Purchasing power parity exchange rates for the poor:

using household surveys to construct PPPs

Angus Deaton

Research Program in Development Studies

Princeton University

This version, August 2006

ABSTRACT

This paper builds a bridge between two literatures, that on purchasing power parity (PPP) exchange rates, which is an extension of national income accounting, and that on poverty measurement, which is based on household survey data on consumption expenditures. It also aims to serve as a manual for those who wish to calculate PPP price indexes using household surveys, particularly, although not exclusively, the PPP price indexes for the poor to be used to construct internationally comparable poverty lines. Because poverty analysts are often unfamiliar with PPP construction, PPP indexes are dealt with from something like first principles. The paper begins with the idea that PPP price indexes, like the usual domestic consumer price indexes, can be computed using weights from household surveys. Section 1 deals with the case of two countries, each with a set of consumer prices, and each with a household survey detailing expenditures on each good for a national sample of households. This first section is concerned with national aggregates, as in standard PPP comparisons, so that the household survey is used only to provide the national average consumption pattern. In this simplified two-country case, where the object of interest is a standard national PPP consumption comparison, it is possible to set up a framework that can be easily extended to deal with many countries and with poverty-weighting. In particular, standard errors are defined and formulas given. Prices are treated as known, so that the source of estimation variance is the sampling variability of the expenditure weights from the household survey, a sampling variability that depends on the sample size and on the survey design. A second type of standard error is distinguished which is new to the literature. In a world of perfect arbitrage and costless trade, relative prices would be the same in all countries, and all methods of computing PPP indexes would give the same answer. Deviations of prices from this ideal give rise to uncertainty about the index. Treating these deviations as random, as in the stochastic approach to price indexes, but with expenditure weights as non-stochastic, gives a second set of standard errors that reflect the uncertainty associated with the failure of arbitrage that is the fundamental reason why we need PPP index numbers. Formulas are given for these standard errors for the usual PPP price index numbers, including the Fisher and Törnqvist versions of the EKS index, as well as weighted CPD indexes. Section 2 provides illustrative calculations for a national consumer PPP index for food, fuel, alcohol, and tobacco for Indonesia in terms of India in 1999–2000. Section 3 extends the two-country analysis to the case of "poverty" PPPs, which are international price indexes calculated for people at or near the poverty line, under the requirement that the ratio of the two poverty lines is equal to the PPP index. It shows that, when the Engel curves take a specific but realistic functional form, there is a closed form solution for the poverty PPP index, and proposes using this case as a starting value for a non-parametric, but iterative, calculation. Section 4 applies this case to the Indonesian to Indian comparison. Section 5 extends the analysis in Sections 1 and 3 to the multiple country case, and Section 6 provides illustrative calculations of food and tobacco PPPs for India, Indonesia, Bangladesh, and Pakistan.

Table of Contents

1. Using household surveys to compute national level consumption PPPs for two countries

1.1 Laspeyres, Paasche, and Fisher

A simple way to begin is with two countries, India and Indonesia, say, which I label 1 and 2. In each, consumers buy N goods, indexed by n, and I start with the *average* consumption bundles for each country, q^1 and q^2, both N-vectors. In this section, I set up notation and concepts in terms of national averages, as would be done for constructing national accounts, rather than for poverty-relevant indexes, which I consider in the next section. Suppose that prices are the same for everyone in each country, and that the two price vectors, expressed in local currencies, are p^1 and p^2. One way to measure the price level in country 2 relative to country 1 is to price out country 1's consumption bundle at country 2's prices, and then compare the relative cost. This gives the familiar *Laspeyres* index of 2 based on 1, which we write

$$P_L^{12} = \frac{p^2 . q^1}{p^1 . q^1} \tag{1}$$

where the "." indicates an inner product of the two vectors. Note also that I adopt the convention that of the two superscripts in a price index the first is the base country and the second is the country being compared. The Laspeyres index in (1) is expressed in (Indonesian) rupiah per (Indian) rupee.

In this context, there is every reason to treat the two countries symmetrically, since neither one has any unique claim to be the base or the comparison. In particular, we could equally well have started with country 2 as base and computed the Laspeyres index for country 1 with country 2 as base. This index, of course, would be expressed in rupees per rupiah, so it should be compared, not to (1) but to its reciprocal. Put another way, the two rupiah to rupee exchange rates considered so far are given by (1), the Laspeyres for country 2 relative to country 1, and the

reciprocal of the Laspeyres for country 1 relative to country 2, which we write

$$\left[P_L^{21} \right]^{-1} = \frac{p^2 \cdot q^2}{p^1 \cdot q^2} \tag{2}$$

Equations (1) and (2) could just as well be replaced by *Paasche* indexes, but we can see that this has already been done. Equation (2) is not only the reciprocal of the Laspeyres index of 1 based on 2, but it is also the Paasche price index comparing 2 with base 1. Similarly, (1) is not only the Laspeyres index for 2 with base 1, but also the reciprocal of the Paasche index for country 1 with 2 as base.

Faced with two estimates of the same thing, (1) and (2), and, by symmetry, no reason to favor one over the other, an obviously sensible procedure is to take an average while keeping track of some measure of their difference; this argument is extensively developed by Erwin Diewert (1997, 2001). It turns out to be convenient to do this in logarithms or, equivalently, by computing a geometric rather than an arithmetic mean, since the geometric mean of the Paasche and the Laspeyres is the Fisher ideal index, which has several convenient properties of its own. The Fisher index for 2 based on 1 is written

$$\ln P_F^{12} = \frac{1}{2}\left(\ln P_L^{12} + \ln P_P^{12} \right) = \frac{1}{2}\left(\ln P_L^{12} - \ln P_L^{21} \right) \tag{3}$$

where P_P is the Paasche index. Note that, by averaging in this way, we have solved the symmetry problem; as is easily checked from (3), the Fisher index for 2 based on 1 is the reciprocal of the Fisher index for 1 based on 2. Transposing 1 and 2 in (3) gives

$$\ln P_F^{12} = -\ln P_F^{21} \tag{4}$$

This property, that we get the same answer whether we compare India with Indonesia, or Indonesia with India, is referred to as the "country reversal" property, and is, in general, satisfied by neither the Paasche nor the Laspeyres separately.

2

1.2 Standard errors

If we maintain the idea that we are computing an average of multiple (here two) estimates of the same thing, we might also compute standard error for our estimate. But with only two estimates, it is unlikely to be useful to calculate a standard error. A better approach is to follow the "stochastic approach" to price index numbers, and to derive standard errors associated with appropriately specified stochastic variation in prices. In this approach, it is explicitly recognized that, in a world in which the law of one price does not hold so that relative prices differ in the two (or many) countries, any index number is an average over many possibilities and that there is a margin of uncertainty reflecting the need to make a single choice. In Section 1.5 below, I will discuss the "country-product dummy" (CPD) model, which can be used as the stochastic basis for calculating standard errors for the Laspeyres, Paasche, and Fisher indexes.

1.3 Cost-of-living indexes and superlative indexes

As shown by Diewert (1976), the Fisher index (3), is a *superlative* index, meaning that it is capable of providing a local second-order approximation to any cost-of-living index, where a cost-of-living index is the price index defined by the ratio of the minimum costs of obtaining the same level of utility at the two sets of prices. Cost-of-living indexes work on the assumption that preferences are identical in the two places being compared. If preferences are homothetic, so that the pattern of demand is the same at all levels of living, the cost-of-living index depends only on the two price vectors, and a superlative index approximates it. If preferences are not homothetic, which is the relevant case, the cost-of-living index depends on the reference level of utility, which in this case, would be either the utility in country 1 or the utility in country 2, and the

superlative index provides a second-order approximation to the cost-of-living index at some intermediate level of utility. Cost-of-living indexes, unlike Paasche or Laspeyres indexes, incorporate the substitution effects that a utility-maximizing consumer would display in response to the differences in relative prices between the two countries. Superlative indexes capture this substitution, at least approximately, by incorporating quantities from both countries, as opposed to Paasche and Laspeyres indexes, which use a commodity bundle from one or the other. The averaging over multiple indexes not only solves the symmetry or country-reversal problem, it also incorporates an allowance for substitution.

Cost-of-living index numbers apply to individual agents, not to the averages with which we have been dealing so far. The aggregation required can be dealt with by appropriate redefinition of the index numbers, and indeed will be less of a problem when we come to think about index numbers for a specific group, such as the poor, or those living close to the poverty line.

More generally, opinions differ on the usefulness of cost-of-living indexes for making price index comparisons between countries, or further still, whether cost-of-living index numbers ought to be the gold standard that PPP exchange rates should approximate. One immediate difficulty is the assumption that consumers in the two countries (or in the general case, in *all* countries) have the same tastes irrespective, not only of local customs and traditions, but of climate and temperature. It is far from clear that it makes sense to assume that China and Brazil have the same tastes, or even that northern and southern India do so. This case can be contrasted with the use of cost-of-living indexes to measure price changes over time within a single country where, at least to an approximation, the people being compared are the same in the two situations. Cost-of-living theory requires some set of reference preferences to evaluate the

4

different price vectors, yet selecting one country violates the symmetry that is an important part of cross-country indexes. With different tastes, a superlative index delivers an approximation to a cost-of-living index evaluated at some intermediate or average tastes, and while this solves the symmetry problem, it is not clear that the cost-of-living index that is being approximated has a great deal of claim on our attention as a cost-of-living index.

Even if we are not concerned with the ability of superlative indexes to approximate cost-of-living index numbers, they have other advantages that are extremely useful when constructing international price indexes. In particular, Diewert (1976) has shown that *all* superlative indexes are symmetric in the sense that they satisfy the country reversal property (4); this is a property shared by all price indexes that correspond to cost-of-living index numbers for homothetic preferences. In consequence of the symmetry property, one way of generating useful PPP index numbers is to start from the class of superlative indexes.

1.4 The Törnqvist index

The Törnqvist index is another superlative index. It works by defining the *budget shares* which are the fractions of total expenditure in each country that is accounted for by good n,

$$s_n^c = \frac{p_n^c q_n^c}{\sum_k p_k^c q_k^c} \tag{5}$$

for countries $c = 1,2$. The logarithm of the Törnqvist index is then the weighted average of the logarithmic price ratios between the two countries, with weights defined as the averages of (5) for the two countries. Hence, we have

$$\ln P_T^{12} = \frac{1}{2} \sum_{n=1}^{N} (s_n^1 + s_n^2) \ln \frac{p_n^2}{p_n^1} \tag{6}$$

This index is a competitor for the Fisher index for calculating the PPP exchange rate between the two countries. Because both are superlative indexes, they will provide close approximations to one another around any specific set of prices and quantities, so that we might hope that they would not be far apart in practice. Note also that, like the logarithm of the Fisher index, the logarithm of the Törnqvist index is the average of two indexes in which the log price relatives are weighted by the shares of each country in turn. These two indexes will be identical for all price relatives if and only if the budget shares are the same in both countries, and at any set of budget shares, they will be identical if the price vector of one country is proportional to the price vector in the other, so that relative prices are the same in the two countries.

1.5 Country product dummy (CPD) indexes

The approach so far has been to start with familiar indexes, such as the Paasche and Laspeyres, and ask how they need to be modified in order to satisfy the requirements of a PPP index, particularly the country reversal test. Or alternatively, to start from a superlative index, which automatically satisfies the test by virtue of being superlative. A different approach is to start from the prices in each country, and to ask what properties they would have to satisfy so that PPP indexes would exist. One case is where the relative prices are the same in both countries and absolute prices differ only by a uniform scale factor that defines the units in which prices are measured in each country, and which defines the exchange rate between them. If this were the case, we could write

$$p_n^c = \theta^c \tilde{p}_n \tag{7}$$

for n from 1 to N and for $c = 1, 2$. The PPP exchange rate for 2 based on 1 is then θ^2 / θ^1 and that

for 1 based on 2 is θ^1 / θ^2. Note that, in this case, the Paasche, Laspeyres, Fisher, and Törnqvist

indexes are identical, as are all indexes that are linearly homogeneous in the prices and that

satisfy the "identity test," which requires that the index takes the value unity when a price is

compared with itself; all lead to the same PPP exchange rate. Equation (7) also corresponds to

the law of one price, which would be guaranteed by arbitrage if all goods were tradable, there

were no tariffs or taxes, and transport costs were zero. Such arbitrage would also guarantee the

country reversal test (otherwise money could be made by trading one currency for the other and

back again) so that the Paasche and Laspeyres indexes would be the same.

Just as we can justify the Fisher index as an average of the Paasche and Laspeyres that deals

with their divergence because of lack of arbitrage, we can calculate new prices that approximate

the actual prices but satisfy (7). This can be done by running a regression of the logarithms of

prices on a set of country and commodity (product) dummies or better, *projecting* the logarithms

of prices on a set of country and commodity dummies, so that the predicted values are the

projection of the log prices on the best fitting price structure that satisfies the law of one price;

this procedure was first suggested by Robert Summers in 1973. We write this regression

$$\ln p_n^c = \alpha^c + \beta_n + u_n^c \qquad (8)$$

where *n* runs from 1 to *N*, and *c* from 1 to 2, although in this case the generalization to many

countries is immediate. With two countries and *N* commodities, there will be one country

dummy and $N-1$ commodity dummies on the right hand side of (8) so that, if we drop the

dummy for country 1 in favor of a constant term, and compare with (7), we see that the estimate

of α^2 is the logarithm of the PPP exchange rate of country 2 relative to country 1.

One further modification is required. As written, the regression (8) minimizes the residual

7

sum of squares, and so gives equal weight to deviations from the law of one price for all goods, irrespective of their importance. Rice and wheat get the same weight as salsify, barbuti beans, or airplane tickets. To deal with this, (8) is run as a weighted (or generalized least squares regression) in which the weights are the budget shares of each good in each country (5); this defines the weighted CPD procedure, developed by Prasada Rao, Selvanathan and Rao (1994), and Rao (1990, 1995, 2001, 2002, 2004), and further investigated by Diewert (2004, 2005).

The weighted regression can be written in standard form

$$\hat{b} = (X'SX)^{-1} X'Sy \tag{9}$$

where X is a $(2N \times N+1)$ matrix of ones and zeroes, with $N-1$ columns for the commodities (N commodities less one) plus a column for country 2, plus a constant, and rows corresponding to the vector y, which is the "stacked" vector of log prices, N for each country. (With C countries, X will have $C+N-1$ columns and CN rows.) The S matrix is a diagonal matrix with the budget shares on the diagonal, N for country 1 followed by N for country 2. The element of the estimated parameter b corresponding to the dummy on country 2 is the estimate of the log of the PPP exchange rate for country 2 in terms of country 1. Note that, although (9) can be thought of as a generalized least-squares estimator, the weighting matrix S is included for substantive reasons to do with the importance of each good in spending patterns, and not because of any supposed relationship between the budget shares and the variances of the error terms in (8). Indeed, (8) should not be thought of as a "true" model of the data generating process; rather (8) and (9) should simply be regarded as a convenient device for projecting the log prices on country and commodity effects in a metric that recognizes the different importance of each commodity in the budget.

A standard error for the weighted CPD exchange rate can be calculated from the variance-covariance matrix of (9). A heteroskedasticity-robust estimator of this variance covariance matrix can be written

$$\widehat{V}(\widehat{b}) = (X'SX)^{-1}(X'S\Sigma SX)(X'SX)^{-1} \tag{10}$$

where Σ is a diagonal matrix whose elements are the squared residuals from the *unweighted* OLS regression corresponding to (8). The square root of the appropriate element of (10) is an estimate of the standard error of the logarithm of the estimated PPP. If there is evidence that there are correlations between the residuals for groups of commodities (e.g. traded goods, or a group of fuels whose prices tend to all be high or all low in any country), or across groups of countries in the many country case, the matrix Σ could be modified appropriately.

Note finally that the standard errors calculated by these procedures are associated with the uncertainty that comes from there being many different ways of calculating a PPP, among which we have to choose. It does not capture any of the uncertainty associated with sampling of the prices or the budget shares. Sampling variances will be dealt with below. Even apart from sampling variances, it treats the logarithms of the prices as stochastic, distributed around a set of common relative prices, while treating the budget shares, the weights, as fixed. Other schemes are clearly possible; for example, it would be possible to consider the case where the weights are also stochastic, responding to the international stochastic variation in prices. By contrast, the standard errors developed here are conditional on the weights, which are treated as fixed (except for sampling variability), and this statistical—not behavioral—approach treats the problem of computing PPP indexes as one in which weights are given, but there is uncertainty about indexes that comes from the fact that prices in different countries are not fully arbitraged so that there is a

range of possible PPP indexes. It is this range that the variances are designed to capture.

1.6 CPD based standard errors for Laspeyres, Paasche, Fisher, and Törnqvist

We can use equation (8) to derive "failure of arbitrage" standard errors for the other indexes. Note that (8) implies that we can write logarithms of the relative prices of country 2 relative to country 1 as

$$\ln(p_n^2 / p_n^1) = \theta + \eta_n = (\alpha^2 - \alpha^1) + (u_n^2 - u_n^1) \tag{11}$$

where θ is the logarithm of the PPP exchange rate of country 2 in terms of country 1. From (11), we have immediately that the log of the Törnqvist index satisfies

$$\ln P_T^{12} = \theta + \frac{1}{2} \sum_{n=1}^{N} (s_n^1 + s_n^2)\eta_n = \theta + \overline{s}\,'\eta \tag{12}$$

in an obvious notation for the mean of the budget shares over the two countries. The variance of this expression can be written

$$V(\ln P_T^{12}) = \overline{s}\,'V_{\eta\eta'}\overline{s} \tag{13}$$

where $V_{\eta\eta'}$ is the variance-covariance matrix of η. We can estimate this using the information from the unweighted CPD regression, in parallel to the estimation of Σ for (10). In particular, from the definition of η in (11), we can rewrite (13) as

$$V(\ln P_T^{12}) = \overline{z}\,'V_u\overline{z} \tag{14}$$

where $\overline{z}\,'=[-\overline{s}\,', \overline{s}\,']$ is a (1 x 2N) matrix formed from the (N x 1) matrix of averaged budget shares $\overline{s}$. This can be estimated using the Σ, the diagonal matrix of squared residuals from the CPD regression by

$$\widehat{V}(\ln P_T^{12}) = \overline{z}\,'\Sigma\overline{z}. \tag{15}$$

Given (11), the log Laspeyres index takes the form

$$\ln P_L^{12} = \theta + \log\left(\sum_{n=1}^{N} s_n^1 \exp \eta_n\right) \tag{16}$$

Ignore squared and higher terms in η_s, (16) can be approximated by

$$\ln P_L^{12} \simeq \theta + \sum_{n=1}^{N} s_n^1 \eta_n \tag{17}$$

So that, parallel to (15), we can approximate the variance of the log Laspeyres by

$$\widehat{V}(\ln P_L^{12}) = z^1 {}' \Sigma z^1 \tag{18}$$

where $z^1{}' = [-s^1{}', s^1{}']$. Following through the same logic for the log Paasche, and remembering that $\ln P_P^{12} = -\ln P_L^{21}$ we have the parallel estimate of the variance

$$\widehat{V}(\ln P_P^{12}) = z^2 {}' \Sigma z^2 \tag{19}$$

where $z^2{}' = [s^2{}', -s^2{}']$. By averaging the log approximations for the Laspeyres and Paasche, it turns out that the corresponding approximation for the log Fisher is identical to the approximation for the log Törnqvist, (12), so that (15) can serve double duty, as an estimator of the variance of both the log Fisher and the log Törnqvist. This also implies that if (8), the CPD approximation to the log prices, is close, the Törnqvist and Fisher indexes will often be closer together than either is to the CPD index.

1.7 Törnqvist, Fisher, and CPD indexes

Both the Fisher and Törnqvist indexes are superlative, and so can be expected to approximate one another. Diewert (2005) has also shown that the weighted CPD index described above is closely related to the Törnqvist index. While the Törnqvist index (6) is a weighted average of the logarithms of the price relatives with weights that are the arithmetic mean of the budget shares in the two countries, Selvanathan and Prasado-Rao (1994, page 25) and Diewert (2005) show that

(in the two country case), the weighted CPD index is a weighted average of the logarithms of the price relatives with weights that are (normalized) *harmonic* mean of the budget shares in the two countries. Instead of the arithmetic mean, define

$$\tilde{s}_n = \frac{h(s_n^1, s_n^2)}{\sum_n h(s_n^1, s_n^2)} \tag{20}$$

Where $h(a, b)$ is the harmonic mean defined by

$$h(a,b) = \left(\frac{1}{2} a^{-1} + \frac{1}{2} b^{-1} \right)^{-1} \tag{21}$$

The weighted-CPD is then a second-order approximation to the Törnqvist around any specific price and quantity combination. The two indexes are likely to be close in practice, at least in this two good case. Note that these results mean that we can also estimate the variance of the two-country CPD index from (15) with $\tilde{s}$ and the corresponding $\tilde{z}$, replacing $\bar{s}$ and $\bar{z}$.

1.8 The Geary-Khamis index

Although we shall not use it in the rest of this paper, we briefly discuss another PPP formula, due to Geary (1958) and Khamis (1970), and which forms the basis of the calculations for the Penn World Table. In the Geary-Khamis system, the prices in each country are compared with those of an imaginary composite country, itself constructed from averaging the countries in the system. In the case of two countries, the Geary-Khamis PPP index is computed as a Paasche index that compares domestic prices with "world" prices, which are the prices of the composite so that, for $c = 1, 2$,

$$P_{GK}^c = \frac{\sum_{n=1}^{N} p_n^c q_n^c}{\sum_{n=1}^{N} \pi_n q_n^c} \tag{22}$$

12

where π_n is the world price of good *n*, which is itself defined as the quantity weighted average of the prices of good *n* in each country, expressed in the global currency. Hence

$$\pi_n = \frac{p_n^1}{P_{GK}^1} \frac{Q_n^1}{Q_n^1 + Q_n^2} + \frac{p_n^2}{P_{GK}^2} \frac{Q_n^2}{Q_n^1 + Q_n^2} \tag{23}$$

where the uppercase *Q*'s denote *aggregate* quantities, as opposed to averages. Note that (22) and (23) need to be solved simultaneously, which can be done iteratively, or as shown by Diewert (1999) as the solution to an eigenvalue problem. When there are many countries, (23) is extended in the obvious way, by summing over all the countries, with weights equal to that country's share of total consumption.

The Geary-Khamis system of PPPs has one great advantage, which is that it preserves aggregation over subgroups of consumption. Because there exists a world price for each good, each item of consumption can be repriced at the world price, and added up to give repriced subgroups or totals. This aggregation property is not shared by the other indexes discussed so far which would give, for example, PPPs for food, non-food, and total consumption that would not be mutually consistent in the sense that the sum of food and non-food in the international currency would not generally add up to total consumption in the international currency. For a system of national accounts, such as the Penn World Table, the absence of aggregation would be inconvenient, although not insuperable; indeed EUROSTAT uses EKS indexes, a multicountry generalization of Fisher indexes, for its PPP national accounts for Euroland. For the poverty work that is our main concern here, where we need only an exchange rate for total consumption, failure of aggregation is not a serious issue.

But the Geary-Khamis system has disadvantages that make it unattractive for our purposes. Unlike the Fisher and Törnqvist indexes, it is not superlative. In consequence, if the two

countries had the same homothetic tastes, the Geary-Khamis index would not be a second-order approximation to the "true" cost-of-living index. If this were the main concern, Geary-Khamis could be replaced by Peter Neary's (2005) GAIA system, based on Geary-Khamis, but which is fully utility consistent, though its evaluation requires the estimation of a full international system of demand equations. However, given that we do not want to assume identical tastes, nor use the cost-of-living framework that would be arguably appropriate if tastes were identical, these objections to Geary-Khamis are not decisive.

To see the central problem, note that the quantity weighting of prices in (12) means that the country with the larger physical volume of consumption of a good gets greater weight in the construction of the composite world prices. If, for example, we used Geary-Khamis to compute a PPP for Bangladesh relative to the US, the world prices would be close to those of the US. In the Penn World Table as a whole, it has been argued by Daniel Nuxoll (1994) that the composite world prices are those that would characterize a middle-income country such as Italy or Hungary. The use of such prices has the effect of overstating the level of consumption in poor countries. For example, many services—haircuts, domestic service, restaurant meals—are cheap in poor countries because people are poor, because such services cannot be traded, and because labor is not free to move around the world. If we use (say) Italian prices to value (for example) Indian consumption, these components of consumption will be valued very highly, and will inflate the value of Indian consumption at international prices. This is called the Gershenkron effect, the overvaluation of one country's consumption when evaluated at another country's prices. Put another way, it is the understatement of the price level in a one country relative to another that comes from using a Paasche index; in (22) the domestic value of consumption in the

numerator is divided by the inflated value of consumption at world prices in the denominator. Of course, the understatement of the Paasche index is an understatement relative to a true cost-of-living index, which is itself not well-defined in the current context. But if we compare the Paasche in (22) with the superlative indexes presented above, the latter always averages weights from *both* countries, whereas the Geary-Khamis index uses only *domestic* weights. Compared with the superlative indexes, Geary-Khamis indexes will understate PPPs in poor countries relative to rich ones, and overstate their living standards. They make the world look too equal, and understate poverty in the poorest countries. Of course, these effects are likely to be much smaller than the effects in the opposite direction that we would get by using market foreign exchange rates. Even so, we do not want to minimize poverty in India by valuing Indian services at anything like first-world prices.

2. Illustrative calculations for indexes at the national level

2.1 Data description

We start with one of the running illustrative examples in this manual, the computation of food, fuel, and tobacco PPP price indexes for India and Indonesia. For India, the household survey data come from the 55th Round of the National Sample Survey, a national household expenditure survey that was in the field between July 1st 1999 and June 30th 2000. For Indonesia, we use the SUSENAS household survey that was fielded in January and February 1999. Both surveys are nationally representative multistage random surveys. In India, there are 48,919 urban and 71,382 rural households in the sample; in Indonesia there are 25,313 urban and 35,970 rural households. Both surveys collect household expenditure data on several hundred items, and this information

is used to calculate the budget share weights for the various indexes.

In standard International Comparison Project (ICP) calculations, the expenditure weights come, not from household surveys, but from the consumption sector of the national accounts. The weights are then combined with prices that are collected, either as part of each country's regular price collection activities, or are specially gathered for the ICP. For the calculation of poverty PPPs, the prices could be the same prices as for the regular calculation, but the expenditure weights would come from household surveys. Prices would be matched to expenditures at the basic heading level, or at some higher level of aggregation, by matching commodities in the surveys to the ICP prices. Because the basic heading categories of the ICP are not likely to be the same as the commodities in the survey (there is rarely a very good match even between survey and NIPA consumption classifications), there is likely to be a good deal of approximation and aggregation, both of the expenditures and of the ICP prices. This process, although of the first importance in practice, is not described in this manual, whose main concern is with the use of the expenditure data.

In the calculations reported here, we use, not the ICP prices, but the prices, or rather unit values that are collected within the household surveys themselves. Households are asked to report both expenditures and physical quantities of goods, so that the ratio of the two yields a unit value for each household recording a purchase. The unit values are contaminated by quality effects, and they show small systematic differences across households so that, for example, better-off households tend to pay slightly more per unit, but they are typically close to prices that are collected directly in market surveys. Unfortunately, they are only available for goods where quantities are sufficiently well-defined to be asked in a survey so that, at best, the expenditure

surveys yield unit values for food, tobacco, alcohol, and fuels. Among the poorest households in the world, these goods often account for two-thirds or more of the budget, but still exclude important items, such as housing, clothing, and transportation.

In the calculations here, there are 63 commodities that can be directly matched between the two countries and for which both surveys have prices. Together, these commodities account for 44 percent of the total budget in India and 39 percent in Indonesia. (Note that these coverage proportions will be much larger when we come to focus on the poor, rather than on national aggregates.) Table 1 lists the aggregate budget shares for a selection of these goods, including the most important, with India in the left-hand panel, and Indonesia in the right-hand panel. The budget shares in the table are calculated from the survey data by estimating the total national expenditure on each good and then dividing by the total national expenditure on *all* goods and services, not just the 63 that we are going to use subsequently. Hence, if h denotes a sample household, and $x_{nh} = p_{nh}q_{nh}$ the expenditure by that household on good n, the budget shares in the table take the form

$$S_n^c = \frac{\sum_{n=1}^{N} w_h x_h \left/ \sum_{n=1}^{N} w_h \right.}{\sum_{n=1}^{N*}\sum_{h=1}^{H} w_h x_{nh} \left/ \sum_{n=1}^{H} w_h \right.} \tag{24}$$

where, as before, c denotes a country, and n a good, here running to $N*$, the total number of goods in the budget, not just the 63 goods in the match. For clarity, we have dropped all of the c suffices on the right-hand side of (24). The household inflation factors w_h come from the survey design, and can be thought of as the number of population households represented by sample household h. Their sum is the total number of households in the population of country c, so that both numerator and denominator of (24) are weighted sample means, and the budget share S_n^c is

a ratio estimator of the share of good n in the aggregate (or average) expenditure pattern of country c; under ideal (but rarely realized conditions), it would coincide with the national accounts' estimate of the share of good n in total consumers' expenditure. The inflation factors, together with the strata and primary sampling units of the sample design, can be used to provide (asymptotic) sampling standard errors of the ratio estimate, for example using the **svyratio** command in STATA. These are the standard errors reported in Table 1.

Table 1 shows that rice is the single most important commodity in both countries, accounting for 11 percent of the total budget in India and 16 percent in Indonesia. Other goods with substantial shares of the budget in both countries include fish, sugar, chicken, moong beans, coconut oil, eggs, potatoes, cabbage, leafy greens, bananas, and beyond the food groups, cigarettes, electricity, and kerosene. There are also some sharp differences between the two countries; wheat and milk are important in India but negligible in Indonesia; soy products, particularly *tempe*, are important in Indonesia but do not exist in India, and therefore cannot be matched. A method for incorporating such goods is developed in Deaton, Friedman, and Alatas (2004) but, given the illustrative nature of the calculations here, we do not pursue these refinements here. Table 1 shows the expenditure patterns for urban and rural areas separately. Although we shall not follow the distinction further in this manual, there are in some cases substantial differences in expenditure patterns between urban and rural consumers. In particular, the share of rice in the budget is almost twice as large in the countryside as in the towns and cities. In the World Bank's poverty calculations, the rural and urban sectors of India (and China), but not Indonesia, are treated separately, and such would also be good practice also for the calculation of PPP's. However, there is nothing in principle wrong with working with national

average weights, provided it is clear that we are doing so, and in the poverty-weighted indexes discussed below, poor households will be given greater weight, so that if most of the poor live in the countryside, the rural sector will automatically be weighted in the appropriate way. What needs to be carefully avoided for poverty work, however, is the combination of *urban* prices (which are the only prices collected in some countries) with *national* or *urban* weights because, as Table 1 makes clear, such indexes may be quite misleading for the poor, particularly when cereal and food prices move differently from other prices, as happened in Indonesia during the Asian financial crisis.

The standard errors associated with the sampling are quite small in Table 1, typically a little less than one percent of the estimated budget shares. These standard errors will become smaller still when we compute price indexes, which are functions of linear combinations of many budget shares, and they will become larger when we focus only on poor households, which reduces the effective sample size.

Table 2 shows the median unit values for the most important 40 goods in the 63-good match. For each of these goods, the table lists its share of total expenditure in the total of matched goods, so that if all 63 goods were shown, the shares would add to unity. These are obtained by computing (24) for each country, but with N^* replaced by $N = 63$, and then averaging over the two countries; they are the weights that appear in the Törnqvist index (6). The second and third columns show the median unit values in each country, in rupiah and rupees respectively, while the last column shows the ratio of the two, which is the commodity-specific PPP exchange rate. Among the goods shown, these vary a good deal, from 761 rupiah per rupee for potatoes and 600 for green chillies, to 70 for coconut oil, and 88 for tea. Given the share of rice in the budget of

19

both countries, the rice PPP-exchange rate of 260 rupiah per rupee is a useful reference point, and will become even more important when we focus on the budgets of the poor. Note also that fuels, liquid petroleum gas (LPG), kerosene, and electricity, are all relatively cheap in Indonesia, an energy producing country whose domestic prices are held below world market prices. Rice, and to a lesser extent, wheat, sugar, and kerosene, are subsidized in India through "fair-price" shops, which make limited amounts of these goods available to households in some parts of India. The unit values in Table 2 incorporate these subsidies, by calculating the weighted average of the median unit values from the shops and the free-market.

In the calculations that follow, we treat the prices in Table 2 as if they were measured without error, so that when we calculate the standard errors of the price indexes, the variance comes in part from the sampling errors of the expenditures that serve as weights, and in part from the reconciliation of competing estimates of the PPP exchange rate, see equations (10), (15), (18) and (19) above. There is no contribution from sampling errors on the median unit values. While the unit values here are indeed sampled from the surveys, and have their own sampling errors, we do not incorporate them so as to replicate as closely as possible what would happen if we were combining prices collected by the ICP with the expenditures from the survey. There does not appear to be any useful framework for thinking about sampling error associated with the ICP prices, and so we here treat the unit values as if they were ICP prices, and measured without error. An important concern with the use of surveys in PPP work is whether the samples are large enough to support the accurate computation of index numbers, particularly when we focus on the poor. Sampling errors from prices, although likely real enough in the ICP, are no part of this concern, so that it makes sense to treat the prices as if they were known.

2.2 Survey formulas and standard errors

2.2.1 Laspeyres and Paasche

We start from the Laspeyres index (1), and rewrite it in a form that makes the calculations transparent, as well as how the standard errors are evaluated. We rewrite (1) in the form

$$P_L^{12} = \frac{\sum_{n=1}^{N} \bar{x}_n^1 (p_n^2 / p_n^1)}{\sum_{n=1}^{N} \bar{x}_n^1} = \frac{\sum_{h=1}^{H} w_h^1 \left(\sum_{n=1}^{N} x_{nh}^1 (p_n^2 / p_n^1) \right) / \sum_{h=1}^{H} w_h^1}{\sum_{h=1}^{H} \left(\sum_{n=1}^{N} x_{nh}^1 \right) / \sum_{h=1}^{H} w_h^1} \tag{25}$$

where an overbar denotes a sample average, and x denotes expenditure, so that x_{nh}^1 is the expenditure on good n by household h in country 1. Note that (25) can be interpreted as the ratio of the (weighted) average over Indian households of expenditure revalued at Indonesian prices to the weighted average of expenditures at Indian prices. Since the relative prices are taken to be measured without error, this is simply a ratio estimate, and given the Indian survey design, its (asymptotic) sampling standard error can be calculated in the usual way.

We can also calculate the variance from the stochastic approach using (17). Since the variance from the stochastic approach is conditional on the true budget shares, while the sampling variance comes from the sampling of the budget shares, we can also compute a total variance by adding the two variances. Which of the three estimates of variance is of interest depends on the interpretation of the indexes. If the Laspeyres is the object of interest, its sampling variance is all that is required. If the Laspeyres is regarded as an average designed to estimate the underling PPP, then the total variance is the relevant one. In both cases, the

21

sampling variance should be taken into account.

2.2.2 Fisher

For further calculations it is convenient to work in logarithms, noting that, asymptotically,

$$s.e.[\log P_L^{12}] = \frac{1}{P_L^{12}} s.e.[P_L^{12}] \tag{26}$$

Given that the logarithm of the Paasche index for Indonesia based on India is minus the

logarithm of the Laspeyres index for India based on India, we can calculate the Paasche index,

its logarithm, and standard error by applying (25) and then (26) to the Indonesian data. And

because the Indonesian and Indian surveys are independent of one another, we immediately have

the sampling variance of the log Fisher index as a quarter of the sum of the variances of the log

Laspeyres and the log Paasche, or indeed of the two log Laspeyres indexes, so that

$$V(\log P_F^{12}) = [V(\log P_L^{12}) + V(\log P_L^{21})]/4 \tag{27}$$

As before, the variance in (27) can be extended by incorporating the variance from the stochastic

approach, equation (14).

2.2.3 Törnqvist

The Törnqvist index (6) can be rewritten in terms of the survey data as

$$\ln P_T^{12} = \frac{1}{2} \frac{\sum_{h=1}^{H_1} w_h^1 \sum_{n=1}^{N} x_{nh}^1 \ln\left(p_n^2/p_n^1\right) \Big/ \sum_{h=1}^{H_1} w_h^1}{\sum_{h=1}^{H_1} \sum_{n=1}^{N} w_h^1 x_{nh}^1 \Big/ \sum_{h=1}^{H_1} w_h^1} + \frac{1}{2} \frac{\sum_{h=1}^{H_2} w_h^2 \sum_{n=1}^{N} x_{nh}^2 \ln\left(p_n^2/p_n^1\right) \Big/ \sum_{h=1}^{H_2} w_h^2}{\sum_{h=1}^{H_2} \sum_{n=1}^{N} w_h^2 x_{nh}^2 \Big/ \sum_{h=1}^{H_2} w_h^2} \tag{28}$$

where H_1 and H_2 are the numbers of sample households in each of the two surveys (which it has

not been necessary to distinguish until now.) Because the prices are being treated as known, the

first term depends only on the first survey, and the second only on the second survey, so the

sampling variance of (28) is the sum of the two variances. Each term is a ratio of means so that, once again, standard software will give the asymptotic variances given the design of each survey. Note that, as for the other indexes, we can also incorporate the additional term (14), the variance from the stochastic approach.

2.2.4 Country Product Dummy

The Country Product Dummy (CPD) index comes from running a weighted regression of the log prices on country and commodity dummies using the budget shares as weights, see (8) and (9). In (9), which is a GLS estimator, the sampling variability comes through the budget shares, so that, given the variance-covariance matrix of the budget shares from the surveys, we can calculate the sampling variance covariance matrix from (9) using the delta method, as described, for example, in Hayashi (2000, pp 93-94). The estimated budget shares for each country are calculated as

$$\widehat{s}_n^c = \sum_{h=1}^{H} w_h^c x_{nh}^c \Bigg/ \sum_{n=1}^{N} \sum_{h=1}^{H} w_h^c x_{nh}^c \tag{29}$$

where $c = 1, 2$ indicates the country. In equation (9), these budget shares are arranged into a diagonal matrix with N shares for country 1 followed by N shares for country 2. As usual, (29) is a ratio estimate, and its variance and covariances with the other commodity shares for the country can be calculated in the usual way. (Note that the **svyratio** command in STATA will return the variance covariance matrix of a list of survey ratios.)

Suppose that we denote s the $2N$ by 1 vector of stacked shares whose elements are estimated from (29). We then write

$$\Omega = E(\widehat{s} - s)(\widehat{s} - s)' \tag{30}$$

23

for the variance-covariance matrix. Note that, because the surveys are independent of one another, this is a block diagonal matrix. (In the case where there are M countries, Ω will be an NM by NM block diagonal matrix.) In the Appendix, I show that application of the delta-method to (9) gives a variance covariance matrix of the estimates

$$\hat{V}_\beta = (X'SX)^{-1} X'E\Omega EX(X'SX)^{-1} \tag{31}$$

where E is a diagonal matrix whose diagonal elements are the $2N$ residuals e defined by

$$e = y - X\beta \tag{32}$$

Paralleling the discussions of the Fisher and Törnqvist above, (31) captures the sampling variance. If we add the variance from the stochastic approach, (10), we get the total variance-covariance matrix for the CPD index

$$\tilde{V}_\beta = (X'SX)^{-1} X'[E\Omega E + S\Sigma S]X(X'SX)^{-1} \tag{33}$$

where, once again, $\Sigma = E^2$ is a diagonal matrix with the squared residuals on the diagonal.

Referring back to equation (8), in which the country dummies appear ahead of the commodity dummies in the CPD regression, and assuming that a constant term comes first, the variance of the logarithm of the PPP index for country two in terms of country one will be element (2,2) of (31) or (33).

2.3 Illustrative PPPs and standard errors

Table 3 lists the five index numbers, together with estimates of their sampling errors, their stochastic-approach standard errors, and the total standard errors. The Paasche and Laspeyres, at 187 and 255 rupiah per rupee, are far apart, while the two superlative indexes, as well as the approximately superlative CPD, are quite close, 218 (Fisher), 216 (Törnqvist), and 219 (CPD).

The fact the price indexes are averages reduces the relative sampling errors below what they were for the average (aggregate) budget shares. While in Table 1 the sampling standard errors were around one percent of the estimates, for the PPPs themselves in Table 3, the standard errors are between 0.15 and 0.30 percent of the estimates. At least for these aggregate national indexes, these sampling errors are effectively negligible.

The same is not true for the standard errors that come from the stochastic approach, and which reflect the deviation of the individual prices from the law of one price. These standard errors are around ten percent of the PPPs for the three relevant indexes. Again, we emphasize that these standard errors do not reflect estimation uncertainty, or sampling, but rather the fact that, as we see in Table 2, the relative prices of the 63 goods are different in the two countries, so that there are many different commodity-specific PPPs over which we are averaging. Note that the "stochastic approach" standard errors for the Törnqvist and Fisher indexes are different because the indexes themselves are different; the estimated stochastic standard errors for their logarithms are identical, see equations (14) and (26) above. The stochastic standard errors dominate the sampling standard errors, so that the totals are essentially identical to the former.

3. Poverty-based indexes: PPPPs

3.1 Framework: budget share weighted indexes

In the theory and calculations so far, we have used national aggregate (or average) quantities to weight the prices to derive the various indexes. The national consumption patterns will often be different from the consumption patterns of the poor. In particular, aggregate consumption is affected more by those who consume more, who are the rich, not the poor, and goods that are

important to the poor, and comprise large shares of their budgets, will be down-weighted in the national averages if they are not similarly important to those who spend more. In this section, we show how to tailor the PPP indexes to the spending patterns of the poor, and carry forward the two country illustrative calculations to that case in Section 4. We shall maintain the assumption that everyone in the country faces the same prices, so that the poverty PPPs will differ from the national PPPs only in the weights that are used to aggregate the prices, not in the prices being aggregated.

When thinking about price indexes for different income groups, it is useful to work with budget shares and prices, rather than quantities and prices. All of the price indexes that we have considered can be thought of as functions of budget shares on the one hand, and relative prices, on the other. Given that we are assuming that the relative prices are the same for everyone, the effects of income on the price indexes work entirely through the budget shares. We therefore start by restating the various index numbers in terms of budget shares s_n^c defined as the fraction of expenditure in country c spent on good n. For the moment, we leave undefined whose budget, whether an individual, or an average of a group. The price indexes then take the forms

$$P_L^{12} = \sum_{n=1}^{N} s_n^1 \left(\frac{p_n^2}{p_n^1} \right) \tag{34}$$

for the Laspeyres.

$$P_P^{12} = \left[\sum_{n=1}^{N} s_n^2 \left(\frac{p_n^1}{p_n^2} \right) \right]^{-1} \tag{35}$$

is the Paasche, and the Fisher is their geometric mean. The Törnqvist, (6), is already in the budget share form

$$\ln P_T^{12} = \frac{1}{2} \sum_{n=1}^{N} (s_n^1 + s_n^2) \ln \frac{p_n^2}{p_n^1} \tag{36}$$

while the weighted CPD index, as we have seen, can be written as

$$\ln P_{CPD}^{12} = \sum_{n=1}^{N} \tilde{s}_n \ln\left(\frac{p_n^2}{p_n^1}\right) \qquad (37)$$

where $\tilde{s}_n$ is the normalized harmonic mean budget share defined in (20) and (21) above.

In the indexes of sections 1 and 2, the budget share of good n was taken to be the share in national aggregate expenditure

$$s_n^c = \frac{\sum_{h=1}^{H} p_n^c q_{nh}^c}{\sum_{k=1}^{N}\sum_{h=1}^{H} p_k^c q_{kh}^c} \qquad (38)$$

where q_{nh}^c is the quantity of good n purchased by household h in country c. In the subsections below, we explore alternative definitions of the budget shares that will allow us to construct price indexes that are more relevant for the poor.

3.2 Three weighting schemes

We shall modify the budget shares (38) in three different ways. First, it is useful to think about budget shares for each household individually, and then to compute a simple average in which each household is given equal weight. We write the budget share for household h as

$$s_{nh}^c = \frac{p_n^c q_{nh}^c}{\sum_{k=1}^{N} p_k^c q_{kh}^c} \qquad (39)$$

then the "democratic" budget share is simply

$$s_n^{Dc} = \frac{1}{H}\sum_{h}^{H} s_{nh}^c \qquad (40)$$

Note that we can write the national budget shares (38) in terms of the individual household budget shares (40) in the form

$$s_n^c = \sum_{h=1}^{H} x_h s_{nh}^c \bigg/ \sum_{h=1}^{H} x_h \qquad (41)$$

where x_h is total expenditure on all N goods by household h. Whereas the democratic budget shares (40) are simple averages over all households, the "plutocratic" budget shares (38) and (41) are weighted by the total expenditure of each household, thus biasing them towards the expenditure patterns of better off households, Sigbert Prais (1959). Note also that the sum in (40) is taken over *all* households, whether or not they purchase the good. A household that does not purchase good n has a budget share of zero, and this zero must be included in the average. Democratic price indexes are averages of price indexes over all or groups of households; a household that does not consume a good is unaffected by a change in its price, and that zero effect should be included along with the effects for households that do consume.

The insertion of these democratic weights into (34) through (37) yields a system of democratic PPP indexes for country 2 relative to country 1.

While democratic PPPs are not biased towards the rich, they include *all* households, poor and non poor alike. We can adapt (40) to cover *only* the poor by averaging over only the poor. If we denote the set of poor households by $\wp$, so that we can write $h \in \wp$ to mean that members of household h are in poverty. If H_p is the number of households in poverty, the "poverty" budget shares are defined by

$$s_n^{Pc} = \frac{1}{H_p} \sum_{h \in \wp}^{H_p} s_{nh}^c \qquad (42)$$

Note that the poverty budget shares in (42) count each poor *household* equally, and it might be argued that this is inappropriate, and that each poor *person* should be counted equally. If so, the simple average in (42) should be replaced by a weighted average, in which the weights are the numbers of persons in each household.

Using the poverty budget shares to calculate PPPs will give indexes that are appropriate for comparing the living standards of the poor in different countries. However, one of the main uses of the PPPs, and perhaps the most important use within the World Bank, is to convert the international poverty line into its local currency equivalent. For this purpose, we need, not the budget shares of all of those *below the poverty line*, but the budget share of those *at the poverty line*. Of course, in practice, any given survey will have few or no households *exactly* at the line, so we follow a standard procedure for this sort of case, which is to calculate a weighted average of the budget share of households *near* the poverty line, in which the weights are larger the closer is the household to the poverty line. Such "kernel"-weighted averages are familiar from nonparametric regression and density estimation in statistics and are constructed as follows.

Define the kernel weight $\omega_{h\tau}(z)$ for household h by

$$\omega_{h\tau}(z) = \frac{1}{\tau} K\left(\frac{\ln(x_h/n_h) - \ln z}{\tau}\right)$$
(43)

where z is the poverty line, defined as the level of per capita total expenditure x_h/n_h, for household size n_h, that is the cutoff for poverty. The function $K(.)$ is a "kernel" function; it integrates to unity, is non-negative, symmetric around zero, and decreasing in the absolute value of its argument. (It can be thought of as a density function, and indeed the standard normal density is often used as a kernel weighting function.) The parameter τ is a "bandwidth," that is ideally set to optimize the trade-off between bias (too large a bandwidth) and variance (too small a bandwidth). If τ is small, only households near to the poverty line will receive much weight while, if it is large, more households will be included. Ideally, we would like τ to be zero, using

only households *at* the line, but this will only work with infinite samples. In general, τ will be smaller the larger is the sample and, in practice, allows us to trade-off precision (by including more households) and closeness to the poverty line (for relevance, and lack of bias.) In general, τ is set to be inversely proportional to the negative fifth power of the data, and will vary directly with its standard deviation.

Note also that we work with the logarithms of total per capita expenditure, which is typically distributed more like a symmetric normal distribution than is per capita expenditure itself. Given that the kernel is symmetric about the bandwidth, the transformation will reduce possible problems of having very different numbers of households on either side of the poverty line. Note that, as the bandwidth goes to zero, the transformation will have no effect.

One convenient kernel function that we shall use in the calculations below is the "biweight" function

$$K(t) = \frac{15}{8}\left(1-t^2\right)^2 \quad \text{for} \quad -1 \le t \le 1$$
$$= 0 \quad \text{otherwise}$$

(44)

The poverty-line budget shares are then defined by the kernel-weighted local average

$$s_n^c(z^c) = \frac{\sum_{h=1}^{H} \omega_{h\tau}^c(z^c) s_{nh}^c}{\sum_{h=1}^{H} \omega_{h\tau}^c}$$

(45)

3.3 Reconciling the poverty lines

Calculation of a democratic PPP index proceeds in exactly the same way as the calculation of the plutocratic indexes in Sections 1 and 2, with the democratic budget shares substituted for the plutocratic shares. The same is true for the two poverty PPPs provided that the poverty lines for both countries are known. Such calculations will be appropriate if we have well-established

30

poverty lines in each country, and we want to calculate a PPP exchange rate that compares prices

for the poor in both countries. However, if the PPP rate is to be used to convert the international

line into local currencies, the two poverty lines should be the same when converted into a

common currency using the poverty-line PPP exchange rate. This requires that

$$PPP^{12} = \frac{z^2}{z^1} \tag{46}$$

where *PPP* stands for whichever of the indexes we choose. For example, when (46) holds, and if

z^1 is the Indian poverty line in rupees, then $z^1 PPP^{12}$ is the Indian poverty line converted into

rupiah using the PPP exchange rate in rupiah per rupee, and this will be equal to the Indonesian

poverty line z^2. If we are assuming that tastes are the same in the two countries, and that one of

the (superlative) PPP indexes as an approximation to the cost-of-living index, the poverty lines

in country *c* would satisfy

$$c(u_z, p^c) = z^c \tag{47}$$

for tastes represented by the cost function $c(u, p)$, and with poverty line utility level u_z, so that

the poverty-line cost-of-living PPP index number would be

$$PPP^{12} = \frac{c(u_z, p^2)}{c(u_z, p^1)} \tag{48}$$

In the two country case, we start with two poverty lines, which we use to calculate the two

sets of poverty-line budget shares, (45), which are then used in the usual way to compute a trial

PPP. This PPP is then used to adjust the poverty line z^2 using (46), and to calculate a new set of

poverty-line budget shares for country 2, and so on until convergence. Alternatively, we could

take the poverty line of country 2, z^2 as fixed and adjust the poverty line and budget shares for

country 1. (In the *M* country case, a poverty line in one country is taken as base—$1 a day, for

example—and the others adjusted around it.) If we take country 1's poverty line as base, and we are working with the Törnqvist index to illustrate, the iteration is described by

$$\ln P_{Tr}^{12} = \frac{1}{2} \sum_{n=1}^{N} \left[s_n^1(z^1) + s_n^2(P_{Tr-1}^{12} z^1) \right] \ln \frac{p_n^2}{p_n^1} \tag{49}$$

where r is the iteration number.

For at least one specific Engel curve, it is possible to solve explicitly for the PPP price index. Consider, for example, the case where the budget shares in each country are linear functions of the logarithm of total expenditures, a functional form that often fits the data well, and that is consistent with choice theory, see for example, Deaton and Muellbauer (1980, Chapter 3.)

$$s_{nh}^c = \xi_{0n}^c + \xi_{1n}^c \ln x_h + \varepsilon_{nh}^c \tag{50}$$

where c is the country, here 1 or 2, ε_{nh} is a disturbance term, and ξ_{0n}^c and ξ_{1n}^c are commodity- and country-specific parameters. Over all of the goods in the budget, and for each country, the ξ_{1n}^c parameters add to zero, and the ξ_{0n}^c parameters to one. If we evaluate (50) at the poverty line for each country, and then average over all households at the poverty line, the poverty-line Törnqvist index can be written

$$\ln P_T^{12} = \frac{\sum_{n=1}^{N} \left(\xi_{0n}^1 + \xi_{0n}^2 + (\xi_{1n}^1 + \xi_{1n}^2) \ln z^1 \right) \ln \frac{p_n^2}{p_n^1}}{2 - \sum_{n=1}^{N} \xi_{1n}^2 \ln \frac{p_n^2}{p_n^1}} \tag{51}$$

provided that the disturbance terms in (50) have expectation zero over households at the poverty line.

At the very least, (52) provides a convenient starting point for the iterative calculation of the Törnqvist using (49), and indeed for the other poverty-weighted PPP indexes. In general, the iterative procedure is likely to be preferred, because it does not make any assumptions about the

functional form of the Engel curves, and because, as the sample sizes go to infinity, the kernel smoothing near the poverty line will converge to the actual average at the poverty line. Of course, in finite samples, the kernel smoothing requires its own assumptions, so that there is no guarantee that it will be superior to (51) in practice.

3.4 Estimation of standard errors

Estimation variances for the PPPs described in this section can be calculated in the same way as for the PPP indexes in Sections 1 and 2. This is true both for the sampling variances and for the variances associated with the stochastic approach. Note that when calculating variances and covariances for the kernel-weighted budget shares (45), we have once again a ratio estimator, in this case the average of the budget shares multiplied by the kernel weight, divided by the average of the kernel weights themselves.

4. Illustrative calculations for two country poverty PPPs

We take India as the base country whose poverty line will be taken as fixed, so that our aim is to calculate poverty-weighted purchasing power parity index numbers for people in India who are close to this poverty line, compared with people in Indonesia who are close to an Indonesian poverty line that is the Indian poverty line multiplied by the purchasing power parity exchange rate for the poor. The government of India publishes two poverty lines for 1999–2000, 327.56 rupees per person per month for rural areas, and 454.11 rupees per person per month for urban areas. At these poverty lines, 26.1 percent of the All Indian population is poor. We set our base Indian poverty line at 345.08 rupees per person per month which is selected so that, if the line is

applied to the whole Indian population regardless of sector, we replicate the official 26.1 percent poverty rate.

The calculations start by using the individual budget share data to estimate regressions on the logarithm of household total expenditure per capita, i.e. (50), for each of the 63 goods in each country, The estimated parameters are then used to calculate (53), which is an approximation to the poverty-weighted Törnqvist index, and will serve as a starting value for the iterative calculations. This yields a rupiah to rupee exchange rate of 229.52, which is 6.4 percent higher than the corresponding plutocratic Törnqvist index, which is itself close to the Fisher and CPD indexes. As we shall see below, the non-parametric iterative calculation changes this number hardly at all.

I next calculate the fully consistent poverty-line PPPs. The Indian poverty line is held fixed. At the first iteration, for all three indexes, Fisher, Törnqvist, and CPD, I calculate a trial Indonesian poverty line as the Indian line (345.08) multiplied by the Törnqvist approximation (229.52), and then calculate kernel weighted budget shares around both poverty lines. These budget shares are then used to calculate new estimates of the three indexes, which are then applied to the Indian line to get a second estimate of the Indonesian poverty line. And so on. In all cases, there was no further change in the first eight digits of the index within five iterations.

Table 4a shows the three indexes for five different values of the bandwidth, together with their standard errors associated with the sampling. The standard errors associated with arbitrage failure are similar to those shown earlier for the plutocratic case, and so are not shown here. As in the earlier case, they are much larger than the sampling standard errors. Nevertheless, we focus on the latter to investigate concerns that, by focusing on households near the poverty line,

34

we reduce the effective sample sizes so much that the indexes cannot be reliably estimated.

The widest bandwidth is one standard deviation of the logarithm of per capita expenditure, and I also calculated indexes with the bandwidth set at 0.5, 0.1, 0.05, and 0.01 standard deviations. In any given index calculation, the bandwidth is the same multiple of the standard deviation for both countries although, of course, the two standard deviations are different. Table 4b shows the number of observations from each survey that fall within the band around the poverty line; this shows how many sample households are actually used in calculating the indexes. As is to be expected, the sampling standard errors in Table 4a are larger the smaller is the bandwidth, and are (up to their own sampling errors) inversely proportional to the square root of the sample sizes in the second panel.

Compared with Table 3, the poverty-weighted indexes in Table 4a are about 5 percent larger than the plutocratic indexes, presumably because of the greater weight of rice in the budgets of the poor, and given that the commodity-specific exchange rate for rice is 260, which is substantially higher than any of the overall indexes. With the widest bandwidth, the sampling standard errors are smaller than those of the plutocratic indexes, even though the latter uses a larger sample of households. This reflects the larger sampling standard errors of the plutocratic compared with democratic or poverty-weighted budget shares.

Perhaps the most encouraging finding in Table 4a is that the sampling standard errors are so small. Even when we use a bandwidth around the poverty line of one percent of a standard deviation of log per capita household expenditure, within which there are only 700 and 400 observations in the Indian and Indonesian samples, respectively, the sampling standard error is less than one percent of the estimated price index. This finding holds out good hope of being

35

able to calculate similar PPPs from much smaller household surveys than those available for India and Indonesia. In large samples, such as the present, the PPPs can be very sharply focused around the poverty line. In India, one percent of a standard deviation of log per capita expenditure is 0.0058, so that the index is computed using only those households whose per capita expenditure is within 0.58 percent of the poverty line. In Indonesia, one percent of the standard deviation is 0.0054.

Table 5 shows budget shares and standard errors for the two countries for 20 of the most important goods. The table shows, first, the plutocratic or aggregate budget shares as in Table 2, followed by the poverty-line budget shares, calculated for the smallest bandwidth considered in Table 4a. As anticipated, the major difference between the plutocratic and poverty-line budget shares is in the share of rice, which is 50 percent (India) and 30 percent (Indonesia) higher among the poor than its share in aggregate expenditure would indicate. There are corresponding reductions in shares devoted to a number of other foods, as well as to electricity, and to cigarettes, for example. (Note that we are here working only with 63 covered goods, most of which are food, so that the greater food share among the poor has relatively little effect here compared with what would be the case when we have a full set of prices for all goods, as will be the case when the ICP prices are merged with household survey budget shares.) Given the very small bandwidth, the sampling standard errors are quite large for some of the commodities. For example, the standard error for rice in India is about four percent of the estimated value, and in Indonesia , about 2 percent. However, once the budget shares and relative prices are averaged into index numbers, we get down to the one percent standard errors in Table 4a.

5. Poverty PPPs (PPPPs) for three or more countries

I have left the multicountry case until last, because it enabled me to discuss other issues first,

such as the basic ideas of PPP exchange rates, and how to extend them to poverty-weighted

indexes. These ideas go through to the many country case, but there are additional issues that

need to be dealt with. In this section, I start with the easiest case, that of the CPD index, where

the previous analysis requires minimal extension. In Section 5.2, I then turn to the Törnqvist and

Fisher indexes, which have to be modified for use in the many country case. Section 6 provides

illustrative calculations for a set of poverty-weighted PPP indexes for India, Indonesia,

Bangladesh, and Pakistan.

5.1 Poverty-weighted CPD indexes for many countries

In the two country case, the CPD index was obtained by running a weighted regression (8) of the

logarithms of prices in the two countries on country dummy and a set of product dummies. This

regression extends immediately to the case of M countries. Instead of (8), the regression is

$$\ln p_n^c = \alpha + \sum_{c=2}^{M} \theta^c + \sum_{n=1}^{N} \beta_n \qquad (52)$$

 so that, with M countries, there are $M-1$ country dummies, together with the N commodity

dummies, which also appeared in the two country case. The omitted country, labeled 1, is the

base country whose currency is used as the denominator for the exchange rates; the particular

choice makes no difference, except to scale all of the estimated exchange rates by the same

amount. As before, the regression is estimated with the budget shares of good n in country c as

weights, so that (52) is a regression of MN observations on $M+N-1$ zero/one dummy variables

plus a constant term. The estimated θ coefficients provide the logarithms of the PPP exchange

37

rates for each country in terms of the currency of the base country.

An important and recurrent issue when dealing with many countries is what happens if not all of the prices are available for all of the countries, something that is nearly always the case, whether dealing with prices collected by the ICP itself, or prices from household surveys. Regional commodities like tofu, or fermented fish paste, are common and important in the budgets of people in several countries in South East Asia, but do not appear at all in the budgets of people in India or Pakistan. In two country comparisons, these commodities must simply be dropped from the comparison. But with many countries, prices can be included when they are available, and dropped when they are not, while still affecting the estimated PPPs. So that instead of running (52) with MN observations, the regression may have substantially fewer observations, though there will always be a commodity dummy on the right hand side for any commodity that appears in at least two countries. (When it only appears in one country, the inclusion of the commodity dummy has the same effect as dropping it from the regression.)

Poverty-weighting works in exactly the same way as before. Given a poverty line for the base country, and a starting set of poverty-weighted PPPs—I shall develop the analog of the two-country starting value procedure below—we calculate a corresponding poverty line for each of the other countries. These poverty lines are then used to calculate budget share weights for people near the poverty line in each country, and those budget shares, together with the poverty-weighted budget shares for the base country, are then used as the weights for running (52). This gives a new set of poverty-weighted PPPs, a new set of poverty lines, and so on until convergence. There is no new principle beyond what was done in the previous section.

Standard errors in the many country case are calculated exactly as in Section 2.2.4 above.

Each country's poverty-weighted budget shares have sampling standard errors that are calculated as in Section 4, and then these are plugged into the equations in 2.2.4 to give the sampling variances (and failure of arbitrage variances) of the PPPP indexes. Once again, no new issues arise.

5.2 Poverty-weighted Törnqvist and Fisher-type indexes for many countries

With M countries, we can calculate a Fisher (or some other superlative) index between each pair, i and j say. For the moment, denote these indexes as π^{ij} forming an M by M matrix. Not all of the elements of this matrix are independent of one another. In particular, the diagonal elements are all unity and, because the indexes are superlative, they satisfy the country reversal property that $\pi^{ij} = 1/\pi^{ji}$. Nevertheless, there are still $M(M-1)/2$ distinct indexes in the matrix. When M is bigger than 2, this is many more index numbers than is needed for a system of exchange rates.

In a system of exchange rates for M countries, we need only $M-1$ exchange rates or indexes. For any two countries, the exchange rate between them needs to be the such that we get the same amount of j's currency for a unit of i's currency whether we make the transaction directly, or whether we first buy the currency of any third country, and then convert that. (These are imaginary conversions and so carry no transactions costs!) Put another way, if the price level in country 2 is twice what it is in country 1, and the price level in country 3 is twice that in country 2, then the price level in country 3 must be 4 times the price level in country 1. For this to happen, we need to have that, for all i, j, and k,

$$\pi_{ij} = \pi_{ik}\pi_{kj} \tag{53}$$

a condition that is not automatically satisfied by either the Törnqvist or Fisher indexes.

To see how we can modify the indexes to satisfy (53), it is convenient to work with the logarithms of the price indexes. We write

$$b^{ij} = \ln \pi^{ij} \qquad (54)$$

so that (53) becomes

$$b^{ij} = b^{ik} + b^{kj} \qquad (55)$$

for all i, j, and k. Suppose that country 1 is the base country, whose currency we are using as a denominator for all of the PPPs. Then because (56) holds for all k, it holds for country 1, so that (56) implies that

$$b^{ij} = b^{i1} + b^{1j} = -b^{1i} + b^{1j} \equiv -a^i + a^j \qquad (56)$$

where I have used the country reversal property that $b^{ij} = -b^{ji}$ and have defined the a^i to be equal to b^{1i}. Because the a's are the logarithms of the price level in each country relative to the base country 1, they are the (logarithms) of the PPPs that we are looking for. Once (56) is satisfied, with $b^{ij} = a^j - a^i$ for some set of a's, (55) will be satisfied, and we will have what we want, a system of exchange rates with only $M - 1$ distinct numbers.

The simplest way to adjust the b's to satisfy (56) is to choose a's to make (56) as good an approximation as possible in a least squares sense, and idea that was first propose in this context by Corrado Gini (1924). If we choose a^i to minimize

$$\phi = \sum_{k=1}^{M} \sum_{j=1}^{M} (b^{kj} - a^k - a^j)^2 \qquad (57)$$

we get first-order conditions

$$a^i = -\frac{1}{M} \sum_{j=1}^{M} b^{ij} + \frac{1}{M} \sum_{j=1}^{M} a^j \qquad (58)$$

Inspection of (58) shows that, given the b's, there is no unique solution for the a's: adding a constant to each of a's does not change (58). The obvious normalization is to set $a^1 = 0$, which

makes the PPP exchange rate is 1 for the numeraire country, so that, by (58), we have

$$\frac{1}{M}\sum_{j=1}^{M}a^{j} = \frac{1}{M}\sum_{j=1}^{M}b^{1j} \tag{59}$$

Substituting (60) into (59) gives us, finally,

$$a^{i} = \frac{1}{M}\sum_{j=1}^{M}(b^{1j} + b^{ji}) \tag{60}$$

we can interpret (60) as an average of the M two-step price indexes that we get by going from

the base country to country i through each of the other countries in turn. Alternatively, we can

exponentiate (60) and write

$$P_{i} = \left(\prod_{j=1}^{M}\frac{\pi^{ji}}{\pi^{j1}}\right)^{\frac{1}{M}} \tag{61}$$

We can interpret (61) in terms of our Indian and Indonesian example. The base country is

India, and we want to compute an index for Indonesian prices relative to Indian prices, in rupiah

per rupee. We can do this directly, as we have done in the previous section. We can also do it

indirectly, for example using Bangladesh taka, so that we compute a price index for Indonesia

base Bangladesh in terms of rupiah per taka, and a price index for Bangladesh base India in

terms of taka per rupee. The ratio of these two, which has units of rupiah per rupee, is another

measure of the Indonesian price level base India. We get as many such rupiah to rupee indexes

as there are countries (note that the own rate is automatically unity), and then take the geometric

mean of all of them to give us a PPP index that behaves in the way that we need.

If we first compute the Fisher indexes and use them in (62), the resulting index is known as

the EKS index, after Eltetö and Köves (1964) and Szulc (1964), but the credit really belongs to

Gini. But as we have seen, the calculation will work for any superlative price index, such as the

Törnqvist, not just the Fisher. Indeed, I will refer to any PPP calculated using (60) or (61) and

41

based on a superlative index as an EKS index, and I will compute EKS indexes from both the Fisher and the Törnqvist indexes.

To compute the index, we need a complete set of superlative indexes linking each country to every other. However, as was the case for the CPD index, we do *not* necessarily require that all commodities be consumed in all the countries because the Fisher (for example) indexes for any pair of countries can be computed from a different set of overlapping goods than would be the case for the Fisher for another pair of countries.

Poverty-weighting is straightforward, but we must begin with a set of Fisher or Törnqvist indexes that are themselves poverty-weighted. Given a trial set of PPPs, and a base country poverty line, we have a trial set of poverty lines that can be used to calculate the poverty-weighted Fisher or Törnqvist indexes which are then plugged into (62) in order to give new PPPs, and so on.

5.3 The closed-form EKS-Törnqvist poverty-line PPP

If we assume that the Engel curves in each country have the functional form (50), with different parameters for each country, then it is possible, one again, to derive an explicit solution for the poverty-line PPP, at least for the EKS-Törnqvist. From (50), and using the fact that the logarithms of the PPP exchange rate for country c relative to country one is a^c, the average budget shares at the poverty line take the form

$$s_n^c = \xi_{on}^c + \xi_{1n}^c \ln z + \xi_{1n}^c a^c \tag{62}$$

where, as before, z is the international poverty line, or poverty line in the base country 1. The log Törnqvist index for j relative to i, b^{ij}, is then

$$b^{ij} = \frac{1}{2} \sum_{n=1}^{N} \left((\xi_{on}^{i} + \xi_{on}^{j}) + (\xi_{1n}^{i} + \xi_{1n}^{j}) \ln z + \xi_{1n}^{i} a^{i} + \xi_{1n}^{j} a^{j} \right) \ln \frac{p_{n}^{j}}{p_{n}^{i}} \tag{63}$$

Or, with appropriate definition of terms

$$b^{ij} = \psi^{ij} + \phi^{ij} a^{i} + \theta^{ji} a^{j} \tag{64}$$

Finally, the system is closed by (60), the formula that links the EKS PPP indexes a^{i} to the

matrix of original indexes b^{ij}, which gives

$$a^{i} \left(1 - \frac{1}{M} \sum_{j=1}^{M} \theta^{ji} \right) - \frac{1}{M} \sum_{j=2}^{M} (\theta^{1j} + \phi^{ji}) a^{j} = \frac{1}{M} \sum_{j=1}^{M} (\psi^{1j} + \psi^{ji}) \tag{65}$$

where I have used the fact that $a^{1} = 0$. Equation (65) is a system of $M-1$ linear equations in the

$M-1$ unknown poverty-line PPPs.

5.4 Standard errors for EKS-Fisher and EKS-Törnqvist indexes from survey data

Equation (60) gives the log EKS PPP rates as a function of the underlying log Fisher or

Törnqvist indexes, whose standard errors we have already discussed. So the only new issue is to

deal with the covariances between the various fundamental indexes in (60). The Törnqvist is the

most straightforward case, so I deal with it first.

We write the Törnqvist index for country j with country i as base in the form

$$b^{ij} = \frac{1}{2} \sum_{n=1}^{N} (s_{n}^{i} + s_{n}^{j}) t_{n}^{ij} = \frac{1}{2} (s^{i} + s^{j}).t^{ij} \tag{66}$$

where

$$t_{n}^{ij} = \log(p_{n}^{j} / p_{n}^{i}) \tag{67}$$

is the vector of the logarithms of the price ratios, and the second term in (66) is a dot product. If

we substitute (63) into (60), we get the logarithm of the Törnqvist indexes

$$a^{i} = \frac{1}{2M} \sum_{j=1}^{M} \left[s^{1}.t^{1j} + s^{j}.(t^{1j} + t^{ji}) + s^{i}.t^{ji} \right] \tag{68}$$

The variance of these log PPPs is driven by the sampling variances of the vectors of budget

shares. Because the survey in each country is drawn independently of one another, there are no cross-country covariances. If we denote the sampling variance-covariance matrix of the budget shares for each country by the N x N matrix V^i, (68) gives an exact sampling variance for the log PPPs as, for i running from 2 to M

$$
\begin{aligned}
4M^2 V(a^i) = (1 - 2\delta_{i1}) \sum_{j=1}^{M} \sum_{k=1}^{M} t^{1j}.V^1 t^{1k} + 2\sum_{j=1}^{M} t^{1j}.V^1 t^{1i} + 2\sum_{j=1}^{M} t^{ji}.V^i t^{1i} \\
+ \sum_{j=1}^{M} \sum_{k=1}^{M} t^{ji}.V^i t^{ki} + \sum_{j=1}^{M} t^{1i}.V^j t^{1i}
\end{aligned}
\tag{69}
$$

The expressions in (69) of the form $x.Vx$ demote quadratic forms in the various country variance covariance matrices. The quantity δ_{i1} is the Kronecker delta and is unity when $i = 1$ and zero otherwise. As is easily checked, the variance of the baseline PPP, for which $a^1 = 0$, is zero.

The pairwise Fisher indexes are not linear functions of the budget shares, so that we need to take an approximate, linearized, approach. The pairwise log Fisher indexes can be written

$$
b^{ij} = \frac{1}{2} \left(\ln s^i.r^{ij} - \ln s^j r^{ji} \right)
\tag{70}
$$

where

$$
r_n^{ij} = p_n^j / p_n^i
\tag{71}
$$

is the vector of price ratios. Equation (67) gives the log Fisher as the average of the log Laspeyres and the log Paasche. The log EKS index (61) is then given by

$$
a^i = \frac{1}{2M} \sum_{j=1}^{M} \left[\ln s^1.r^{1j} - \ln s^j.r^{j1} - \ln s^i.r^{ij} + \ln s^j.r^{ji} \right]
\tag{72}
$$

We then apply the delta-method to derive the approximate variances. Define the vectors whose nth element, $n = 1, \ldots, N$, is given by

$$
\beta_n^{ij} = \frac{r_n^{ij}}{w^i.r^{ij}}
\tag{73}
$$

If we follow through the algebra in a way that closely mirrors the calculations for the variance of

the Törnqvist index in (70), we eventually reach

$$4M^2V(a^i) = (1-2\delta_{1i})\sum_{j=1}^{M}\sum_{k=1}^{M}\beta^{1j}.V^1\beta^{1k} + \sum_{j=1}^{M}\sum_{k=1}^{M}\beta^{ij}.V^i\beta^{ik} + \sum_{j=1}^{M}(\beta^{ji}-\beta^{j1}).V^j(\beta^{ji}-\beta^{j1})$$

$$+2\sum_{j=1}^{M}\beta^{1j}.V^1(\beta^{1i}-\beta^{11})-2\sum_{j=1}^{M}(\beta^{ii}-\beta^{i1}).V^i\beta^{ij}$$

(74)

As before, we can compute variances that reflect the failure of arbitrage, or that the PPPs are compromise indexes that "average" over many possible candidate indexes. We follow Section 1.6 in deriving CPD-based variances for the modified Törnvist and Fisher (EKS) indexes. Again, the Törnqvist is the more straightforward. From the CPD formulation, equation (8), the pairwise log Törnqvist indexes are

$$b^{ij} = \alpha^j - \alpha^i + \frac{1}{2}\sum_{n=1}^{N}(s_n^i + s_n^j)(u_n^j - u_n^i)$$

(75)

The log Törnqvist PPPs are, from (60),

$$a^i = \alpha^i + \frac{1}{2}\sum_{j=1}^{M}(s^i + s^j).(u^i - u^j) + \frac{1}{2}\sum_{j=1}^{M}(s^i + s^1).(u^j - u^1)$$

(76)

I assume that the variances of the u^i satisfy

$$E(u^i u^j) = \delta_{ij}\Omega^i$$

(77)

where δ_{ij} is the Kronecker delta and, if we follow our previous practice, Ω^i is a diagonal matrix whose elements are replaced in the calculations by the squares of the residuals for that country from the CPD regression. The variance of (77) is then obtained by squaring the expression for $a^i - \alpha^i$ in (76), and taking expectations using (77) and treating the vectors of budget shares as fixed. This "failure-of-arbitrage" variance of the EKS-Törnqvist index takes the form

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

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

(78)

In Section 1.6, we showed that, given the CPD model, the Fisher and Törnqvist indexes were

identical to the first-order of approximation so that, to this order, the "failure of arbitrage" variance of the log PPP from the modified Fisher, the EKS-Fisher index, is the same as that of the modified Törnqvist. This result holds here too, so that (78) serves as the "failure of arbitrage" variance for both the EKS-Fisher and the EKS-Törnqvist. And as was the case with two countries, the sampling variances and the "failure of arbitrage" variances are independent, so that the "total" variance of the EKS-Törnqvist is given by the sum of (69) and (78), and that for the EKS-Fisher is given by the sum of (74) and (78).

6. Illustrative calculations for India, Indonesia, Bangladesh, and Pakistan

To illustrate the multiple country calculations, I now add data from Pakistan and Bangladesh to those from India and Indonesia. The Pakistan data come from the Pakistan Integrated Household Survey of 1998-99, while those for Bangladesh come from the Household Income and Expenditure Survey of 2000. Both of these are substantially smaller samples than either the Indian or Indonesian surveys; Pakistan has just over 16,000 households, Bangladesh 7,440 households. Moving from two countries to four makes it more difficult to find commodities that are represented in all of the surveys, and for which there exists both price and expenditure information. For example, the Bangladeshi survey, although it contains expenditure data on fuels, does not contain information on quantities, so that it is not possible to use the survey to calculate unit values for use as prices. For the more substantive ICP analysis, for which this is a template, this will not be a problem, because the prices are collected separately, and the expenditure information, which comes from the survey, covers the complete budget. Outside of the ICP, where prices must be taken from the surveys, it would be better to use the methods of

Deaton, Friedman and Alatas (2004), and to construct groups by functional status, such as cereals, tubers, dairy products, and so on.

In principle, both the EKS and weighted CPD methods can deal with the case where not all goods appear in every survey. At the first stage, the construction of the pairwise b^{ij} indices, the calculations can use the largest number of matched commodities for countries i and j, with a different match for each pair of countries. I have not followed that procedure here for a number of reasons. One purely practical reason is that the formulas for variances developed in the previous section assume that the same goods are covered in all countries. While the formulas could be reworked in principle, there are also theoretical reasons for being cautious. In particular, there does not appear to be a good theoretical understanding of the properties of indexes constructed with missing goods. Consider, for example, the fact that the Bangladeshi survey has no (price or quantity) information on fuels. Suppose that the prices of fuels differed markedly from world prices in the missing country; indeed we have already seen this in the India and Indonesia comparison. The PPP for the country with the missing data, country i, say, calculated according to (60) would then be an average over j of terms of the form $b^{1j} + b^{ji}$. The first term would contain the fuels in the comparison, while the second would not, so that it is unclear that this hybrid is of interest. Presumably something similar happens in the case of the weighted CPD index.

For the illustrative calculations in this manual, I have identified 34 commodities or commodity groupings for which there is information in all four surveys. These exclude fuels and alcohol (on which there is no information from the Muslim countries), but include food and tobacco. In consequence, they are not directly comparable to the earlier pairwise comparison of

India and Indonesia, where the PPPs were for food, tobacco, alcohol, and fuels. Averaging the budget shares over households, the 34 commodities comprise 50.4 percent of all expenditures in India, 46.5 percent in Indonesia, 51.2 percent in Bangladesh, and 31.4 percent in Pakistan.

Table 6 shows the results of the calculations for the plutocratic PPP indexes for the four countries. The top panel shows the Törnqvist indexes, the second panel the Fisher indexes, and the third panel the weighted CPD indexes. For the Törnqvist and Fisher indexes, I start with the matrix of pairwise indexes, the b^{ij}, followed by the corresponding matrix of levels, $\exp(b^{ij})$, followed by the EKS-version of corresponding PPPs, derived from the b^{ij} from (60). For the weighted CPD, I start with the PPPs. In all cases, I present the standard errors for the logarithms of the indexes, both from the sampling and from the variation in prices induced by the failue of arbitrage.

Starting from the first panel of the table, it is possible to check that the matrix of indexes does not quite satisfy the "circularity" condition (53), that the conversion from one currency to another does not depend on whether it is done directly or indirectly. For example, according to the Törnqvist index, an Indian rupee is worth 252 Indonesian rupiah. However, if the rupee is converted into Bangladeshi taka first, it is worth 1.39 taka, which will yield 260 rupiah when converted at the Bangladeshi to Indonesian PPP. If the transaction goes from Indian rupees to Indonesian rupiah through Pakistan rupees, the Indian rupee is worth only 235 rupiah, but is worth 243 rupiah if the transaction goes through both Bangladesh and Pakistan. It is this sort of ambiguity that is eliminated by applying (60) to the pairwise PPPs, which yields the next row of the panel, in which an Indian rupee is worth 249 rupiah, 1.37 taka, or 1.23 Pakistani rupees. These EKS-Törnqvist PPPs are very close to the EKS-Fisher PPPs in the next panel, which

estimate that an Indian rupee is worth 253 rupiah, 1.40 taka, or 1.19 Pakistani rupees. For both

EKS indexes, the sampling of the budget shares from the household surveys generates very small

sampling standard errors for the indexes. These are always substantially less than a tenth of one

per cent for the Indonesian to Indian PPP, and less than 0.15 percent for the Indian to

Bangladesh and Pakistan surveys, where the sample sizes are substantially smaller. These

sampling standard errors are about half the size of those in the pairwise comparison in the earlier

sections, presumably because, with four countries, the effective sample size is larger.

The CPD estimates are given in the last panel of the table. Although these are similar to the

EKS-Törnqvist and EKS-Fisher estimates, and have very similar standard errors, they are further

from the two other sets of indexes than the latter are from one another. According to the

weighted CPD PPPs, the Indian rupee is worth about 12 percent less against the rupiah than is

estimated by the EKS indexes. In addition, according to the CPD, the Pakistan to Bangladesh

PPP is essentially unity, while both EKS indexes estimate that the taka is worth about ten percent

more than a Pakistani rupee. It is not clear which of the two sets of indexes to prefer. As we have

seen, the fact that the Fisher and Törnqvist indexes share the same approximation around the

arbitraged vector of prices provides a theoretical rationale for them to be close to one another.

One argument in favor of either the Fisher or the Törnqvist index is that both are superlative in

pairwise comparisons, and while this result has no counterpart for the EKS-versions that are

needed for multilateral comparisons, the EKS adjustment has very little effect in this particular

example because the untreated indexes are close to satisfying the circularity condition.

For the poverty-line PPPs, I once again use an Indian poverty line of 345.08 rupees per

person per month as the international poverty line to be converted into the other three currencies.

49

The results for this poverty line, and for a bandwidth of one standard deviation of the logarithm of per capita expenditure, are presented in Table 7. Somewhat disappointingly, these estimates are very close to the plutocratic estimates in Table 6. This is perhaps not too surprising given that the PPPs are calculated only for food and tobacco, so that the effect of Engel's Law on the weights is lost or much attenuated. Perhaps there will be a larger difference between plutocratic and poverty-line estimates in the full ICP calculation, when it will be possible to use the complete budget. The only substantive difference between Tables 6 and 7 is that the sampling standard errors are somewhat larger, as is to be expected given the restriction to households around the poverty line. When the bandwidth is made smaller, these standard errors become larger, as was the case in the two country comparison in Table 4, but even with a bandwidth of 0.05 of a standard deviation of log pce, the effect is modest; for example, the sampling standard error of 0.0025 for the weighted-CPD index for Pakistan (see bottom right of Table 7) increases to 0.0076.

List of works cited:

Deaton, Angus, Jed Friedman, and Vivi Alatas (2004), "Purchasing power parity exchange rates from household survey data: India and Indonesia," Princeton University, Research Program in Development Studies, processed.

Deaton, Angus, and John Muellbauer (1980), *Economics and Consumer Behavior*, New York. Cambridge University Press.

Diewert, W. Erwin (2001), "The Consumer Price Index and Index Number Purpose", *Journal of Economic and Social Measurement* 27, 167-248

Diewert, W. Erwin (1997), "Commentary on Mathew D. Shapiro and David W. Wilcox, Alternative Strategies for Aggregating Price in the CPI", *The Federal Reserve Bank of St. Louis Review* 79:3 (May/June1997), 127-137.

Diewert, W. Erwin (1976), "Exact and Superlative Index Numbers," *Journal of Econometrics* 4:2 (June 1976), 115-145.

Diewert, W. Erwin (1999), "Axiomatic and Economic Approaches to International Comparisons", pp. 13-87 in *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth, Volume 61, Chicago: The University of Chicago Press

Diewert, W. Erwin (2004), "On the Stochastic Approach to Linking the Regions in the ICP", Department of Economics, Discussion Paper 04-16, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1.

Diewert, W. Erwin (2005), "Weighted Country Product Dummy Variable Regressions and Index Number Formulae", *Review of Income and Wealth*, Series 51:4, 561-569.

Eltetö, O., and P. Köves (1964), "On a problem of index number computation relating to international comparison," *Statisztikai Szemle*, 42, 507-18.

Geary, Roy G. (1958), "A Note on Comparison of Exchange Rates and Purchasing Power between Countries", *Journal of the Royal Statistical Society*, Series A, 121, 97-99.

Gini, Corrado (1924), "Quelques considérations au sujet de la construction des nombres indices des prices et des questions analogues," *Metron*, 3-162.

Hayashi, Fumio (2000), *Econometrics*, Princeton and Oxford. Princeton University Press.

Khamis, S. H., (1972), "A new system of index numbers for national and international purposes," *Journal of the Royal Statistical Society,* Series A, 135, 96-121.

Neary, J. Peter, (2005), "Rationalizing the Penn World Table: true multilateral indexes for international comparisons of real income," *American Economic Review*, 94, 1411-28.

Nuxoll, Daniel A., (1994), "Differences in relative prices and international differences in growth rates,' *American Economic Review*, 84, 1423-36.

Prais, Sigbert J. (1959), "Whose cost of living?"*Review of Economic Studies*, 26, 126-34.

Rao, D.S. Prasada (1990), "A System of Log-Change Index Numbers for Multilateral Comparisons", pp. 127-139 in *Comparisons of Prices and Real Products in Latin America*, J. Salazar-Carillo and D.S. Prasada Rao (eds.), New York: Elsevier Science Publishers.

Rao, D.S. Prasada (1995), "On the Equivalence of the Generalized Country-Product-Dummy (CPD) Method and the Rao System for Multilateral Comparisons", Working Paper No. 5, Centre for International Comparisons, University of Pennsylvania, Philadelphia.

Rao, D.S. Prasada (2001), "Weighted EKS and Generalized CPD Methods for Aggregation at the Basic Heading Level and Above Basic Heading Level", Paper presented at the Joint World Bank and OECD Seminar on Purchasing Power Parities, January 30-February 2, Washington D.C.

Rao, D.S. Prasada (2002), "On the Equivalence of Weighted Country Product Dummy (CPD) Method and the Rao System for Multilateral Price Comparisons", School of Economics, University of New England, Armidale, Australia, March.

Rao, D.S. Prasada (2004), "The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power parities in the ICP", paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, June 30-July 3, 2004, Vancouver, Canada.

Selvanathan, E. A. and D. S. Prasada Rao (1994), *Index Numbers: A Stochastic Approach*, Ann Arbor: The University of Michigan Press.

Szulcz, B. (1964), "Indices for multiregional comparisons," *Przeglad Statystyczny*, 3, 239-54.

Appendix: Derivation of Equation (31)

The weighted CPD estimator is defined by equation (9) which can be rewritten

$$(X'SX)\hat{b} = X'Sy \tag{A.1}$$

The matrix S is a diagonal matrix containing the budget shares for the N goods in the M countries. It is the effect of the sampling variability of S on $\hat{b}$ that is of interest. Applying the delta-method, we can write

$$(X'dSX)b + (X'SX)db = X'dSy \tag{A.2}$$

where b is the value of $\hat{b}$ at the true value of S. From (A.2), we have

$$(X'SX)db = X'dS(y - Xb) = X'dSe \tag{A.3}$$

so that, transposing, post-multiplying, and taking expectations, we have

$$(X'SX)V_b(X'SX) = X'E[dSee'dS]X \tag{A.4}$$

for expectation operator $E[.]$. Because dS is a diagonal matrix,

$$dSee'dS = EdSdS'E \qquad (A.5)$$

where E is a diagonal matrix with the elements of e on the diagonal. Combining (A.4) and (A.5), and using (30) gives equation (31) in the text.

Table 1: Selected aggregate budget shares, India 1999–2000 and Indonesia 1999

India				Indonesia			
Indian NSS description	Rural	Urban	All India	Indonesian Susenas description	Rural	Urban	All Indonesia
Rice	13.09 (0.12)	6.92 (0.11)	10.80 (0.09)	Rice	21.26 0.15	11.63 (0.16)	16.25 (0.13)
Wheat/atta	5.75 (0.05)	3.93 (0.06)	5.08 (0.04)	Wheat flour	0.16 0.01	0.15 (0.01)	0.15 (0.00)
Moong beans	0.53 (0.01)	0.43 (0.01)	0.49 (0.01)	Mungbean	0.10 0.01	0.10 (0.00)	0.10 (0.00)
Liquid milk	7.90 (0.08)	7.33 (0.08)	7.69 (0.06)	Pure milk, packaged milk	0.00 0.00	0.00 (0.00)	0.00 (0.00)
Coconut oil	0.13 (0.00)	0.10 (0.00)	0.12 (0.00)	Coconut oil	1.26 0.03	0.71 (0.02)	0.98 (0.02)
Other edible oils	0.60 (0.01)	0.77 (0.02)	0.67 (0.01)	Other cooking oils	1.31 0.02	1.09 (0.02)	1.20 (0.02)
Eggs	0.37 (0.01)	0.39 (0.01)	0.38 (0.02)	Chicken eggs	1.59 0.02	1.67 (0.02)	1.63 (0.02)
Fish & prawns	1.28 (0.02)	0.94 (0.03)	1.16 (0.00)	Fresh fish & shrimp	4.50 0.06	3.74 (0.05)	4.10 (0.04)
Chicken	0.42 (0.01)	0.47 (0.01)	0.43 (0.01)	Chicken	1.03 0.03	1.45 (0.03)	1.25 (0.02)
Potatoes	1.38 (0.02)	0.74 (0.01)	1.14 (0.01)	Potatoes	0.15 0.01	0.24 (0.01)	0.20 (0.00)
Cabbage	0.21 (0.00)	0.23 (0.00)	0.22 (0.00)	Cabbage	0.20 0.01	0.11 (0.00)	0.15 (0.00)
Palak & other greens	0.40 (0.01)	0.31 (0.01)	0.37 (0.00)	Spinach, water spinach	0.72 0.01	0.54 (0.01)	0.63 (0.01)
Green chilies	0.35 (0.04)	0.23 (0.00)	0.31 (0.00)	Green chilies	0.13 0.01	0.08 (0.01)	0.10 (0.00)
Bananas	0.39 (0.01)	0.51 (0.01)	0.44 (0.00)	Bananas	0.90 0.01	0.54 (0.02)	0.71 (0.00)
Coconuts	0.39 (0.01)	0.35 (0.01)	0.38 (0.01)	Coconuts	1.15 0.02	0.53 (0.00)	0.83 (0.01)
Sugar	2.06 (0.01)	1.55 (0.02)	1.87 (0.01)	Granulated sugar	2.20 0.02	1.40 (0.02)	1.78 (0.02)
Salt	0.22 (0.00)	0.16 (0.00)	0.20 (0.00)	Salt	0.29 0.00	0.15 (0.00)	0.21 (0.00)
Tea (leaf)	1.25 (0.01)	1.14 (0.01)	1.21 (0.01)	Tea (leaf)	0.41 0.01	0.29 (0.00)	0.35 (0.00)
Coffee (powder)	0.07 (0.00)	0.15 (0.01)	0.10 (0.00)	Coffee	0.91 0.01	0.53 (0.01)	0.71 (0.01)
Cigarettes	0.18 (0.01)	0.43 (0.02)	0.27 (0.01)	Cigarettes (clove & tobacco)	5.75 0.06	4.40 (0.08)	5.05 (0.05)
Electricity	1.15 (0.02)	3.46 (0.53)	2.01 (0.20)	Electricity	1.24 0.01	2.12 (0.06)	1.70 (0.03)
Kerosene	0.91 (0.01)	0.94 (0.02)	0.92 (0.01)	Kerosene	1.15 0.02	1.19 (0.02)	1.17 (0.01)

Notes: Budget shares are computed as aggregate total expenditure on the good divided by the aggregate total expenditure on all goods and services, both computed from the household surveys. Shown as percentages. The asymptotic standard errors of the ratios are calculated taking into account the survey designs. The Indian sample has 71,382 rural and 48,919 urban households. The Indonesian sample has 35,970 rural and 25,513 urban households.

Table 2: Average shares, prices, and commodity-specific PPPs

Commodity (unit)	Share	P(1)	P(2)	PPP	Commodity	Share	P(1)	P(2)	PPP
Rice	30.32	2600	10	260	Beef	1.00	22000	35	629
Milk (lt)	8.76	2000	12	167	Powdered milk	0.91	10579	114	93
Wheat	6.60	3400	8	425	Coffee powder (gm)	0.91	15	0.14	107
Cigarettes (no)	5.92	156	1	156	Tomatoes	0.89	3000	8	375
Fish	5.87	7000	40	175	Garlic (gm)	0.72	8	0.04	200
Electricity (kwh)	4.17	156	1.43	109	Biscuits	0.71	10000	40	250
Sugar	4.10	3600	14.7	245	Eggplant	0.68	1400	7	200
Kerosene (lt)	2.34	450	4.53	99	Moong bean	0.67	5000	28	179
Eggs (no)	2.24	480	1.5	320	French beans	0.59	2000	10	200
Other edible oils (lt)	2.09	4000	35	114	Salt	0.46	1667	4	417
Chicken	1.88	12000	60	200	Green chillies	0.46	10000	16.7	600
Tea (gm)	1.76	12	0.14	88	Oranges (no)	0.44	500	2	250
Tea & coffee (cups)	1.64	500	1.5	333	Cabbage	0.42	2000	8	250
Potatoes	1.51	3804	5	761	Pumpkin/gourd	0.37	1200	5.43	221
Coconut (no)	1.35	2000	6	333	Tobacco (leaf)	0.35	20000	60	333
LPG	1.29	1467	11.97	123	Ground nuts	0.32	8000	28	286
Bananas (no)	1.29	231	0.84	277	Mangoes	0.32	3000	15	200
Coconut oil (lt)	1.22	4188	60	70	Black pepper (gm)	0.31	35	0.24	147
Goat meat	1.14	16000	100	160	Tamarind (gm)	0.29	5	0.03	156
Palak & greens	1.11	1300	6	217	Apples	0.28	8000	30	267

Notes: Units are kilograms, unless otherwise noted. Kwh are kilowatt hours, lt liters, and gm grams. The shares in the first column are percentages and are obtained by first normalizing within each country to add to 100 over the 63 commodities in the comparison, and then averaged over the two countries. Goods appear in order of decreasing importance by this share; only the first 40 of the 63 goods are shown. P(1) is the price (median unit value) in Indonesia in rupiah and p(2) is the price in rupees in India. The ratio of the rupiah to the rupee price appears in the column headed PPP, and is the commodity specific exchange rate.

Table 3: Alternative price indexes and standard errors from sampling and stochastics

	Estimate	Standard Error Sampling	Standard Error Stochastic	Standard Error Total
Laspeyres	255.26	0.76	31.1	31.1
Paasche	186.72	0.21	17.4	17.4
Fisher	218.32	0.35	20.1	20.1
Törnqvist	215.67	0.41	19.8	19.8
CPD	219.35	0.48	22.2	22.2

Notes: Units are rupiah per rupee. The sampling standard errors arise from sampling errors in estimating the budget shares that are used as weights. The "stochastic" standard errors come from assuming that log prices satisfy the CPD formulation (8), and then calculating the approximate variance in the index associated with variation in the stochastic component of (8). These standard errors are associated with the uncertainty that comes from having to choose a single index in a world in which the law of one price does not hold across countries.

Table 4a Poverty weighted price indexes and sampling errors for alternative bandwidths

Bandwidth	Fisher	Standard error	Törnqvist	Standard error.	CPD	Standard error
1.0	232.20	0.232	229.65	0.219	227.60	0.233
0.5	233.44	0.266	231.00	0.248	228.46	0.265
0.1	233.43	0.432	231.10	0.410	228.65	0.446
0.05	233.09	0.597	230.91	0.576	228.71	0.595
0.01	231.43	1.364	229.06	1.388	228.48	1.381

Notes: Bandwidths for each country are the number shown multiplied by the standard deviation of the logarithm of per capita household expenditure in that country. Standard errors are for sampling only.

Table 4b: Number of sample households included at each bandwidth

Bandwidth	India	Indonesia Fisher	Indonesia Törnqvist	Indonesia CPD
1.0	63,206	35,977	35,597	35,229
0.5	34,691	20.,430	20,148	19,827
0.1	7,156	4,228	4,143	4,110
0.05	3,669	2,093	2,110	2,088
0.01	739	401	460	442

Table 5: Aggregate (plutocratic) and poverty-line budget shares, with standard errors
(percentages)

| | INDIA | | | | INDONESIA | | | |
| | Plutocratic | | Poverty Line | | Plutocratic | | Poverty line | |
	share	s.e.	share	s.e.	share	s.e.	share	s.e.
Rice	24.5	0.20	35.6	1.30	36.1	0.17	48.3	0.90
Milk	17.5	0.01	10.8	0.86	0.06	0.00	0.19	0.05
Wheat	12.9	0.11	14.1	0.06	0.33	0.01	0.03	0.09
Cigarettes	0.62	0.01	0.05	0.12	11.2	0.01	8.79	0.06
Fish	2.62	0.01	2.19	0.01	9.12	0.01	5.42	0.01
Electricity	4.56	0.00	1.76	0.55	3.78	0.00	2.63	0.04
Sugar	4.24	0.14	4.32	0.05	3.97	0.00	4.63	0.00
Kerosene	2.09	0.01	2.66	0.86	2.59	0.00	3.15	0.03
Eggs	0.86	0.01	0.67	0.25	3.62	0.04	2.39	0.22
Edible oils	1.51	0.01	1.72	0.08	2.66	0.01	3.07	0.00
Chicken	0.99	0.01	0.84	0.06	2.78	0.04	0.37	0.07
Loose tea	2.75	0.03	2.41	0.03	0.77	0.04	0.76	0.24
Tea & coffee	2.34	0.01	1.79	0.10	0.95	0.03	0.64	0.25
Potatoes	2.59	0.04	3.80	0.00	0.44	0.08	0.14	0.00
Coconut	0.85	0.03	0.50	0.21	1.84	0.01	2.10	0.47
LPG	1.99	0.02	0.23	0.21	0.60	0.04	0.00	0.23
Bananas	0.99	0.01	0.58	0.01	1.59	0.02	1.39	0.05
Coconut oil	0.26	0.02	0.07	0.27	2.17	0.05	2.39	0.02
Goat meat	2.19	0.03	1.72	0.01	0.09	0.01	0.02	0.07
Palak greens	0.83	0.01	1.12	0.08	1.39	0.00	1.66	0.20

Notes: These are the 20 leading commodities by order of average plutocratic budget share, as in Table 2. The average of column1 and column 5 is the same as the first column of Table 2. The shares are shares of total expenditure on the 63 covered goods. The poverty-line budget shares are calculated using the biweight kernel and a bandwidth of one percent of the s.d. of log total household expenditure per capita in both countries. The Indonesian and Indian poverty lines are matched using the Fisher index corresponding to the one percent bandwidth.

Table 6: Plutocratic PPPs for India, Indonesia, Bangladesh, and Pakistan

	India	Indonesia	Bangladesh	Pakistan
Törnqvist				
Logarithms of Törnqvist indexes				
India	0	5.530	0.330	0.178
Indonesia	−	0	−5.233	−5.277
Bangladesh	−	−	0	−0.114
Pakistan	−	−	−	0
Törnqvist indexes				
India	1	252	1.39	1.20
Indonesia	−	1	0.00534	0.00571
Bangladesh	−	−	1	0.892
Pakistan	−	−	−	1
Logarithms, standard errors, and levels of EKS-Törnqvist indexes				
Log PPPs	0	5.519	0.312	0.206
Sampling s.e.	0	(0.00066)	(0.00086)	(0.00106)
No arbitrage s.e.	0	(.0766)	(0.1078)	(0.1230)
Level PPPs	1	249	1.37	1.23
Fisher				
Logarithms of Fisher indexes				
India	0	5.519	0.357	0.143
Indonesia	−	0	−5.236	−5.306
Bangladesh	−	−	0	−0.177
Pakistan	−	−	−	0
Fisher indexes				
India	1	256	1.43	1.15
Indonesia	−	1	0.00532	0.00496
Bangladesh	−	−	1	0.8379
Pakistan	−	−	−	1
Logarithms, standard errors, and levels of EKS-Fisher indexes				
Log PPPs	0	5.533	0.336	0.177
Sampling s.e.	0	(0.00066)	(0.00078)	(0.00096)
No arbitrage s.e.	0	(0.0766)	(0.1078)	(0.1230)
Level PPPs	1	253	1.40	1.19
Weighted CPD				
Logarithms, standard error, and levels of weighted CPD index				
Log PPPs	0	5.406	0.194	0.191
Sampling s.e.	0	0.00078	0.00123	0.00137
No arbitrage s.e.	0	0.0802	0.1208	0.1340
Level PPPs	1	223	1.21	1.21

Table 7: Poverty line PPPs for India, Indonesia, Bangladesh, and Pakistan

	India	Indonesia	Bangladesh	Pakistan
Törnqvist				
Logarithms of Törnqvist indexes				
India	0	5.524	0.320	0.187
Indonesia	−	0	−5.250	−5.292
Bangladesh	−	−	0	−0.136
Pakistan	−	−	−	0
Törnqvist indexes				
India	1	250	1.38	1.21
Indonesia	−	1	0.00525	0.00503
Bangladesh	−	−	1	0.873
Pakistan	−	−	−	1
Logarithms, standard errors, and levels of EKS-Törnqvist indexes				
Starting values	0	5.537	0.270	0.285
Log PPPs	0	5.526	0.311	0.193
Sampling s.e.	0	(0.00147)	(0.00161)	(0.00252)
No arbitrage s.e.	0	(.0721)	(0.1304)	(0.1456)
Level PPPs	1	254	1.310	1.330
Fisher				
Logarithms of Fisher indexes				
India	0	5.537	0.351	0.153
Indonesia	−	0	−5.248	−5.316
Bangladesh	−	−	0	−0.203
Pakistan	−	−	−	0
Fisher indexes				
India	1	254	1.42	1.17
Indonesia	−	1	0.00526	0.00491
Bangladesh	−	−	1	0.816
Pakistan	−	−	−	1
Logarithms, standard errors, and levels of EKS-Fisher indexes				
Log PPPs	0	5.536	0.337	0.169
Sampling s.e.	0	(0.00081)	(0.00083)	(0.00100)
No arbitrage s.e.	0	(0.0725)	(0.1275)	(0.1415)
Level PPPs	1	254	1.40	1.18
Weighted CPD				
Logarithms, standard error, and levels of weighted CPD index				
Log PPPs	0	5.422	0.175	0.218
Sampling s.e.	0	0.00095	0.00107	0.00250
No arbitrage s.e.	0	0.0817	0.1516	0.1560
Level PPPs	1	226	1.19	1.24

STATA code for multilateral plutocratic and poverty-line PPPs

1. Introduction and layout

The code provided here required data sets that are defined as follows. For each country, the household expenditures are stored in a data set name expsurv_*iso*.dta, where *iso* is the three letter isocode for the country. This is a household level data set, with one record per household, and which contains the following variables:

hhid	the household i.d.
x1, x2, x3 etc	expenditure by household on goods 1, 2, 3 , up to *N*
hhs	household size
hh_totx	household total expenditure on *all* goods
sx	the sum of expenditures on the *N* goods
psu	identifier for the primary sampling unit
mult	a household level multiplier or inflation factor
stratum	the sampling stratum to which the household belongs

The second data set contains the price information. In the calculations here, I use unit values for prices, so that this data set is called uvs_match.dta It contains

kgp	the number of the good or group of goods
uv_1, uv_2,	unit values for country _1, _2, etc up to *M*.

expsurv_iso.dta has as many records as there are households in each survey, and there are *M* such datasets. There is only one dataset uvs_match.dta, which has *N* elements.

Finally, there is a data set vcv_w_*iso*.dta for each country, labeled by *iso*, which contains the plutocratic budget shares of the *N* goods and their sampling variance variance covariance matrix (*N* x *N*). For the present, these were derived separately, though they could easily have been calculated from expsurv_*iso*.dta.

The programs (do files) are as follows, in the order in which they need to be run. I have tried as much as possible to write the code where the number of countries and number of goods is specified at the beginning, but there are a few places, which are marked, where this was not possible, and where the code will have to be adapted (always in an obvious way) to change the number of goods or number of countries.

cpd_index calculates the plutocratic weighted CPD index and its standard errors

super_index calculates the plutocratic EKS-Fisher, EKS-Törnqvist, and standard errors

torn_approx_all estimates Engel curves for each country, and calculates the poverty-line EKS-Törnqvist index for this case.

poverty_indexes_cpd calculates the weighted CPD indexes at the poverty line, starting from the Törnqvists calculated above (entered by hand here). It does so iteratively, but for a range of bandwidths expressed as fractions of the standard deviation of the logarithm of per capita expenditure. The converged estimates, together with the bandwidths, are saved in a data set pppp_cpd.dta

poverty_indexes_fish and **poverty_indexes_torn** do the same as **poverty_indexes_cpd**, but for the EKS-Fisher and EKS-Törnqvist indexes respectively. The results, and bandwidths, are saved in pppp_fish.dta and pppp_torn.dta

poverty_indexes_cpd_sterrs, **poverty_indexes_fish_sterrs**, and **poverty_indexes_torn_sterrs** pick up the saved data from the iterations and calculate the various standard errors for each of the indexes.

1. **cpd_index.do**

```
version 9.2
#delimit ;
clear;
set logtype text;
cap log close;
log using cpd_index, replace;
set more off;
set matsize 5000;


use vcv_w_ind;
rename w_est w_ind;
mkmat vcv_w*, mat(vcv1);
gen kgp=_n;
sort kgp;
save temp, replace;

use vcv_w_idn;
rename w_est w_idn;
mkmat vcv_w*, mat(vcv2);
gen kgp=_n;
sort kgp;
merge kgp using temp, unique;
tab _m; drop _m;
sort kgp;
save temp, replace;

use vcv_w_bgd;
rename w_est w_bgd;
mkmat vcv_w*, mat(vcv3);
gen kgp=_n;
sort kgp;
merge kgp using temp, unique;
tab _m; drop _m;
sort kgp;
save temp, replace;
```

```
use vcv_w_pak;
rename w_est w_pak;
mkmat vcv_w*, mat(vcv4);
gen kgp=_n;
sort kgp;
merge kgp using temp, unique;
tab _m; drop _m;
sort kgp;
save temp, replace;

merge kgp using uvs_match, unique;
tab _m; drop _m;
sort kgp;

label data "uvs and pluto shares";
save uvs_w_match, replace;

ren uv_ind uv1;
ren uv_idn uv2;
ren uv_bgd uv3;
ren uv_pak uv4;

ren w_ind w1;
ren w_idn w2;
ren w_bgd w3;
ren w_pak w4;

reshape long uv w, i(kgp) j(cno);
des;

gen lnp=log(uv);

xi: regress lnp i.cno i.kgp;

xi: regress lnp i.cno i.kgp [aweight=w];


**now doing the standard errors etc;

scalar ng=34;
scalar nc=4;
scalar nt=nc*ng;

mkmat lnp;
mkmat w;
matrix w=diag(w);
matrix x=J(nt,1,1);
forvalues i=2(1)4 {;
        mkmat _Icno_`i';
        matrix x=x,_Icno_`i';
        matrix drop _Icno_`i';
};

forvalues i=2(1)34 {;
        mkmat _Ikgp_`i';
        matrix x=x,_Ikgp_`i';
```

```
        matrix drop _Ikgp_`i';
};


matrix aa=syminv(x'*w*x);
matrix b=aa*x'*w*lnp;
matrix list b;




**sampling standard errors;

matrix bu=syminv(x'*x)*x'*lnp;
matrix g=lnp-x*bu;
matrix g=diag(g);



matrix zz=J(ng,ng,0);
matrix
vv=(vcv1,zz,zz,zz)\(zz,vcv2,zz,zz)\(zz,zz,vcv3,zz)\(zz,zz,zz,vcv4);
matrix vs=aa*x'*g*vv*g*x*aa;
display("Sampling standard errors");
forvalues i=1(1)4 {;
display sqrt(vs[`i',`i']);
};



**failure of arbitrage standard errors;
matrix gg=g*g;
matrix vf=aa*x'*w*gg*w*x*aa;
display("Failure of arbitrage standard errors");
forvalues i=1(1)4 {;
display sqrt(vf[`i',`i']);
};

matrix va=vf+vs;

display("Total standard errors");
forvalues i=1(1)4 {;
display sqrt(va[`i',`i']);
};

log close;
```

2. super_index.do

```
**this does the eks indexes: tornqvist and fisher

version 9.2
#delimit ;
clear;
set logtype text;
cap log close;
log using super_index, replace;
set more off;
```

```
set matsize 5000;

global cn=4;
**number of countries;
global gn=34;
**number of goods;

global iso="ind idn bgd pak";
global i=1;
foreach cn of global iso {;
        drop _all;
        use vcv_w_`cn';
        mkmat vcv_w*, mat(vcv$i);
        drop _all;
        use uvs_w_match;
        ren uv_`cn' p$i;
        ren w_`cn' w$i;
        gen lnp$i=log(p$i);
        mkmat lnp$i;
        mkmat p$i;
        mkmat p$i;
        mkmat w$i;
        global i=$i+1;
};


drop _all;


matrix fish=J($cn,$cn,0);
matrix torn=J($cn,$cn,0);


forvalues i=1(1)$cn {;
        forvalues j=1(1)$cn {;
                matrix fish[`i',`j']=((w`i')'*syminv(diag(p`i'))*p`j');
                matrix fish[`i',`j']=log(fish[`i',`j']);
                matrix torn[`i',`j']=0.5*(w`i'+w`j')'*(lnp`j'-lnp`i');
        };
};

matrix fish=0.5*(fish-fish');

display("Log of Tornqvist indexes");
matrix list torn;
display("Log of Fisher indexes");
matrix list fish;

display("Now making India the base and averaging");

matrix eks_fish=J($cn,1,0);
matrix eks_torn=J($cn,1,0);


forvalues i=1(1)$cn {;
        matrix eks_fish[`i',1]=0;
        matrix eks_torn[`i',1]=0;
```

5

```
        forvalues j=1(1)$cn {;
                matrix
eks_torn[`i',1]=eks_torn[`i',1]+torn[1,`j']+torn[`j',`i'];
                matrix
eks_fish[`i',1]=eks_fish[`i',1]+fish[1,`j']+fish[`j',`i'];
        };
};

matrix eks_fish=eks_fish/$cn;
matrix eks_torn=eks_torn/$cn;

display("Fisher EKS indexes");
matrix list eks_fish;

display("Tornqvist EKS indexes");
matrix list eks_torn;


**sampling variance of the Tornqvist EKS

matrix vs_t=J($cn,1,0);

**first term;
scalar a1=0;
forvalues j=1(1)$cn {;
        matrix t1j=lnp`j'-lnp1;
        forvalues k=1(1)$cn {;
                matrix t1k=lnp`k'-lnp1;
                matrix aa=t1j'*vcv1*t1k;
                scalar a1=a1+aa[1,1];
        };
};

**second and subsequent terms;

matrix a2=J($cn,1,0);
matrix a3=J($cn,1,0);
matrix a4=J($cn,1,0);
matrix a5=J($cn,1,0);

forvalues i=1(1)$cn {;
        local a2=0;
        local a3=0;
        local a4=0;
        local a5=0;
        matrix t1i=lnp`i'-lnp1;
        matrix vit1i=vcv`i'*t1i;
        forvalues j=1(1)$cn {;
                matrix aa=t1i'*vcv1*(lnp`j'-lnp1);
                local a2=`a2'+aa[1,1];
                matrix aa=(lnp`i'-lnp`j')'*vit1i;
                local a3=`a3'+aa[1,1];
                matrix tjivi=(lnp`i'-lnp`j')'*vcv`i';
                matrix aa=t1i'*vcv`j'*t1i;
                local a5=`a5'+aa[1,1];
                forvalues k=1(1)$cn {;
                        matrix aa=tjivi*(lnp`i'-lnp`k');
```

```
                              local a4=`a4'+aa[1,1];
                         };
               };
          matrix a2[`i',1]=2*`a2';
          matrix a3[`i',1]=2*`a3';
          matrix a4[`i',1]=`a4';
          matrix a5[`i',1]=`a5';
};

matrix fd=-1\J($cn-1,1,1);

matrix vs_t=(a1*fd+a2+a3+a4+a5)/(4*$cn*$cn);
matrix list vs_t;

**abritrage failure variance of the tornqvist;

*first running the unweighted cpd regressions in order to get the
variance matrices;

**note that the lines below will have to be modified depending on the
number of countries;

matrix lnp=lnp1\lnp2\lnp3\lnp4;
matrix i1=J($gn,1,1);
matrix i0=J($gn,1,0);
matrix x=(i1\i0\i0\i0),(i0\i1\i0\i0),(i0\i0\i1\i0),(i0\i0\i0\i1);
matrix aa=J(1,$gn-1,0)\I($gn-1);
matrix x=x,(aa\aa\aa\aa);
matrix eu=lnp-x*syminv(x'*x)*x'*lnp;
matrix e2=diag(eu)*eu;

**these are the diagomal matrix of squared residuals for each country;

matrix om1=diag(e2[1..$gn,1]);
matrix om2=diag(e2[$gn+1..2*$gn,1]);
matrix om3=diag(e2[2*$gn+1..3*$gn,1]);
matrix om4=diag(e2[3*$gn+1..4*$gn,1]);


matrix a1=J($cn,1,0);
matrix a2=J($cn,1,0);
matrix a3=J($cn,1,0);
matrix a4=J($cn,1,0);
matrix a5=J($cn,1,0);

forvalues i=1(1)$cn {;
          local a1=0;
          local a2=0;
          local a3=0;
          local a4=0;
          local a5=0;
          forvalues j=1(1)$cn {;
          matrix aa=(w`i'+w`j')'*om`i'*(w1-w`i');
          local a2=`a2'+2*aa[1,1];
          matrix aa=(w1-w`i')'*om1*(w1+w`j');
          local a3=`a3'-2*aa[1,1];
          matrix aa=(w1-w`i')'*om`j'*(w1-w`i');
```

```
        local a4=`a4'+aa[1,1];
                forvalues k=1(1)$cn {;
                matrix aa=(w`i'+w`j')'*om`i'*(w`i'+w`k');
                local a1=`a1'+aa[1,1];
                matrix aa=(w1+w`j')'*om1*(w1+w`k');
                local a5=`a5'+aa[1,1];
        };};
        matrix a1[`i',1]=`a1';
        matrix a2[`i',1]=`a2';
        matrix a3[`i',1]=`a3';
        matrix a4[`i',1]=`a4';
        matrix a5[`i',1]=`a5';
        if `i'==1 {;
                matrix a1[`i',1]=-`a1';
        };
};

matrix va_t=(a1+a2+a3+a4+a5)/(4*$cn^2);

matrix list va_t;


forvalues i=2(1)4 {;
        display("sampling s.e. country `i'");
        display sqrt(vs_t[`i',1]);
        display("arbitrage s.e. country `i'");
        display sqrt(va_t[`i',1]);
        display("total s.e. country `i'");
        display sqrt(vs_t[`i',1]+va_t[`i',1]);
};


**sampling variance for the Fisher EKS;

**making the betaij's;

forvalues i=1(1)$cn {;
        forvalues j=1(1)$cn {;
                matrix r`i'`j'=J($gn,1,0);
                forvalues n=1(1)$gn{;
                        matrix r`i'`j'[`n',1]=p`j'[`n',1]/p`i'[`n',1];
                };
        matrix aa=w`i''*r`i'`j';
        matrix beta`i'`j'=r`i'`j'/aa[1,1];
        matrix drop r`i'`j';
        };
};



matrix vs_f=J($cn,1,0);

**first term;
scalar a1=0;
forvalues j=1(1)$cn {;
        forvalues k=1(1)$cn {;
                matrix aa=beta1`j''*vcv1*beta1`k';
```

```
                    scalar a1=a1+aa[1,1];
        };
};

**second and subsequent terms;

matrix a2=J($cn,1,0);
matrix a3=J($cn,1,0);
matrix a4=J($cn,1,0);
matrix a5=J($cn,1,0);

forvalues i=1(1)$cn {;
        local a2=0;
        local a3=0;
        local a4=0;
        local a5=0;
        forvalues j=1(1)$cn {;
                matrix aa=beta`i'`j''*vcv`i'*(beta`i'`i'-beta`i'1);
                local a2=`a2'-2*aa[1,1];
                matrix aa=beta1`j''*vcv1*(beta1`i'-beta11);
                local a3=`a3'+2*aa[1,1];
                matrix aa=(beta`j'`i'-beta`j'1)'*vcv`j'*(beta`j'`i'-
beta`j'1);
                local a4=`a4'+aa[1,1];
                forvalues k=1(1)$cn {;
                        matrix aa=beta`i'`j''*vcv`i'*beta`i'`k';
                        local a5=`a5'+aa[1,1];
                };
        };
        matrix a2[`i',1]=`a2';
        matrix a3[`i',1]=`a3';
        matrix a4[`i',1]=`a4';
        matrix a5[`i',1]=`a5';
};

matrix fd=-1\J($cn-1,1,1);

matrix vs_f=(a1*fd+a2+a3+a4+a5)/(4*$cn*$cn);
matrix list vs_f;


**note that arbitrage variance for Fisher is same as for Tornqvist;
matrix va_f=va_t;


display ("Standard errors for EKS-Fisher index");

forvalues i=2(1)4 {;
        display("sampling s.e. country `i'");
        display sqrt(vs_f[`i',1]);
        display("arbitrage s.e. country `i'");
        display sqrt(va_f[`i',1]);
        display("total s.e. country `i'");
        display sqrt(vs_f[`i',1]+va_f[`i',1]);
};

log close;
```

3. torn_approx_all.do

```
**makes the tornvist approximation to the PPPP
** written for the m-country case

version 9.2
#delimit ;
clear;
set logtype text;
cap log close;
log using torn_approx, replace;
set more off;

global cns="ind idn bgd pak";
global gn=34;
scalar pl0=345.08;
global cn=4;
scalar lnz=log(pl0);




forvalues i=1(1)$cn {;
        matrix xei`i'=J($gn,2,0);
};

global cno=1;
foreach iso of global cns {;
        forvalues i=1(1)$gn {;
                drop _all;
                use x`i' sx mult hhs hh_totx using expsurv_`iso';
                gen lnpce=log(hh_totx/hhs);
                gen w=x`i'/sx;
                qui regress w lnpce [aweight=mult];
                matrix xei$cno[`i',1]=_b[_cons];
                matrix xei$cno[`i',2]=_b[lnpce];
        };
global cno=$cno+1;
};


drop _all;
use uvs_w_match;
global cno=1;
foreach iso of global cns {;
        ren uv_`iso' p$cno;
        gen lnp$cno=log(p$cno);
        mkmat lnp$cno;
        global cno=$cno+1;
};

drop _all;




matrix zeta=J($cn,$cn,0);
matrix phi=J($cn,$cn,0);
```

```
matrix psi=J($cn,$cn,0);
forvalues i=1(1)$cn {;
        forvalues j=1(1)$cn {;
                matrix
zeta[`i',`j']=0.5*(xei`i'[1..$gn,1]+xei`j'[1..$gn,1]+xei`i'[1..$gn,2]*l
nz+xei`j'[1..$gn,2]*lnz)'*(lnp`j'-lnp`i');
                matrix phi[`i',`j']=0.5*xei`i'[1..$gn,2]'*(lnp`j'-
lnp`i');
                matrix psi[`i',`j']=0.5*xei`j'[1..$gn,2]'*(lnp`j'-
lnp`i');
        };
};

matrix uns=J($cn,1,1);

matrix a=(phi'+uns*psi[1,1..$cn])/$cn+(I($cn)-(1/$cn)*diag(psi'*uns));

matrix ax=J($cn,$cn,0);
matrix ax[1,1]=(phi[1,1..$cn]*uns)/$cn;
matrix a=a+ax;
matrix b=((zeta[1,1..$cn]*uns)*uns+zeta'*uns)/$cn;

matrix ans=inv(a)*b;

display("PPPP Tornqvist approximations");
matrix list ans;

log close;
```

4. poverty_indexes_cpd.do

```
**this makes cpd poverty weighted indexes for a range of bandwidths 1
s.d. 0.5 s.d. 0.1 s.d

version 9.2
#delimit ;

set logtype text;
cap log close;
log using poverty_indexes_cpd, replace;
clear;
set more off;

global cns="ind idn bgd pak";
global cn=4;
global gn=34;
global nits=10;

**bandwidth for both kernels, in units of sd of logs of pce;
scalar h=1;
**Indian poverty line, more generally, base country PL;
scalar pl0=345.08;
**starting value for Tornqvist PPPP;
matrix torns=0\5.5365\0.2704\0.2854;
scalar lnz=log(pl0);
```

```
matrix bws=(1\0.5\0.1\0.05);

**making the logarithm of price vectors;

use uvs_w_match;
global cno=1;
foreach iso of global cns {;
        ren uv_`iso' p$cno;
        gen lnp$cno=log(p$cno);
        mkmat lnp$cno;
        global cno=$cno+1;
};

drop _all;

**constructing the X matrix for the CPD regressions;
**NB this would have to be modified depending on the number of
countries;

matrix lnp=lnp1\lnp2\lnp3\lnp4;
matrix i1=J($gn,1,1);
matrix i0=J($gn,1,0);
matrix x=(i1\i0\i0\i0),(i0\i1\i0\i0),(i0\i0\i1\i0),(i0\i0\i0\i1);
matrix aa=J(1,$gn-1,0)\I($gn-1);
matrix x=x,(aa\aa\aa\aa);


global cno=1;
foreach iso of global cns {;
        use expsurv_`iso';
        drop if hh_totx==.;
        gen pce=hh_totx/hhs;
        keep x* mult pce sx;
        forvalues i=1(1)$gn {;
                gen w`i'=x`i'/sx;
                drop x`i';
        };
        keep w* pce mult;
        save temp$cno, replace;
        global cno=$cno+1;
};


cap program drop mk_wpov;
program define mk_wpov;
        use temp`1';
        gen lnx=log(pce);
        qui summ lnx [aw=mult];
        scalar hbw=`2'*r(sd);
        gen tau=abs(lnx-`3')/hbw;
        gen kwt=(15/(16*hbw))*(1-tau^2)^2;
        replace kwt=0 if tau >= 1;
        qui count if tau < 1;
        display "No of observations in band is " r(N);
        display "Fraction of sample in band is " r(N)/_N;
        scalar nin_s=r(N);
```

```
        gen pwt=mult*kwt;
        drop mult;
        collapse w* [aweight=pwt];
        gen ct=1;
        reshape long w, i(ct) j(kgp);
        drop ct;
        sort kgp;
        mkmat w, matrix(`4');
        drop _all;
end;


matrix ans=J(4,1,0);

matrix nin=J($cn,5,0);

global cns_1="idn bgd pak";

forvalues ibw=1(1)4 {;
scalar hhh=bws[`ibw',1];
display "Bandwith is " hhh " standard deviations of log pce";
mk_wpov 1 hhh lnz "wpov1";
matrix nin[1,`ibw']=nin_s;

matrix torn_ans=torns;

**starting from the trial tornqvist number for the pppp;


matrix torns_1=torns;

forvalues it=1(1)$nits {;
        global cno=2;
        foreach iso of global cns_1 {;
                scalar ptorn=torns_1[$cno,1];
                scalar lnpl$cno=lnz+ptorn;
                mk_wpov "$cno" hhh lnpl$cno "wpov$cno";
                matrix nin[$cno,`ibw']=nin_s;
                global cno=$cno+1;
        };
        matrix w=wpov1;
        forvalues i=2(1)$cn {;
                matrix w=w\wpov`i';
        };
        matrix w=diag(w);
        matrix aa=syminv(x'*w*x);
        matrix b=aa*x'*w*lnp;
        matrix tt=b[2..$cn,1]-J($cn-1,1,1)*b[1,1];
        matrix torns_1=0\tt;
        display "New CPD indexes";
        matrix list torns_1;
        matrix torn_ans=torn_ans,torns_1;
};

display"Bandwidth is " hhh;
display ("Log indexes at each iteration");
matrix list torn_ans;
```

```
if `ibw'==1 {;
        matrix ans=torns_1;
};
if `ibw' > 1 {;
        matrix ans=ans,torns_1;
};
};


display("Numbers in the bands in final iteration");
matrix list nin;
display("Final estimates of CPD insexes at each bandwidth");
matrix list ans;

svmat ans;
gen cno=_n;
reshape long ans,i(cno) j(bwn);
gen bw=1;
replace bw=0.5 if bwn==2;
replace bw=0.1 if bwn==3;
replace bw=0.05 if bwn==4;

label data "CPD PPPPs";
save pppp_cpd, replace;


log close;
```

5. poverty_indexes_fish.do

```
**this makes fisher poverty weighted indexes for a range of bandwidths
1 s.d. 0.5 0.1 0.05 s.d.
**I have dropped the smallest bandwidth, which is too small for BGD and
PAK

version 9.2
#delimit ;

set logtype text;
cap log close;
log using poverty_indexes_fish, replace;
clear;
set more off;

global cns="ind idn bgd pak";
global cn=4;
global gn=34;
global nits=10;

**bandwidth for both kernels, in units of sd of logs of pce;
scalar h=1;
**Indian poverty line, more generally, base country PL;
scalar pl0=345.08;
**starting value for Tornqvist PPPP;
matrix torns=0\5.5365\0.2704\0.2854;
```

```
scalar lnz=log(pl0);

matrix bws=(1\0.5\0.1\0.05);

**making the logarithm of price vectors;

use uvs_w_match;
global cno=1;
foreach iso of global cns {;
        ren uv_`iso' p$cno;
        gen lnp$cno=log(p$cno);
        mkmat lnp$cno;
        mkmat p$cno;
        global cno=$cno+1;
};


drop _all;



global cno=1;
foreach iso of global cns {;
        use expsurv_`iso';
        drop if hh_totx==.;
        gen pce=hh_totx/hhs;
        keep x* mult pce sx;
        forvalues i=1(1)$gn {;
                gen w`i'=x`i'/sx;
                drop x`i';
        };
        keep w* pce mult;
        save temp$cno, replace;
        global cno=$cno+1;
};



cap program drop mk_wpov;
program define mk_wpov;
        use temp`1';
        gen lnx=log(pce);
        qui summ lnx [aw=mult];
        scalar hbw=`2'*r(sd);
        gen tau=abs(lnx-`3')/hbw;
        gen kwt=(15/(16*hbw))*(1-tau^2)^2;
        replace kwt=0 if tau >= 1;
        qui count if tau < 1;
        display "No of observations in band is " r(N);
        display "Fraction of sample in band is " r(N)/_N;
        scalar nin_s=r(N);
        gen pwt=mult*kwt;
        drop mult;
        collapse w* [aweight=pwt];
        gen ct=1;
        reshape long w, i(ct) j(kgp);
        drop ct;
```

```
        sort kgp;
        mkmat w, matrix(`4');
        drop _all;
end;


matrix ans=J(5,3,0);

matrix nin=J($cn,5,0);

global cns_1="idn bgd pak";

forvalues ibw=1(1)4 {;
scalar hhh=bws[`ibw',1];
display "Bandwith is " hhh " standard deviations of log pce";
mk_wpov 1 hhh lnz "wpov1";
matrix nin[1,`ibw']=nin_s;

matrix torn_ans=torns;

**starting from the trial tornqvist number for the pppp;


matrix torns_1=torns;
matrix b=J($cn,$cn,0);
matrix uns=J($cn,1,1);

forvalues it=1(1)$nits {;
        global cno=2;
        foreach iso of global cns_1 {;
                scalar ptorn=torns_1[$cno,1];
                scalar lnpl$cno=lnz+ptorn;
                mk_wpov "$cno" hhh lnpl$cno "wpov$cno";
                matrix nin[$cno,`ibw']=nin_s;
                global cno=$cno+1;
        };
        forvalues i=1(1)$cn {;
                forvalues j=1(1)$cn {;
                        matrix aa=(wpov`i''*invsym(diag(p`i'))*p`j');
                        matrix b[`i',`j']=log(aa[1,1]);
                };
        };

        matrix b=0.5*(b-b');
        matrix cc=b'*uns/$cn;
        matrix aa=b[1,1..$cn]*uns/$cn;
        matrix torns_1=cc+aa[1,1]*uns;
        display "New Fisher indexes";
        matrix list torns_1;
        matrix torn_ans=torn_ans,torns_1;
};

display"Bandwidth is " hhh;
display ("Log indexes at each iteration");
matrix list torn_ans;

if `ibw'==1 {;
```

```
        matrix ans=torns_1;
};
if `ibw' > 1 {;
        matrix ans=ans,torns_1;
};


};


display("Numbers in the bands in final iteration");
matrix list nin;
display("Final estimates of Fisher indexes at each bandwidth");
matrix list ans;

svmat ans;
gen cno=_n;
reshape long ans,i(cno) j(bwn);
gen bw=1;
replace bw=0.5 if bwn==2;
replace bw=0.1 if bwn==3;
replace bw=0.05 if bwn==4;

label data "Fisher EKS PPPPs";
save pppp_fish, replace;

log close;
```

6. poverty_indexes_torn.do

```
**this makes tornqvist poverty weighted indexes for a range of
bandwidths 1 s.d. 0.5 0.1 0.05 s.d.
**I have dropped the smallest bandwidth, which is too small for BGD and
PAK

version 9.2
#delimit ;

set logtype text;
cap log close;
log using poverty_indexes_torn, replace;
clear;
set more off;

global cns="ind idn bgd pak";
global cn=4;
global gn=34;
global nits=10;

**bandwidth for both kernels, in units of sd of logs of pce;
scalar h=1;
**Indian poverty line, more generally, base country PL;
scalar pl0=345.08;
**starting value for Tornqvist PPPP;
matrix torns=0\5.5365\0.2704\0.2854;
scalar lnz=log(pl0);
```

```
matrix bws=(1\0.5\0.1\0.05);

**making the logarithm of price vectors;

use uvs_w_match;
global cno=1;
foreach iso of global cns {;
        ren uv_`iso' p$cno;
        gen lnp$cno=log(p$cno);
        mkmat lnp$cno;
        global cno=$cno+1;
};

drop _all;



global cno=1;
foreach iso of global cns {;
        use expsurv_`iso';
        drop if hh_totx==.;
        gen pce=hh_totx/hhs;
        keep x* mult pce sx;
        forvalues i=1(1)$gn {;
                gen w`i'=x`i'/sx;
                drop x`i';
        };
        keep w* pce mult;
        save temp$cno, replace;
        global cno=$cno+1;
};



cap program drop mk_wpov;
program define mk_wpov;
        use temp`1';
        gen lnx=log(pce);
        qui summ lnx [aw=mult];
        scalar hbw=`2'*r(sd);
        gen tau=abs(lnx-`3')/hbw;
        gen kwt=(15/(16*hbw))*(1-tau^2)^2;
        replace kwt=0 if tau >= 1;
        qui count if tau < 1;
        display "No of observations in band is " r(N);
        display "Fraction of sample in band is " r(N)/_N;
        scalar nin_s=r(N);
        gen pwt=mult*kwt;
        drop mult;
        collapse w* [aweight=pwt];
        gen ct=1;
        reshape long w, i(ct) j(kgp);
        drop ct;
        sort kgp;
        mkmat w, matrix(`4');
        drop _all;
```

```
end;


matrix ans=J(5,3,0);

matrix nin=J($cn,5,0);

global cns_1="idn bgd pak";

forvalues ibw=1(1)4 {;
scalar hhh=bws[`ibw',1];
display "Bandwith is " hhh " standard deviations of log pce";
mk_wpov 1 hhh lnz "wpov1";
matrix nin[1,`ibw']=nin_s;

matrix torn_ans=torns;

**starting from the trial tornqvist number for the pppp;


matrix torns_1=torns;
matrix b=J($cn,$cn,0);
matrix uns=J($cn,1,1);

forvalues it=1(1)$nits {;
        global cno=2;
        foreach iso of global cns_1 {;
                scalar ptorn=torns_1[$cno,1];
                scalar lnpl$cno=lnz+ptorn;
                mk_wpov "$cno" hhh lnpl$cno "wpov$cno";
                matrix nin[$cno,`ibw']=nin_s;
                global cno=$cno+1;
        };
        forvalues i=1(1)$cn {;
                forvalues j=1(1)$cn {;
                        matrix aa=0.5*(wpov`i'+wpov`j')'*(lnp`j'-
lnp`i');
                        matrix b[`i',`j']=aa[1,1];
                };
        };

        matrix cc=b'*uns/$cn;
        matrix aa=b[1,1..$cn]*uns/$cn;
        matrix torns_1=cc+aa[1,1]*uns;
        display "New Tornqvist indexes";
        matrix list torns_1;
        matrix torn_ans=torn_ans,torns_1;
};

display"Bandwidth is " hhh;
display ("Log indexes at each iteration");
matrix list torn_ans;

if `ibw'==1 {;
        matrix ans=torns_1;
};
if `ibw' > 1 {;
```

```
            matrix ans=ans,torns_1;
};


};


display("Numbers in the bands in final iteration");
matrix list nin;

display("Final estimates of Tornqvist indexes at each bandwidth");
matrix list ans;

svmat ans;
gen cno=_n;
reshape long ans,i(cno) j(bwn);
gen bw=1;
replace bw=0.5 if bwn==2;
replace bw=0.1 if bwn==3;
replace bw=0.05 if bwn==4;

label data "Tornqvist PPPPs";
save pppp_torn, replace;

log close;
```

7. poverty_indexes_cpd_sterrs.do

```
**this makes the poverty weighted cpd index with standard errors: uses
last iteration from poverty_indexes_cpd
**plus code from pppp_sterrs to incorporate standard errors, together
with appropriate bits from cpd_index

version 8.2
#delimit ;

set logtype text;
cap log close;
log using poverty_indexes_cpd_sterrs, replace;
clear;
set more off;

global cns="ind idn bgd pak";
global cn=4;
global gn=34;

**Indian poverty line, more generally, base country PL;
scalar pl0=345.08;
scalar lnz=log(pl0);

**Retrieving the calculated estimates of the indexes at each bandwidth;
use pppp_cpd;
reshape wide ans, i(bwn) j(cno);
mkmat ans*, mat(torn);
matrix torn=torn';
```

```
mkmat bw, mat(bws);
drop _all;


scalar lnz=log(pl0);


**making the logarithm of price vectors;


use uvs_w_match;
global cno=1;
foreach iso of global cns {;
        ren uv_`iso' p$cno;
        gen lnp$cno=log(p$cno);
        mkmat lnp$cno;
        global cno=$cno+1;
};


drop _all;


**constructing the X matrix for the CPD regressions;
**NB this would have to be modified depending on the number of
countries;

matrix lnp=lnp1\lnp2\lnp3\lnp4;
matrix i1=J($gn,1,1);
matrix i0=J($gn,1,0);
**this is with the constant term;
matrix x=(i1\i1\i1\i1),(i0\i1\i0\i0),(i0\i0\i1\i0),(i0\i0\i0\i1);
matrix aa=J(1,$gn-1,0)\I($gn-1);
matrix x=x,(aa\aa\aa\aa);


**making the data sets that are used to calculate the poverty weighted
budget shares;

global cno=1;
foreach iso of global cns {;
        use expsurv_`iso';
        drop if hh_totx==.;
        gen pce=hh_totx/hhs;
        keep x* mult pce sx stratum psu;
        forvalues i=1(1)$gn {;
                gen w`i'=x`i'/sx;
                drop x`i';
        };
        keep w* pce mult stratum psu;
        svyset [pweight=mult], strata(stratum) psu(psu);
        save temp$cno, replace;
        global cno=$cno+1;
};


**this program makes the variance covariance matrix of the sampling
estimates;
**NB it will need to be altered to suit the number of goods;


cap program drop mkw_vcv;
program define mkw_vcv;
```

```
        use temp`1';
        gen lnx=log(pce);
        qui summ lnx [aw=mult];
        scalar hbw=`2'*r(sd);
        gen tau=abs(lnx-`3')/hbw;
        gen kwt=(15/(16*hbw))*(1-tau^2)^2;
        replace kwt=0 if tau >= 1;
        qui summ kwt;
        forvalues i=1(1)$gn{;
                replace w`i'=w`i'*kwt;
        };
        svyratio w1/kwt w2/kwt w3/kwt w4/kwt w5/kwt w6/kwt w7/kwt
w8/kwt w9/kwt w10/kwt w11/kwt w12/kwt w13/kwt w14/kwt w15/kwt w16/kwt
         w17/kwt w18/kwt w19/kwt w20/kwt w21/kwt w22/kwt w23/kwt
w24/kwt w25/kwt w26/kwt w27/kwt w28/kwt w29/kwt w30/kwt w31/kwt w32/kwt
         w33/kwt w34/kwt;
        matrix w_est=e(est);
        matrix w_est=w_est';
        matrix vcv_w=e(V_db);
        drop _all;
end;

matrix zz=J($gn,$gn,0);

matrix ses_s=J(4,4,0);
matrix ses_f=J(4,4,0);
matrix ses_a=J(4,4,0);

forvalues ibw=1(1)4 {;
        scalar bw=bws[`ibw',1];
        display "Bandwidth is " bw;
        forvalues cno=1(1)$cn {;
                local lnpppp=torn[`cno',`ibw'];
                local lnpl=lnz+`lnpppp';
                mkw_vcv `cno' bw `lnpl';
                matrix w`cno'=w_est;
                matrix vcv`cno'=vcv_w;
        };
        matrix w=w1;
        forvalues j=2(1)$cn {;
                matrix w=w\w`j';
        };
        matrix w=diag(w);
        matrix aa=syminv(x'*w*x);
        matrix b=aa*x'*w*lnp;
        display("Original indexes");
        matrix tt=torn[1..$cn,`ibw'];
        matrix list tt;
        display("Confirming the PPPP indexes");
        matrix bb=b[2..4,1];
        matrix list bb;
        **sampling standard errors;
        matrix bu=syminv(x'*x)*x'*lnp;
        matrix g=lnp-x*bu;
        matrix g=diag(g);
        matrix
vv=(vcv1,zz,zz,zz)\(zz,vcv2,zz,zz)\(zz,zz,vcv3,zz)\(zz,zz,zz,vcv4);
```

```
            matrix vs=aa*x'*g*vv*g*x*aa;
            display("Sampling standard errors");
            forvalues i=1(1)4 {;
                    display sqrt(vs[`i',`i']);
                    matrix ses_s[`i',`ibw']=sqrt(vs[`i',`i']);
            };
            **failure of arbitrage standard errors;
            matrix gg=g*g;
            matrix vf=aa*x'*w*gg*w*x*aa;
            display("Failure of arbitrage standard errors");
            forvalues i=1(1)4 {;
                    display sqrt(vf[`i',`i']);
                    matrix ses_f[`i',`ibw']=sqrt(vf[`i',`i']);
            };
            matrix va=vf+vs;
            display("Total standard errors");
            forvalues i=1(1)4 {;
                    display sqrt(va[`i',`i']);
                    matrix ses_a[`i',`ibw']=sqrt(va[`i',`i']);
            };
};

matrix ses_s=ses_s[2..$cn,1..4];
matrix ses_f=ses_f[2..$cn,1..4];
matrix ses_a=ses_a[2..$cn,1..4];


display "Standard errors by country and bandwidth: sampling";
matrix list ses_s;
display "Standard errors by country and bandwidth: failure of
arbitrage";
matrix list ses_f;
display "Standard errors by country and bandwidth: total";
matrix list ses_a;


log close;
```

8. poverty_indexes_fish_sterrs.do

```
**this makes fisher poverty weighted indexes for a range of bandwidths
1 s.d. 0.5 0.1 0.05 s.d.
**I have dropped the smallest bandwidth, which is too small for BGD and
PAK

version 8.2
#delimit ;

set logtype text;
cap log close;
log using poverty_indexes_fish_sterrs, replace;
clear;
set more off;

global cns="ind idn bgd pak";
global cn=4;
```

```
global gn=34;

**Indian poverty line, more generally, base country PL;
scalar pl0=345.08;
scalar lnz=log(pl0);

**Retrieving the calculated estimates of the indexes at each bandwidth;
use pppp_fish;
reshape wide ans, i(bwn) j(cno);
mkmat ans*, mat(torn);
matrix torn=torn';
mkmat bw, mat(bws);
drop _all;

**making the logarithm of price vectors;

use uvs_w_match;
global cno=1;
foreach iso of global cns {;
        ren uv_`iso' p$cno;
        gen lnp$cno=log(p$cno);
        mkmat lnp$cno;
        mkmat p$cno;
        global cno=$cno+1;
};


drop _all;

**constructing the X matrix for the CPD regressions;
**NB this would have to be modified depending on the number of
countries;

matrix lnp=lnp1\lnp2\lnp3\lnp4;
matrix i1=J($gn,1,1);
matrix i0=J($gn,1,0);
**this is with the constant term;
matrix x=(i1\i1\i1\i1),(i0\i1\i0\i0),(i0\i0\i1\i0),(i0\i0\i0\i1);
matrix aa=J(1,$gn-1,0)\I($gn-1);
matrix x=x,(aa\aa\aa\aa);

**making the data sets that are used to calculate the poverty weighted
budget shares;

global cno=1;
foreach iso of global cns {;
        use expsurv_`iso';
        drop if hh_totx==.;
        gen pce=hh_totx/hhs;
        keep x* mult pce sx stratum psu;
        forvalues i=1(1)$gn {;
                gen w`i'=x`i'/sx;
                drop x`i';
        };
        keep w* pce mult stratum psu;
        svyset [pweight=mult], strata(stratum) psu(psu);
        save temp$cno, replace;
```

```
            global cno=$cno+1;
};


**this program makes the variance covariance matrix of the sampling
estimates;
**NB it will need to be altered to suit the number of goods;


cap program drop mkw_vcv;
program define mkw_vcv;
        use temp`1';
        gen lnx=log(pce);
        qui summ lnx [aw=mult];
        scalar hbw=`2'*r(sd);
        gen tau=abs(lnx-`3')/hbw;
        gen kwt=(15/(16*hbw))*(1-tau^2)^2;
        replace kwt=0 if tau >= 1;
        qui summ kwt;
        forvalues i=1(1)$gn{;
                replace w`i'=w`i'*kwt;
        };
        svyratio w1/kwt w2/kwt w3/kwt w4/kwt w5/kwt w6/kwt w7/kwt
w8/kwt w9/kwt w10/kwt w11/kwt w12/kwt w13/kwt w14/kwt w15/kwt w16/kwt
         w17/kwt w18/kwt w19/kwt w20/kwt w21/kwt w22/kwt w23/kwt
w24/kwt w25/kwt w26/kwt w27/kwt w28/kwt w29/kwt w30/kwt w31/kwt w32/kwt
         w33/kwt w34/kwt;
        matrix w_est=e(est);
        matrix w_est=w_est';
        matrix vcv_w=e(V_db);
        drop _all;
end;



matrix ses_s=J(4,4,0);
matrix ses_f=J(4,4,0);
matrix ses_a=J(4,4,0);
matrix eks_fish=J($cn,1,0);
matrix fish=J($cn,$cn,0);

forvalues ibw=1(1)4 {;
        scalar bw=bws[`ibw',1];
        display "Bandwidth is " bw;
        forvalues cno=1(1)$cn {;
                local lnpppp=torn[`cno',`ibw'];
                local lnpl=lnz+`lnpppp';
                mkw_vcv `cno' bw `lnpl';
                matrix w`cno'=w_est;
                matrix vcv`cno'=vcv_w;
        };

        forvalues i=1(1)$cn {;
                forvalues j=1(1)$cn {;
                        matrix
fish[`i',`j']=((w`i')'*syminv(diag(p`i'))*p`j');
                        matrix fish[`i',`j']=log(fish[`i',`j']);
                };
```

```
        };

        matrix fish=0.5*(fish-fish');

        forvalues i=1(1)$cn {;
                matrix eks_fish[`i',1]=0;
                forvalues j=1(1)$cn {;
                        matrix
eks_fish[`i',1]=eks_fish[`i',1]+fish[1,`j']+fish[`j',`i'];
                };
        };
        matrix eks_fish=eks_fish/$cn;
        display ("Poverty line Fisher matrix");
        matrix list fish;
        display("Original EKS Fisher indexes");
        matrix tt=torn[1..$cn,`ibw'];
        matrix list tt;
        display("Fisher EKS indexes");
        matrix list eks_fish;

        **sampling variance for the Fisher EKS;
        **making the betaij's;
        forvalues i=1(1)$cn {;
                forvalues j=1(1)$cn {;
                        matrix r`i'`j'=J($gn,1,0);
                        forvalues n=1(1)$gn{;
                                matrix
r`i'`j'[`n',1]=p`j'[`n',1]/p`i'[`n',1];
                        };
                matrix aa=w`i''*r`i'`j';
                matrix beta`i'`j'=r`i'`j'/aa[1,1];
                matrix drop r`i'`j';
                };
        };
        matrix vs_f=J($cn,1,0);
        **first term;
        scalar a1=0;
        forvalues j=1(1)$cn {;
                forvalues k=1(1)$cn {;
                        matrix aa=beta1`j''*vcv1*beta1`k';
                        scalar a1=a1+aa[1,1];
                };
        };

        **second and subsequent terms;
        matrix a2=J($cn,1,0);
        matrix a3=J($cn,1,0);
        matrix a4=J($cn,1,0);
        matrix a5=J($cn,1,0);

        forvalues i=1(1)$cn {;
                local a2=0;
                local a3=0;
                local a4=0;
                local a5=0;
                forvalues j=1(1)$cn {;
```

```
                              matrix aa=beta`i'`j''*vcv`i'*(beta`i'`i'-
beta`i'1);
                              local a2=`a2'-2*aa[1,1];
                              matrix aa=beta1`j''*vcv1*(beta1`i'-beta11);
                              local a3=`a3'+2*aa[1,1];
                              matrix aa=(beta`j'`i'-
beta`j'1)'*vcv`j'*(beta`j'`i'-beta`j'1);
                              local a4=`a4'+aa[1,1];
                              forvalues k=1(1)$cn {;
                                      matrix
aa=beta`i'`j''*vcv`i'*beta`i'`k';
                                      local a5=`a5'+aa[1,1];
                              };
                      };
                  matrix a2[`i',1]=`a2';
                  matrix a3[`i',1]=`a3';
                  matrix a4[`i',1]=`a4';
                  matrix a5[`i',1]=`a5';
          };

          matrix fd=-1\J($cn-1,1,1);
          matrix vs_f=(a1*fd+a2+a3+a4+a5)/(4*$cn*$cn);
          matrix list vs_f;


          **abritrage failure variance of the tornqvist which is same as
Fisher;

          *first running the unweighted cpd regressions in order to get
the variance matrices;

          matrix aa=J(1,$gn-1,0)\I($gn-1);
          matrix x=x,(aa\aa\aa\aa);
          matrix eu=lnp-x*syminv(x'*x)*x'*lnp;
          matrix e2=diag(eu)*eu;

          **these are the diagomal matrix of squared residuals for each
country;

          matrix om1=diag(e2[1..$gn,1]);
          matrix om2=diag(e2[$gn+1..2*$gn,1]);
          matrix om3=diag(e2[2*$gn+1..3*$gn,1]);
          matrix om4=diag(e2[3*$gn+1..4*$gn,1]);

          matrix a1=J($cn,1,0);
          matrix a2=J($cn,1,0);
          matrix a3=J($cn,1,0);
          matrix a4=J($cn,1,0);
          matrix a5=J($cn,1,0);

          forvalues i=1(1)$cn {;
                  local a1=0;
                  local a2=0;
                  local a3=0;
                  local a4=0;
                  local a5=0;
                  forvalues j=1(1)$cn {;
```

```
                    matrix aa=(w`i'+w`j')'*om`i'*(w1-w`i');
                    local a2=`a2'+2*aa[1,1];
                    matrix aa=(w1-w`i')'*om1*(w1+w`j');
                    local a3=`a3'-2*aa[1,1];
                    matrix aa=(w1-w`i')'*om`j'*(w1-w`i');
                    local a4=`a4'+aa[1,1];
                    forvalues k=1(1)$cn {;
                            matrix
aa=(w`i'+w`j')'*om`i'*(w`i'+w`k');
                            local a1=`a1'+aa[1,1];
                            matrix aa=(w1+w`j')'*om1*(w1+w`k');
                            local a5=`a5'+aa[1,1];
                    };
            };
            matrix a1[`i',1]=`a1';
            matrix a2[`i',1]=`a2';
            matrix a3[`i',1]=`a3';
            matrix a4[`i',1]=`a4';
            matrix a5[`i',1]=`a5';
            if `i'==1 {;
                    matrix a1[`i',1]=-`a1';
            };
        };

        matrix va_t=(a1+a2+a3+a4+a5)/(4*$cn^2);
        matrix list va_t;
        matrix va_f=va_t;



        forvalues i=2(1)4 {;
                display("sampling s.e. country `i'");
                display sqrt(vs_f[`i',1]);
                matrix ses_s[`i',`ibw']=sqrt(vs_f[`i',1]);
                display("arbitrage s.e. country `i'");
                display sqrt(va_f[`i',1]);
                matrix ses_f[`i',`ibw']=sqrt(va_f[`i',1]);
                display("total s.e. country `i'");
                display sqrt(vs_f[`i',1]+va_f[`i',1]);
                matrix ses_a[`i',`ibw']=sqrt(vs_f[`i',1]+va_f[`i',1]);
        };


        **note that arbitrage variance for Fisher is same as for
Tornqvist;
        matrix va_f=va_t;
        display ("Standard errors for EKS-Fisher index");

        forvalues i=2(1)4 {;
                display("sampling s.e. country `i'");
                display sqrt(vs_f[`i',1]);
                matrix ses_s[`i',`ibw']=sqrt(vs_f[`i',1]);
                display("arbitrage s.e. country `i'");
                display sqrt(va_f[`i',1]);
                matrix ses_f[`i',`ibw']=sqrt(va_f[`i',1]);
                display("total s.e. country `i'");
                display sqrt(vs_f[`i',1]+va_f[`i',1]);
```

```
                matrix ses_a[`i',`ibw']=sqrt(va_f[`i',1]+vs_f[`i',1]);
        };

};


display "Standard errors by country and bandwidth: sampling";
matrix list ses_s;
display "Standard errors by country and bandwidth: failure of
arbitrage";
matrix list ses_f;
display "Standard errors by country and bandwidth: total";
matrix list ses_a;


log close;
```

9. poverty_indexes_torn_sterrs.do

```
**this makes the poverty weighted tornqvist index with standard errors:
uses last iteration from poverty_indexes_cpd
**plus code from pppp_sterrs to incorporate standard errors, together
with appropriate bits from cpd_index

version 8.2
#delimit ;

set logtype text;
cap log close;
log using poverty_indexes_torn_sterrs, replace;
clear;
set more off;

global cns="ind idn bgd pak";
global cn=4;
global gn=34;

**Indian poverty line, more generally, base country PL;
scalar pl0=345.08;
scalar lnz=log(pl0);


**Retrieving the calculated estimates of the indexes at each bandwidth;
use pppp_torn;
reshape wide ans, i(bwn) j(cno);
mkmat ans*, mat(torn);
matrix torn=torn';
mkmat bw, mat(bws);
drop _all;


**making the logarithm of price vectors;

use uvs_w_match;
global cno=1;
```

```
foreach iso of global cns {;
        ren uv_`iso' p$cno;
        gen lnp$cno=log(p$cno);
        mkmat lnp$cno;
        global cno=$cno+1;
};

drop _all;


**constructing the X matrix for the CPD regressions;
**NB this would have to be modified depending on the number of
countries;

matrix lnp=lnp1\lnp2\lnp3\lnp4;
matrix i1=J($gn,1,1);
matrix i0=J($gn,1,0);
**this is with the constant term;
matrix x=(i1\i1\i1\i1),(i0\i1\i0\i0),(i0\i0\i1\i0),(i0\i0\i0\i1);
matrix aa=J(1,$gn-1,0)\I($gn-1);
matrix x=x,(aa\aa\aa\aa);

**making the data sets that are used to calculate the poverty weighted
budget shares;

global cno=1;
foreach iso of global cns {;
        use expsurv_`iso';
        drop if hh_totx==.;
        gen pce=hh_totx/hhs;
        keep x* mult pce sx stratum psu;
        forvalues i=1(1)$gn {;
                gen w`i'=x`i'/sx;
                drop x`i';
        };
        keep w* pce mult stratum psu;
        svyset [pweight=mult], strata(stratum) psu(psu);
        save temp$cno, replace;
        global cno=$cno+1;
};

**this program makes the variance covariance matrix of the sampling
estimates;
**NB it will need to be altered to suit the number of goods;


cap program drop mkw_vcv;
program define mkw_vcv;
        use temp`1';
        gen lnx=log(pce);
        qui summ lnx [aw=mult];
        scalar hbw=`2'*r(sd);
        gen tau=abs(lnx-`3')/hbw;
        gen kwt=(15/(16*hbw))*(1-tau^2)^2;
        replace kwt=0 if tau >= 1;
        qui summ kwt;
        forvalues i=1(1)$gn{;
```

```
                    replace w`i'=w`i'*kwt;
            };
            svyratio w1/kwt w2/kwt w3/kwt w4/kwt w5/kwt w6/kwt w7/kwt
w8/kwt w9/kwt w10/kwt w11/kwt w12/kwt w13/kwt w14/kwt w15/kwt w16/kwt
            w17/kwt w18/kwt w19/kwt w20/kwt w21/kwt w22/kwt w23/kwt
w24/kwt w25/kwt w26/kwt w27/kwt w28/kwt w29/kwt w30/kwt w31/kwt w32/kwt
            w33/kwt w34/kwt;
            matrix w_est=e(est);
            matrix w_est=w_est';
            matrix vcv_w=e(V_db);
            drop _all;
end;



matrix ses_s=J(4,4,0);
matrix ses_f=J(4,4,0);
matrix ses_a=J(4,4,0);

forvalues ibw=1(1)4 {;
        scalar bw=bws[`ibw',1];
        display "Bandwidth is " bw;
        forvalues cno=1(1)$cn {;
                local lnpppp=torn[`cno',`ibw'];
                local lnpl=lnz+`lnpppp';
                mkw_vcv `cno' bw `lnpl';
                matrix w`cno'=w_est;
                matrix vcv`cno'=vcv_w;
        };
        matrix b_torn=J($cn,$cn,0);
        forvalues i=1(1)$cn {;
                forvalues j=1(1)$cn {;
                        matrix
b_torn[`i',`j']=0.5*(w`i'+w`j')'*(lnp`j'-lnp`i');
                };
        };
        matrix eks_torn=J($cn,1,0);
        forvalues i=1(1)$cn {;
                matrix eks_torn[`i',1]=0;
                forvalues j=1(1)$cn {;
                        matrix
eks_torn[`i',1]=eks_torn[`i',1]+b_torn[1,`j']+b_torn[`j',`i'];
                };
        };
        matrix eks_torn=eks_torn/$cn;

        display("Matrix of pairwise Torwnvist indexes");
        matrix list b_torn;
        display "Original Tornqvist EKS indexes";
        matrix tt=torn[1..$cn,`ibw'];
        matrix list tt;
        display("Tornqvist EKS indexes");
        matrix list eks_torn;
        **sampling variance of the Tornqvist EKS
        matrix vs_t=J($cn,1,0);
        **first term;
        scalar a1=0;
```

```
     forvalues j=1(1)$cn {;
             matrix t1j=lnp`j'-lnp1;
             forvalues k=1(1)$cn {;
                     matrix t1k=lnp`k'-lnp1;
                     matrix aa=t1j'*vcv1*t1k;
                     scalar a1=a1+aa[1,1];
             };
     };
     **second and subsequent terms;
     matrix a2=J($cn,1,0);
     matrix a3=J($cn,1,0);
     matrix a4=J($cn,1,0);
     matrix a5=J($cn,1,0);
     forvalues i=1(1)$cn {;
             local a2=0;
             local a3=0;
             local a4=0;
             local a5=0;
             matrix t1i=lnp`i'-lnp1;
             matrix vit1i=vcv`i'*t1i;
             forvalues j=1(1)$cn {;
                     matrix aa=t1i'*vcv1*(lnp`j'-lnp1);
                     local a2=`a2'+aa[1,1];
                     matrix aa=(lnp`i'-lnp`j')'*vit1i;
                     local a3=`a3'+aa[1,1];
                     matrix tjivi=(lnp`i'-lnp`j')'*vcv`i';
                     matrix aa=t1i'*vcv`j'*t1i;
                     local a5=`a5'+aa[1,1];
                     forvalues k=1(1)$cn {;
                             matrix aa=tjivi*(lnp`i'-lnp`k');
                             local a4=`a4'+aa[1,1];
                     };
             };
             matrix a2[`i',1]=2*`a2';
             matrix a3[`i',1]=2*`a3';
             matrix a4[`i',1]=`a4';
             matrix a5[`i',1]=`a5';
     };
     matrix fd=-1\J($cn-1,1,1);
     matrix vs_t=(a1*fd+a2+a3+a4+a5)/(4*$cn*$cn);
     matrix list vs_t;

     **abritrage failure variance of the tornqvist;

     *first running the unweighted cpd regressions in order to get the
variance matrices;

     matrix aa=J(1,$gn-1,0)\I($gn-1);
     matrix x=x,(aa\aa\aa\aa);
     matrix eu=lnp-x*syminv(x'*x)*x'*lnp;
     matrix e2=diag(eu)*eu;

     **these are the diagomal matrix of squared residuals for each
country;

     matrix om1=diag(e2[1..$gn,1]);
     matrix om2=diag(e2[$gn+1..2*$gn,1]);
```

```
        matrix om3=diag(e2[2*$gn+1..3*$gn,1]);
        matrix om4=diag(e2[3*$gn+1..4*$gn,1]);

        matrix a1=J($cn,1,0);
        matrix a2=J($cn,1,0);
        matrix a3=J($cn,1,0);
        matrix a4=J($cn,1,0);
        matrix a5=J($cn,1,0);

        forvalues i=1(1)$cn {;
                local a1=0;
                local a2=0;
                local a3=0;
                local a4=0;
                local a5=0;
                forvalues j=1(1)$cn {;
                        matrix aa=(w`i'+w`j')'*om`i'*(w1-w`i');
                        local a2=`a2'+2*aa[1,1];
                        matrix aa=(w1-w`i')'*om1*(w1+w`j');
                        local a3=`a3'-2*aa[1,1];
                        matrix aa=(w1-w`i')'*om`j'*(w1-w`i');
                        local a4=`a4'+aa[1,1];
                        forvalues k=1(1)$cn {;
                                matrix aa=(w`i'+w`j')'*om`i'*(w`i'+w`k');
                                local a1=`a1'+aa[1,1];
                                matrix aa=(w1+w`j')'*om1*(w1+w`k');
                                local a5=`a5'+aa[1,1];
                        };};
                matrix a1[`i',1]=`a1';
                matrix a2[`i',1]=`a2';
                matrix a3[`i',1]=`a3';
                matrix a4[`i',1]=`a4';
                matrix a5[`i',1]=`a5';
                if `i'==1 {;
                        matrix a1[`i',1]=-`a1';
                };
        };

        matrix va_t=(a1+a2+a3+a4+a5)/(4*$cn^2);
        matrix list va_t;
        forvalues i=2(1)4 {;
                display("sampling s.e. country `i'");
                display sqrt(vs_t[`i',1]);
                matrix ses_s[`i',`ibw']=sqrt(vs_t[`i',1]);
                display("arbitrage s.e. country `i'");
                display sqrt(va_t[`i',1]);
                matrix ses_f[`i',`ibw']=sqrt(va_t[`i',1]);
                display("total s.e. country `i'");
                display sqrt(vs_t[`i',1]+va_t[`i',1]);
                matrix ses_a[`i',`ibw']=sqrt(vs_t[`i',1]+va_t[`i',1]);
        };
};


display "Standard errors by country and bandwidth: sampling";
matrix list ses_s;
```

```
display "Standard errors by country and bandwidth: failure of
arbitrage";
matrix list ses_f;
display "Standard errors by country and bandwidth: total";
matrix list ses_a;


log close;
```