 2, including all of western Europe, Canada, Australia, and New Zealand (but not Mexico or Japan, though both are close by), the standard errors are much smaller, between 5 and 8 percent. For these countries, whose economic structure is similar to that of the US, we can make much more precise comparisons of prices and GDP, though not without substantial uncertainty.

4. Multilateral weighted CPD and GEKS indexes

Multilateral indexes based on Törnqvist or Fisher indexes are known as GEKS indexes, and these take the pairwise country indexes for all pairs as a starting point and then adjust them in order to impose transitivity over countries, and to generate a single set of PPPs, one for each country. If we denote the (arbitrarily chosen) base country by 1, the Törnqvist-based GEKS index for country c is written

$$\ln PPP_c^T = \frac{1}{M}\sum_{d=1}^{M}(\ln P_{1d}^T + \ln P_{dc}^T) \tag{14}$$

In order to calculate the variance of (14), we substitute in (11) to get

$$\ln PPP_c^T = \alpha_c + \frac{1}{2M}\sum_{j=1}^{M}\left[(s_1 + s_j).(\varepsilon_j - \varepsilon_1) + (s_j + s_c)(\varepsilon_c - \varepsilon_j)\right] \tag{15}$$

where the "." Indicates an inner product and s and ε with single (country) suffices are vectors with one element for each good. The variance of the Törnqvist PPP is then computed by squaring the second term on the right hand side of (15), and taking expectations. Again, a simple version comes from assuming that the ε are independent over goods and countries. More generally, we have

$$4M^2V(\ln p^i) = (1 - 2\delta_{1i})\sum_{j=1}^{M}\sum_{k=1}^{M}(s^i + s^j)\Omega^i(s^i + s^k) + \sum_{j=1}^{M}\sum_{k=1}^{M}(s^1 + s^j)\Omega^1(s^1 + s^k) +$$

$$\sum_{j=1}^{M}(s^1 - s^j)\Omega^j(s^1 - s^j) + 2\sum_{j=1}^{M}(s^i + s^j)\Omega^i(s^1 - s^i) - 2\sum_{j=1}^{M}(s^1 - s^i)\Omega^1(s^1 + s^j) \tag{16}$$

The variance-covariance matrices Ω are diagonal matrices with the commodity-specific variances on the diagonal. In practice, I evaluate them using the squared residuals from the original CPD regression (2). The formulas (14), (15), and (16) are also valid for the Fisher version of the GEKS PPPs with only the substitution of the Fisher for the Törnqvist index in (14).

The second multilateral index that I consider here is the weighted country-product dummy (CPD) index. This method works directly from equation (2), and estimates the country effects a_c by regressing the log prices on a set of country and commodity dummies. The method recognizes the importance of the expenditure shares by using them as weights to calculate a weighted regression, rather than a simple OLS regression. We define X as an MN by $N + M - 1$ matrix of 1's and 0's, where there is one row for each country commodity combination, and the columns are a constant, $N - 1$ commodity columns with 1's marking the position of each commodity, and $M - 1$ columns with 1's marking each country. If we write y for the MN vector of the log prices, and S for an MN by MN diagonal matrix containing the expenditure shares, the weighted CPD estimator comes from selecting the coefficients on the country dummies from the estimate

$$b = (X'SX)^{-1}X'Sy \tag{17}$$

Note that this is *not* a GLS estimator in the usual sense because S is there to control the projection, and is not related to the variance covariance matrix of the residuals in

(2). Rather, we must use the so-called "outer product" variance covariance matrix for (16), which takes the form

$$V(b) = (X'SX)^{-1}X'S\Sigma SX(X'SX)^{-1} \qquad (18)$$

where Σ is the *MN* by *MN* variance covariance matrix of the ε, the deviations of the log prices from their country product predictions. If we follow the assumptions so far, this will be a diagonal matrix with the σ_c^2 on the diagonal.

Using the data from ICP 2005, the three multilateral indexes, the GEKS Fisher, the GEKS Törnqvist, and the weighted CPD, are close to one another, with correlations of the logarithms over countries in excess of 0.999. Even deflated by the average foreign exchange rate for 2005 so that we are correlating the price levels of domestic absorption, the correlations are all above 0.978. Nor are the two sets of standard errors (one for the Fisher and Törnqvist and one for the weighted CPD) very different; the cross-country correlation is 0.928 (excluding the US, whose standard error is zero by construction). This negative result is important because it makes it clear that the uncertainty in the PPPs comes, not from the choice of index number formula, at least for the index number formulas considered here, but from the dispersion of relative prices. The latter has similar effects across index formulas, at least for those considered here.

The more interesting results are those akin to Figure 2, on the size of the standard errors and on the differences in precision across groups of countries. The new standard errors—of the multilateral GEKS—are compared with the old—the standard errors of the bilateral Törnquist or Fisher indexes—in Figure 3, which allows assessment of both size and precision by country. The new multilateral standard

14

errors are plotted on the vertical axis against the original bilateral ones on the horizontal axis. On average, the multilateral standard errors are larger, 0.15, as
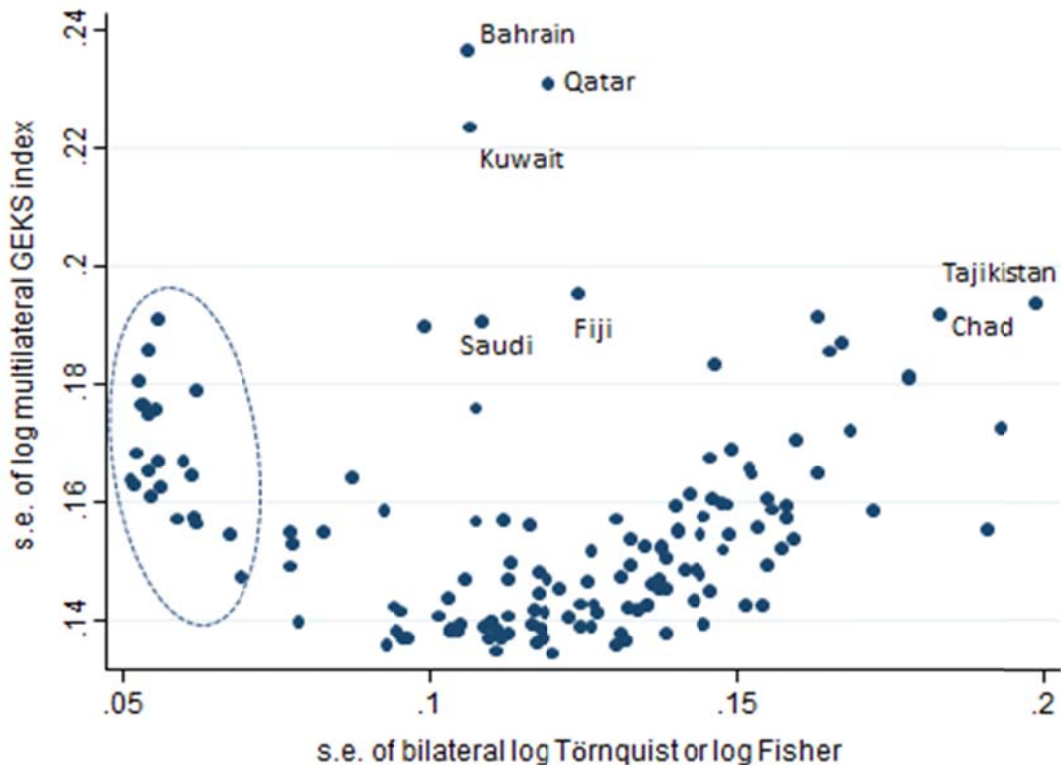

Figure 3: Standard errors of bilateral and multilateral price indexes

opposed to 0.12, but their dispersion is lower, with the standard deviation of the standard errors reduced from 0.036 to 0.022. There are a number of factors at work here. The multilateral indexes introduce a great deal more information, particularly the comparisons between countries other than the US all of which are ignored by the bilateral indexes. More information, by itself, should contribute to precision, but it depends on the nature of that information, and the multilateral indexes use *all* country comparisons, some of which will involve huge variability in relative prices,

think Japan versus Oman, or Fiji versus Finland. The enforcement of transitivity reduces variance of averages, but the raw material going into the averages is more variable to start with.

These offsetting forces work differently for different groups of countries. On the left of Figure 3, inside the dotted ellipse, are the same Western European and European offshoot countries that were grouped in Figure 2. For them, the standard errors have been increased by a large amount, typically by a factor of 3. For these structurally similar economies, the introduction of relative prices from all pairs of countries, and the enforcement of transitivity so that, for example, the price level of Canada relative to the US must now be the same whether it is calculated directly, or through any possible chain of countries—from the US to Oman to Chad to Tajikistan to Canada—introduces much uncertainty that is avoided by the bilateral comparison. The lesson from this is that, if we want a bilateral comparison, we should make a bilateral comparison; nothing but avoidable noise comes from introducing countries that are irrelevant for the problem at hand. More broadly, if we only need comparisons between Western Europe, the US, Canada, Australia, and New Zealand (and possibly Mexico and Japan), we should use multilateral indexes for them that ignore the rest of the world. Of course, the ICP does exactly this by enforcing the principle of fixity, for which the current analysis provide considerable support.

There are 13 countries whose multilateral index with the US has lower variance than their bilateral index. Beyond that, there is a large group of countries, in the bottom middle of Figure 3, where the standard errors are similar, and the costs of the additional information are offset by the benefits of enforcing transitivity. For

these non-OECD countries, it is not clear that the comparison in Figure 3 is the relevant one, though it is presumably still the case that the regional comparisons that are incorporated into the ICP through fixity are likely to be more accurate than the full international multilateral comparison. Even so, we should not expect for them the tight bounds that we get for the OECD countries. Several of the regional groupings contain countries that are much more different from one another than any of the countries of the OECD.

Finally I note that, as was the case for the bilateral indexes, the standard errors of the multilateral indexes are large. There is very substantial uncertainty attached to multilateral price indexes, so that we know much less about international comparisons of national income than is commonly recognized. The large revisions associated with successive rounds of the ICP may reflect this uncertainty, at least in part. In any case, we would do better to recognize the margins of uncertainty in making international comparisons in general, especially those whose structures are dissimilar.

References

Deaton, Angus, and Olivier Dupriez, 2011, "Purchasing power parity exchange rates for the global poor," *American Economic Journal Applied*, 3:1, 137–66.

Hill, Robert J., 1999, "Comparing price levels across countries using minimum spanning trees," *Review of Economics and Statistics*, 81:1, 135–42.

Hill, Robert J., and Marcel P. Timmer, 2006, "Standard errors as weights in

    multilateral price indexes," *Journal of Business and Economic Statistics*, 24:3,

    366–77.

Summers, Robert, 1973, "International comparisons with incomplete data," *Review of*

    Income and Wealth, 29:1, 1–16.