International Comparison Program

[09.01]

Using unit-values to assess spatial price differences:

evidence from India and Brazil

Angus Deaton & Olivier Dupriez

5th Technical Advisory Group Meeting
April 18-19, 2011

Washington DC

Using unit-values to assess spatial price differences: evidence from India and Brazil

Angus Deaton

Olivier Dupriez

Princeton University and World Bank

April 2011

1. Introduction

The 2005 round of the International Comparison Program left unresolved the question of within-country differences in price levels, something that is plausibly important in large poor countries. To date, the ICP's target has been national price levels but in some countries, fewer rural prices were collected, which compromises comparisons between countries with and without complete coverage, for example between a small country like Singapore and large countries like China, Brazil, or India, Deaton and Heston (2010). More fundamentally, the ICP could conceivably move from national price levels to regional or sectoral price indexes, effectively treating regions, provinces, or rural and urban sectors as separate "countries." In a large country such as India, consumption patterns are sharply different across areas, and in some cases, such as the calculation of poverty rates, the Government of India has always worked under the assumption that rural price levels are lower than urban price levels. More generally, the World Bank's poverty counts also recognize urban/rural price differences in counting the number of people in poverty, at least for large countries. The Bank's adjustments sometimes rely on within country differences in urban and rural poverty lines, whose basis is sometimes unclear, and whose reliability is questionable. Nevertheless, there is a real issue here: people living below the global poverty line are more likely to be in rural than in urban areas, and we should like to take into account differences in the prices that they face. And for the same reasons that we expect price levels to

be lower in poorer countries—the Balassa-Samuelson theorem—we would expect prices to be lower in poorer rural areas than in richer urban areas.

In principle, information on geographical price variation within countries ought to be available from national data, given that most countries collect price data throughout their territories. Yet these data are collected, not to calculate geographical price differences, but to calculate inflation rates for different areas, without attempting to estimate spatial price differences in the base year. In these cases, either the raw price quotes are unavailable, or even if available, are not collected in a form that allows spatial comparisons.

Information on prices can also be gleaned from household expenditure surveys that collect information on both expenditures and quantities because unit values can be calculated as ratios of one to the other. Such information is not available for all countries, and does not cover all of consumption, typically only food and (sometimes) a few other items (cigarettes, alcohol, clothing, fuels) where units are readily defined. Nor is a unit value the same thing as a price. In particular, the categories in household surveys are typically broader than the narrowly defined items for which prices are collected in the ICP, so that quality is not held constant across different people or places. Richer people tend to buy higher qualities and report higher unit values, so that there is a danger that what look like higher prices in the city are in fact quality differences driven by higher incomes. In spite of these drawbacks, the attraction of unit values is the large number of observations that are collected in household surveys, and that they can be

stratified by location and by characteristics of households. In the Indian household survey that we explore below, there are more than 5 million unit values, each of which is tied to a specific household whose characteristics we know. The World Bank's poverty calculations also use household survey data to measure living standards but rely on purchasing power parity exchange rates that are based on national prices that are not linked or tailored to the household data that provide the nominal level of living standards.

This paper looks at the unit values from two large countries, India and Brazil.[1] We tabulate differences in unit values between urban and rural areas, as well as (in India) between major states, and we attempt to sort out how much of the differences is attributable to differences in quality driven by differences in incomes, and how much comes from genuine differences in prices. For India, Dikhanov (2010) has carried out a similar analysis, and argues that, once income effects are taken into account, there is only a negligible difference (3.2 percent)

---

[1] Data for India are from the *Socio-Economic Survey, sixty-first round*, conducted by the National Sample Survey Organization (NSSO) from July 2004 to June 2005 on a sample of 124,644 households. Data on food consumption were collected using a 30-day recall period. In the Consumer Expenditure Schedule, household consumption data are broken down into 340 goods and services, including 146 food items and non-alcoholic beverages.

Data for Brazil are from the *Pesquisa de Orçamentos Familiares 2002-2003*, conducted by the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística, IBGE) from July 2002 to June 2003 on a sample of 48,470 households. Data on food consumption were collected using the diary method. In the household questionnaire and in the household and individual diaries, consumption data are broken down into 6,927 products and services, including 3,131 food items and non-alcoholic beverages.

Both surveys have nationwide coverage.

between urban and rural prices, a difference that is small enough to ignore for the purposes of the ICP. This conclusion stands apart from a long line of calculations from Indian NSS data over many years, including the latest official poverty calculations, in which researchers have typically found a difference in prices between urban and rural of about 15 percent. None of these earlier investigations adjust unit values for income effects, so that the 15 percent for India is likely too large. However, in a series of papers beginning with Deaton (1988), one of us estimated income effects on unit values for a number of countries, including for India, Deaton (1997, Chapter 5), and typically found only small income effects. On the basis of this, subsequent work on price comparisons for India simply ignored the income effects, Deaton and Tarozzi (2005), and since this work (with only minor modifications) has been adopted by GOI (2009) as the basis for their new poverty lines, it is important to revisit this question and to assess the size of the quality effects.

In the analysis that follows, we do find substantial income effects in both India and Brazil, so that correction for income-driven quality differences reduces the urban to rural price differential that would be calculated from treating unit values as if they were prices. However, even with allowance for quality differences, we find substantially higher urban than rural prices, around 10 percent for food. In the Indian case, a major factor driving the differences between our results and Dikhanov's, is the treatment of (imputed) prices for own-produced goods. In both India and Brazil, there is a substantial fraction of consumption that is not obtained through cash purchases, the value of which was imputed at the ex farm or ex

factory rate. These prices are lower than unit values for the same category bought for cash because such goods are obtained net of transportation and distribution margins, and because, in some cases, the goods may be of lower quality. Home-produced goods are much more important in rural than in urban areas, so that including their unit values gives a larger urban price differential than when only cash items are used. Similarly, the ICP, which collects no prices for home-produced items, will find understate urban to rural price differentials, even in those cases where it comprehensively samples retail outlets rural areas. Note that for national income accounting purposes, home produced goods are within the production boundary, and so are included in consumption and GDP, and that prices excluding distribution and transportation costs are the appropriate prices for them. Accessing cheaper local produce from one's own or neighbors' plots genuinely reduces the cost of living for rural people.

2. Unit values and prices for major foods in India

Table 1 lists the calculations for the eleven most important food commodities in India, as ranked by their average budget shares over the whole country. The budget shares and unit values in this table include all consumption items, whether purchased for cash, home-grown, or received in some other way, for example as a gift or in barter exchange. We also include purchases through the public distribution system at subsidized prices in "fair price" shops; the important items here are rice, wheat or wheat flour, and sugar. The first two columns list the budget shares for each sector, rural and urban. Median and mean log unit values are

shown, together with the difference between them. Note that the median of the log is identical to the log of the median, so that when working with medians, it is of no consequence whether we work with prices or with their logarithms.
One issue about medians which will become apparent later is that they are sometimes identical across sectors, even when one distribution is clearly to the right of the other. This happens because many quantities and expenditures are reported as whole numbers, so that there is sometimes a single unit value that attracts a substantial fraction of the sample. Suppose that 20 percent of the sample reports x, 40 percent reports x–1, and 40 percent reports x+1, so that the median is x. If the distribution becomes 35 percent below x, 20 percent at x, and 45 percent above x, a clear upward shift, the median remains at x. For this reason, we focus on the means of logarithms of prices, which differs from the procedures reported in Deaton and Tarozzi (2005), for example.

For these eleven goods, urban prices are 17 percent higher than rural prices using the medians, and 19 percent higher using the means of logs; these averages come from the differences for each commodity weighted by their budget shares. These estimates are in line with standard Indian numbers, and as we show below, they do not change very much when we include all foods. Both estimates are two percentage points lower if the publicly distributed items are excluded. As we shall see below, foods obtained through fair price shops are much more important to rural than to urban consumers. It is worth also noting that the urban rural difference is not the same for all goods. For example, for the two major cereals, rice and wheat, both of which are produced in the rural sector, the difference in

the mean log unit values is 28 and 26 percent. The differences for milk and for fish are 19 and 18 percent. Processed foods like sugar and mustard oil show little or no difference. Tea leaf costs pretty much the same in urban and rural, but tea cups, which embody more labor costs, are cheaper in the poorer sector.

These calculations make no allowance for income driven quality effects in the unit values. These are investigated in Table 2 which shows the results of running a household-level regression of the log of each unit value on the logarithm of household per capita expenditure and an urban dummy. The coefficient on the latter is an estimate of the difference in log unit values between the urban and rural sectors once income effects have been removed and can be compared with the unadjusted difference in logs in Table 1, reproduced here as the third column in Table 2. Note that the ability to run these household-level regressions is another advantage of working with the logs of unit values, rather than with medians, which are obviously not available at the household level.

The income elasticities of the unit values are in some cases substantial, more than 26 percent for rice, 18 percent for cups of tea, 17 percent for wheat/atta, and 11 percent for fish, all categories where there is much quality variation within the category and scope for better-off people to buy more expensive varieties. These estimates are generally higher than those estimated in Deaton (1997, Chapter 5) perhaps because of an expansion of varieties as India has become richer compared with Maharashtra in 1983.) Because of the income effects, the adjusted urban/rural differences in column 2 are smaller than the unadjusted differences in column 3, by almost a half in the case of rice, and by around a third over all these

goods. The overall price difference, weighted by an average of the rural and urban budget shares, falls from 19.2 percent with no adjustment to 11.5 percent with the adjustment.

3. Unit values by source in India

The analysis in Section 2 used all unit values for food consumption, irrespective of how the item was obtained. Here we look more closely at unit values by source, with particular attention to the difference between cash purchases and other sources, mostly home production, and between free-market prices and the prices of goods purchased in fair price shops and provided by the Public Distribution System. Table 3 shows the nine cases where there is a non-trivial budget share for non-cash or for subsidized purchases. Sugar, wheat or wheat flour, and rice are provided through the Public Distribution System. There are also significant amounts of non-cash consumption (nearly all of which is of home-produced items) of rice, wheat, and milk (where the non-cash share is larger than the cash share in the rural sector), and much smaller amounts of mustard oil, sugar, and fish. There are substantial unit value differences across sources. Sugar is about 30 percent cheaper from the PDS, and rice and wheat are less than half the price, with the percentage discount larger in urban areas; that said, the budget shares of the PDS commodities are very small among urban households. Even for rural households, the share of expenditure from the PDS is less than ten percent of the share from other sources. As is to be expected, home-produced goods are also recorded at

lower prices, though the differences are relatively modest, 14 percent for rice, 26 percent for wheat, and 9 percent for milk.

We can also calculate the urban price differentials that would be found if we were to use only the unit values from cash purchases. While this is not a desirable calculation in itself—the prices of subsidized and own-produced goods should be included in the price indexes—it is comparable to the differentials that would be observed if only "shop" prices were collected, as in the ICP itself. Calculations of urban/rural price differentials have been carried out by Hill and Syed (2010) using ICP 2005 data for a group of Asian countries for all goods and services (not just food) and they find only small differentials, between two and three percent averaged over five countries, estimates that they regard as implausibly low.

Table 4 reproduces the Indian results in Table 2, but using only the cash prices, and omitting purchases from the Public Distribution System. As anticipated, this reduces both the raw and adjusted differences in the means of the logarithms of the unit values. For example, for wheat and *atta*, where the unadjusted and adjusted differences in Table 2 are 26 and 16 percent respectively, the corresponding numbers for cash only purchases are only 14 and 5 percent. For rice, the numbers fall from 28 and 14 percent to 22 and 9 percent. These two groups of goods are the ones that are most affected, but together they comprise a substantial fraction of the food budget, so the move to cash prices has a correspondingly substantial effect on the overall index of urban to rural prices. At the end of the last section, we reported the overall unadjusted and adjusted indexes as 19.2

and 11.5 percent. Using cash prices only, these estimates fall to 13.5 and 6.8 percent.

4. Extension to all foods in India

We began with selected important food items so that the calculations could be transparent. However, the NSS survey provides unit values for 147 foods which together account for 61.3 percent of the rural budget, and for 50.9 percent of the urban budget, both figures calculated from the household budget shares and averaged over households. The survey collects unit values on a few other items, such as tobacco, some fuels, and some clothing, but the total of these items is small, and some of the unit values have proved problematic in the past. For much of the non-food budget, no unit values are collected. So for the moment, we confine ourselves to the foods. Even among those, there are likely some where the unit values are unreliable, but each of the remaining budget shares is small, and so the effect will be limited.

For all 147 foods, the average of the unadjusted urban/rural price difference, with averaged budget shares as weights (the Törnqvist index) is 17.6 percent, only a little less than the average for the eleven goods. With adjustment for income effects, following the same procedures as in Table 2, the difference is 10.2 percent, so that without correction for income effects, the price differential is overstated by about 70 percent. Allowing for income is important, but does not remove the need to make an adjustment, particularly for a country such as India

where a small change in the poverty line has dramatic consequences for the poverty counts.

As in the previous section, we have also recalculated the food price for urban using only the cash purchases and treating the subsidized PDS goods as separate commodities, rather than combining their unit values with the unit values for the corresponding market goods. As before, this treatment reduces the size of the urban differential, to 13.4 percent unadjusted and 6.7 percent when adjusted for income driven quality effects. This last figure is closer to but remains larger than Dikhanov's estimate of 3.2 percent using cash-only prices (we shall return to the resolution of the discrepancy in Section 7 below). Our cash estimate is close to the 6.9 percent that Hill and Syed (2010) calculate for all prices in Malaysia, but substantially higher than their estimates for Vietnam, Indonesia, Philippines, and Sri Lanka. (It is hard to know what to expect from adding in non-foods; manufactured goods will often be somewhat more expensive in rural areas, while services—like cups of tea—will often be substantially cheaper in the countryside.) Even so, the main point of these alternative calculations is to show that the omission of non-cash and subsidized purchases can bias downwards the estimate of urban prices relative to a rural base. In the current Indian case, a cash only treatment reduces the measured urban price differential by a third. Of course, there is no reason to suppose that such a result will hold everywhere. While own-produced foods will always be cheaper, and always be more common in the countryside, the Indian food subsidy system—which favors rural consumers—will not carry through to other countries.

5. Alternative calculations using decile groups and median unit values

An alternative, and in some ways more flexible procedure, is used by Dikhanov (2010) who groups the population into deciles of per capita total expenditure, with urban and rural households combined so that households in different sectors but in the same decile have roughly the same per capita expenditure, and then to calculate median or mean unit values for each decile group in each sector. The comparison then yields an urban/rural price difference for each decile group, and these can be averaged to give an estimate for the country as a whole. While this method does not allow analysis at the household level, it has other potential advantages, including the less parametric treatment of income effects—allowing them to be different at different points of the distribution—and possibly also the use of medians, which are potentially more robust than working with logs, though logs will also downweight some extreme values.

The results are shown in Table 5. The urban rural difference in log medians does indeed vary by decile groups, and tends to be particularly large among the best off households, perhaps because it is this group that most takes advantage of the richer set of varieties that are available in urban areas. This also raises the possibility that there are further income effects *within* this group, and suggests that there might be payoffs to a more flexible specification of income effects in the previous analysis. Note also the previously noted problem with medians highlighted by the exact zeroes in the table. In some of those cases, there are positive differences in means of logs within the decile group, so that it is possible

that the medians are not sufficiently sensitive to differences in the distribution; this can happen even when the differences are not zero, because the calculation can focus on the difference between the two most common unit values in each sector. Even so, the overall number in the bottom right of the table is 8.6 percent, not much less than the 10.2 percent estimated from the earlier analysis using the means of logs and regression to control income effects.

We have also replicated Table 5 with differences of log means, rather than differences in log medians (results not shown.) The results differ in detail, but little overall. As in Table 5, the largest difference is in the top decile, and both show a 16 percent overall difference. The overall urban/rural price difference is now estimated to be 9.0 percent, as opposed to 8.6 percent in Table 4 using the medians. In spite of the non-parametric advantages of the deciles, our current preference is for the parametric approach using the logarithm of per capita expenditure to control the income effects, and the associated urban/rural price difference for all foods of 11.5 percent.

6. Differences in food prices across states

In a large country like India, geographical differences in prices may be as important as differences within states between urban and rural. This issue is addressed for India's largest states in Table 6.

The left hand panel shows unadjusted results. These were calculated by averaging the budget shares and the logs of the unit values by commodity by state by sector, and then using the averages to calculate multilateral (Törnqvist GEKS)

indexes. Each sector of each state is considered a separate geographical unit ("country") in the calculations which therefore automatically yield a set of urban rural differentials by state. (We excluded rural Delhi from the calculations, so there is no urban rural comparison for Delhi.) Rural Jammu and Kashmir is taken as the numeraire with a price level of 1.00.

In the rural sector, the northern states of Jammu and Kashmir and Himachal Pradesh have the highest food prices, while the lowest prices are recorded in some of the poorest states, Orissa, Bihar, UP, MP, Rajasthan and Jharkhand. The food price level in Orissa is eighty percent of the price level in Jammu and Kashmir. In the urban sector, Delhi has the highest food prices, 1.15 times the rural price level of Jammu and Kashmir. The lowest, 0.95, is again in Orissa, so that the gap between lowest and highest is about the same in both sectors. These estimates are unadjusted for income-related quality differences.

To correct for quality, we ran a series of household level regressions, one for each commodity, of the logarithm of the unit value on the logarithm of per capita total household expenditure, and on a set of dummies, one for each sector of each state. We then used the estimates to calculate an adjusted average log price for each good in each state and sector by calculating what the average log unit value would be for that state and sector if everyone in the state and sector had the Indian average of log per capita total expenditure. We then inserted these adjusted prices into the multilateral calculation in the same way as for the unadjusted prices.

The results are shown in the right-hand panel. Apart from some shrinkage of the variance, with the poorer states price levels somewhat elevated, the adjusted price levels are quite similar to the unadjusted price levels. The range from top to bottom is now 18 percentage points in the rural sector (UP to Gujarat), compared with 20 percentage points in the unadjusted numbers, and 19 points (Orissa to Assam and Maharashtra) compared with 25 in the urban sector. As with the urban rural price differential, the state differences are surely large enough to be concerned about, at least in some contexts.

These estimates also show that the urban prices, although always higher than rural prices, even after the adjustment, are differentially so in different states. In Kerala, where urban and rural sectors are not sharply defined, the difference is only 4 percent. In Maharashtra, by contrast, which covers Mumbai on the one hand, and deep rural areas on the other, urban prices are 18 percent higher than rural prices.

Given that there has been a concern about the lack of consistency of definition of rural versus urban in different countries, these estimates are of some interest because the methodology does not require consistency. By treating each sector of each state as a separate "country," it simply works with whatever are the local definitions and delivers a price differential that matches those definitions, which is what would be required for policy calculations, for example on levels of poverty. In the Indian context, the urban rural distinction is very different in Kerala from Maharashtra, and that is reflected in the estimated price differentials.

7. Comparative results from Brazil

We have replicated the Indian analysis with data from a 2002 expenditure survey from Brazil. Like India, Brazil is a large country, where the neglect of the urban-rural distinction might have important consequences for the ICP. Its per capita GDP is around three times that of India, so that food is a less important share of the budget. Higher incomes also imply that processed foods are relatively more important than in India. The lesser importance of food affects the relevance of the unit-value based methods used here, which are effectively available only for foods. At the same time, the increasing importance of processed food is likely to move urban and rural prices closer together, or perhaps even to lead to prices being higher in the countryside if they are transported in rather than transported out. In rich countries, like the US, spatial differences in food prices are typically thought to be unimportant, and the urban-rural price differential is largely about housing, not food.

Table 7 presents, in abbreviated form, the same information that was earlier shown for India. The table lists the eleven most important foods, ranked by their budget shares averaged over urban and rural. These are the largest of 38 foods that we consider, which together comprise 16 percent of the rural, and 9 percent of the urban total consumption budget. The first and second columns list the average budget shares, while the third lists the difference between the means of the logs of the unit values for each of the food groups. For four of the eleven goods—soybean oil, crystal sugar, ground coffee, and frozen chicken—unit values are *lower* in the urban sector. For the seven other categories—all of which are or

can be locally produced—rural prices are lower, sometimes very substantially so. Fresh milk unit values are 23 percent lower, and cassava flour unit values 18 percent lower in the countryside. If we take all 38 foods together, and weight by the averages of the budget shares to give a Törnqvist index, these unadjusted unit values are 7.09 percent higher in the cities, the number listed at the foot of the third column. We followed our earlier procedures to adjust these premia, running household level regressions of the each food's unit values on the logarithm of per capita total consumption and on the logarithm of household size as well as an urban dummy, and we report the coefficient on the latter in column 4. As in India, this quality correction accounts for some of the urban premia, and the overall Törnqvist index of urban to rural prices falls to 4.54, shown at the foot of the fourth column.

The fifth and sixth columns repeat these calculations using only those cases where the good was obtained for cash, rather than home-grown or received as a gift. As was the case in India, the difference between urban and rural unit values is smaller when home-grown produce is excluded. In general, column 5 shows the same pattern of premia over goods, but is attenuated relative to column 3. The urban to rural Törnqvist using the cash unit values is only 4.15 percent, which falls to only 1.40 percent once the unit values are adjusted for quality effects. In Brazil, even more so than in India, the difference between urban and rural food prices largely boils down to the importance of home-grown food. The relative unimportance of quality effects in Brazil is probably also attributable in part to

differences in survey design; the Brazilian survey distinguishes 3,131 products compared with "only" 146 in India.

8. Conclusions and theory

Unit values contain useful information about prices, particularly for foods. However, unit values are not prices, and in particular they are contaminated by quality effects, and the fact that better-off households buy more expensive varieties. And because urban households are on average better-off, a comparison of unit-values without adjustment will overstate the urban/rural price difference. According to the calculations here, a suitable correction removes about a third of the difference, reducing a crude estimate for the all foods index from 19.2 percent to 11.5 percent. In Brazil, the unadjusted price difference is smaller, 7.1 percent, and is reduced after adjustment to 4.5 percent.

In both India and Brazil, home produced goods—which are cheaper than purchased goods—make an important contribution to the differential, and the differentials are smaller—in the case of Brazil, which has an extraordinary level of detail in its consumption questionnaire, small enough to ignore—when we work only with unit values of items bought for cash. In India, the subsidized prices in fair price shops are also important, especially in rural areas. For ICP purposes, the prices of home-grown (and subsidized) items ought to be included. That they are lower through the absence of transport and distribution margins confers a genuine benefit to people who grow and consume these goods, and that benefit is not available to people who must purchase those goods in shops and markets.

We suspect that these issues are less important for rich countries, and indeed the comparison of Brazil and India suggests that Brazil lies somewhere between the US and India in this respect. But there are a number of large, poor countries, such as India, China, Pakistan, Bangladesh, Indonesia, and Nigeria, where rural urban price differentials are likely to be important for the ICP. Some of these countries have data that would permit a replication of the methods of this paper although it should be noted that not all consumption surveys collect quantity as well as expenditure information. There are also a number of surveys, especially in Africa, where there exists quantity information, but it is difficult or impossible to use because the units are not standardized across all observations.

We conclude with a brief reflection on the theory underlying what we have done, something that arguably should have appeared at the beginning. It is clearly correct that unit values contain information about prices, and it is equally clearly correct that households that have higher incomes—or higher total expenditures—usually have higher unit values, a fact that can reasonably be attributed to their buying higher quality items within a heterogeneous group of foods, or buying similar items in more upscale outlets. What is less clear is whether the regressions that we use for the adjustment, which pool urban and rural households to regress the log unit value on an urban dummy and on the log of per capita total household expenditure, provide the right estimates of urban to rural price differentials. The key issue is that per capita expenditure may not mean the same thing in urban and rural sectors, in part because of the very fact that prices are different, but more fundamentally because the patterns of urban life are different from rural life.

Suppose that the unit value of good *i* is the sum of a log price component and a log quality component, and that, in the simplest possible case, log quality is a linear function of real income, which is money income deflated by the price index for the sector in which the household lives. If so, we can write

$$\ln v_{ih} = \ln p_{is} + \alpha_i + \beta_i(\ln y_h - \ln P_s) \quad (1)$$

where *h* is a household, *i* is the good, *s* is the sector for urban and rural, *v* is unit value, *y* is income, and P_s is the sector price index for all goods, which is taken to be 1 in the rural sector. The sector price for good *i*, p_{is}, is also taken to be 1 in the rural sector, and is the commodity-specific urban price premium. Equation (1) can be rearranged to correspond to and to interpret the regressions that we have run in this paper. In particular, we have

$$\ln v_{ih} = (\ln p_{is} - \beta_i \ln P_s) + \alpha_i + \beta_i \ln y_h \quad (2)$$

Equation (2) implies that when we run the unit value regressions on an urban dummy and on the logarithm of income, the estimated dummy is *not* the log of the urban price premium for that good, but the log of the urban price premium minus a term that is the product of the income elasticity of the unit value and the log of the overall price index of urban prices. Hence

$$\ln \tilde{p}_{is} = \ln p_{is} - \beta_i \ln P_s \quad (3)$$

where the superimposed tilde indicates an estimated amount. In the typical case, where the income elasticity of the unit value is positive, and prices are higher in cities, this will mean that the procedures in this paper underestimate the urban price premium for each good, at least if (1) is the true model. Without knowing

what the overall price index is, it is difficult to guess the size of the bias. One example is rice in Table 2, where the estimated price index is 0.14, and the income elasticity of the unit value is 0.26, so that if the true all-good index were twice as high in urban as in rural India, for example, the correct price differential for rice would be 0.14 plus 0.26 log 2, or 0.32, and our estimate is less than a half of the truth.

The overall price index is in part determined by the food prices, about which we do know something. In particular, the overall price index can be written

$$\ln P_s = \sum w_i^f \ln p_{is} + (1 - w^f) \ln \pi_s^n \tag{4}$$

where w_i^f is the share of the food *i* in the total budget, w^f is the share of total food, and π_s^n is the urban price index for non-food. Combining (3) and (4) gives an expression for the overall urban price index

$$\ln P_s = \frac{w^f \ln P_s^T + (1 - w^f) \ln \pi_{sn}}{1 - \sum w_i^f \beta_i} \tag{5}$$

where $\ln P_s^T$ is the log Törnqvist index for food that we have been estimating here. Equation (5) shows that the transition from the calculated food index to the overall index depends not only on the price of the omitted non-foods, but also on the size of the quality corrections. However, if the denominator in (5) is small, which will typically be the case, the formula reduces to the standard one of the weighted average of the estimated food index and the unobserved non-food index.

Differential housing costs are clearly an important part of the overall price index for urban relative to rural. Housing is, of course, a difficult commodity for

the ICP, and the surveys are also unhelpful. For example, in India, most rural households report no rent of any kind, and while it is sometimes possible to use survey data to impute user costs, the calculations typically depend on hedonic regressions estimated using very few likely unrepresentative households. In India, the NSS statisticians provide some measure of overall rent for urban households. We have not included this in household total expenditure in the calculations in this paper, but if we were to do so, as did Dikhanov (2010), the price differentials for the foods would have been estimated to be much lower. This is exactly the same effect as shown in equations (2) and (3), but with the sign changed, because the inclusion of urban rents scales up urban incomes.

The general problem here is that we have no way of knowing when an urban income is equivalent to a rural income, in part because we do not have an estimate of the overall cost of living in urban relative to rural, but also because we have no way of allowing for other costs of living in a city, such as the costs of transportation.

One solution is to identify some observable quantity that indicates welfare across the sectors. An imaginative procedure along these lines has been used for by Coondoo, Majumdar, and Chattopadhyay (2011) for India, and by Almas and Johnsen (2010) for China. These papers go back to work by Costa (2001) and by Hamilton (2001) that measures price indexes by assuming that the budget share of food indicates welfare, so that rural households and urban households with the same food shares have the same level of welfare; comparison of their incomes thus estimates the price index. As Coondoo et al argue, this procedure allows the

calculation of price indexes without any information on prices, though whether or not this is an advantage or disadvantage is arguable. But there is a long literature that argues that the food share is *not* a good indicator of well-being. For example, if children consume relatively more food than adults, fully compensated households with more children will have larger food shares than households with fewer children, so that it is not correct to compare food shares across rural and urban sectors if the former have more children. More seriously still, rural households are more likely to do heavy manual labor in agriculture, which requires a large calorie intake, and correspondingly high food shares. Once again, we can imagine two equally well-off households, one rural and one urban, with the former working hard in agriculture, and the latter employed in some unpleasant but relatively sedentary urban occupation, or perhaps just condemned to spend many hours commuting to work. By assumption, they are equally well-off, but the former will have a much larger food share. It would take a very large urban price index to bring the urban worker's food share up to the rural worker's food share, so that the method will overestimate urban prices. The procedure has also been criticized on similar grounds in the poverty line literature by Ravallion and Bidani (1994). Note too that Engel's Law postulates that the food share declines with real income. The proposition that households with identical food shares are equally well-off was also asserted by Engel, and is sometimes referred to as Engel's second law to distinguish it from Engel's first law, from which it does not follow, Deaton and Paxson (1998). But Engel's assertion is false.

Unfortunately, it is easier to see why Engel's assertion is wrong than to propose a soundly based alternative. So we remain some way from having a sound method of measuring urban rural differences in the cost-of-living. This paper has had a more modest aim, to show that unit values contain information about prices, and that urban unit values are higher than rural unit values in large poor countries, and that the difference is reduced but not eliminated by a simple income correction.

9. Citations

Almas, Ingvild, and Ashild Auglænd Johnsen, 2010, "China's growth miracle—still awaiting the great leap forward," available at http://www.iariw.org

Coondoo, Dipankor, Amita Majumder and Somnath Chattopadhyay, 2011, "Estimating spatial consumer price indices through Engel curve analysis," *Review of Income and Wealth*, 57(1), 138–55.

Costa, Dora, 2001, "Estimating real income in the United States from 1888 to 1994: correcting CPI bias using Engel curves," *Journal of Political Economy*, 109(6), 1288–1310.

Deaton, Angus, 1988, "Quality, quantity, and spatial variation in price," *American Economic Review*, 78 (3): 418–30.

Deaton, Angus, 1997, *The analysis of household surveys*, Johns Hopkins University Press for the World Bank.

Deaton, Angus and Alan Heston, 2010, "Understanding PPPs and PPP-based national accounts," *American Economic Journal Macroeconomics*, 2(4), 1–35.

Deaton, Angus, and Christina Paxson, 1998, "Economies of scale, household size, and the demand for food," *Journal of Political Economy*, 106(5), 897–930.

Deaton, Angus and Alessandro Tarozzi, 2005, "Prices and poverty in India," Chapter 16 in Angus Deaton and Valerie Kozel, eds., *The great Indian poverty debate*, New Delhi, Macmillan. 381–411.

Dikhanov, Yuri, 2010, "Income-effect and urban-rural price differentials from the household survey perspective," ICP Global Office, February.

Government of India, 2009, Report of the expert group to review the methodology for the estimation of poverty, Planning Commission, November.

Hamilton, Bruce W., 2001, "Using Engel's law to estimate CPI bias," *American Economic Review*, 91(3), 619–30.

Hill, Robert J., and Iqbal Syed, 2010, "Improving international comparisons of real output: the ICP benchmark and its implications for China," University of New South Wales, Australian School of Business, Research Paper No. 2010 ECON 25, http://ssrn.com/abstract=1728428

Ravallion, Martin and Benu Bidani, 1994, "How robust is a poverty profile?" *World Bank Economic Review,* 8(1), 75–102.

Table 1

Budget shares, median and mean log unit values for urban and rural India

	Budget Shares in total expenditure		Median log unit value			Mean log unit values		
	Rural	Urban	Rural	Urban	Difference	Rural	Urban	Difference
Rice	12.99	7.32	2.28	2.48	0.21	2.23	2.51	0.28
Milk	7.40	7.44	2.48	2.64	0.15	2.45	2.64	0.19
Wheat/atta	6.38	5.28	1.97	2.30	0.33	2.01	2.07	0.26
Mustard oil	2.63	1.46	4.01	4.01	0	4.00	4.00	–0.00
Sugar	2.24	1.84	2.89	2.89	0	2.89	2.89	0.02
Potatoes	1.73	1.01	1.95	2.08	0.13	1.87	1.99	0.11
Other oil	1.51	1.55	3.91	3.95	0.04	3.91	3.97	0.06
Fish	1.37	1.00	3.69	3.91	0.22	3.68	3.86	0.18
Arhar	1.24	1.12	3.40	3.42	0.02	3.37	3.42	0.06
Tea (leaf)	1.05	1.04	–2.04	–1.97	0.07	–2.04	–1.96	0.08
Tea (cups)	0.81	1.21	0.41	0.69	0.29	0.45	0.71	0.26
Total/Mean	39.36	30.28			0.17			0.19

Notes: The budget shares are shares in total expenditure, averaged over all households in each sector. The sums in the last row are for these eleven goods only. The means at the foot of the two difference columns are the averages of the differences weighted by the averages of the rural and urban budget shares in columns 1 and 2. Purchases made in fair price shops are excluded.

Table 2

Income effects on unit values and adjusted urban rural differences: selected goods

	Regression coefficient on lnpce	Regression coefficient on urban dummy	Unadjusted difference from Table 1
Rice	0.262	0.14	0.28
	(154)	(60.7)	(125)
Milk	0.082	0.15	0.19
	(44.0)	(64.9)	(88.7)
Wheat/atta	0.174	0.16	0.26
	(73.9)	(52.1)	(88.6)
Mustard oil	–0.008	0.00	–0.00
	(6.7)	(1.2)	(1.5)
Sugar	–0.003	0.00	0.00
	(3.0)	(2.4)	(1.3)
Potatoes	0.084	0.07	0.11
	(40.4)	(40.4)	(45.0)
Other oil	0.072	0.02	0.06
	(52.7)	(8.4)	(35.3)
Fish	0.106	0.12	0.18
	(29.9)	(25.0)	(40.4)
Arhar	0.037	0.04	0.06
	(37.9)	(28.1)	(48.5)
Tea (leaf)	0.069	0.05	0.08
	(35.9)	(18.5)	(35.4)
Tea (cups)	0.179	0.17	0.26
	(57.7)	(41.8)	(71.3)

Notes: Each row comes from a regression of the logarithm of unit value on the logarithm of household per capita expenditure and on an urban dummy. Absolute *t*-values are shown in parentheses. The regressions are run at the household level, with one observation for each household recording a purchase of the good. The second column is an estimate of the extent to which urban prices are higher than rural prices, and can be compared with the unadjusted difference in means in the final column, taken from the last column of Table 1.

Table 3: Budget shares and unit values by source and sector

	RURAL				URBAN			
	Cash purchase		Other		Cash purchase		Other	
	Share	Ln uv	Share	Ln uv	Share	Ln uv	Share	Ln uv
Rice from PDS	1.03	1.54	--	--	0.48	1.55	--	--
Rice market	7.60	2.36	4.35	2.22	6.50	2.58	0.34	2.34
Milk	3.00	2.49	4.39	2.40	6.81	2.65	0.62	2.48
Wheat PDS	0.33	1.38	--	--	0.16	1.61	--	--
Wheat market	3.69	2.17	2.34	1.91	4.84	2.31	0.27	1.96
Mustard oil	2.44	4.01	0.19	3.87	1.44	4.00	0.02	3.89
Sugar PDS	0.18	2.62	--	--	0.11	2.62	--	--
Sugar market	2.06	2.92	0.25	2.91	1.73	2.91	0.00	2.92
Fish	1.12	3.69	0.21	3.59	0.96	3.87	0.04	3.49

Table 4

Income effects on unit values and adjusted urban rural differences: selected goods Cash purchases only

	Regression coefficient on lnpce	Regression coefficient on urban dummy	Unadjusted difference
Rice	0.225	0.09	0.22
	(151)	(45.2)	(110)
Milk	0.085	0.12	0.16
	(40.7)	(64.9)	(68.6)
Wheat/atta	0.155	0.05	0.14
	(71.0)	(19.1)	(54.5)
Mustard oil	0.001	–0.01	–0.01
	(0.4)	(8.3)	(8.9)
Sugar	–0.010	–0.00	–0.01
	(11.7)	(3.4)	(8.9)
Potatoes	0.088	0.05	0.09
	(43.2)	(16.7)	(38.1)
Other oil	0.070	0.01	0.06
	(52.8)	(7.6)	(34.4)
Fish	0.091	0.13	0.18
	(24.6)	(25.2)	(38.4)
Arhar	0.034	0.02	0.04
	(39.2)	(15.9)	(36.0)
Tea (leaf)	0.068	0.04	0.08
	(35.5)	(18.1)	(34.9)
Tea (cups)	0.178	0.16	0.26
	(55.5)	(40.0)	(68.5)

Notes: Each row comes from a regression of the logarithm of unit value on the logarithm of household per capita expenditure and on an urban dummy. Absolute *t*-values are shown in parentheses. The regressions are run at the household level, with one observation for each household recording a purchase of the good. The second column is an estimate of the extent to which urban prices are higher than rural prices, and can be compared with the unadjusted difference in means in the final column, taken from the last column of Table 1.

Table 5

Differences in log median unit values (urban minus rural) by decile groups

Decile	rice	milk	wheat	mustard oil	sugar	All food
1	0.06	0.18	0	–0.04	0	0.04
2	0.10	0.18	0.13	–0.02	0	0.08
3	0.11	0.17	0.13	–0.04	0	0.08
4	0.05	0	0.19	–0.04	0	0.06
5	0.02	0.08	0.25	–0.01	0	0.07
6	0.09	0.15	0.18	–0.02	0	0.08
7	0.18	0.15	0.12	0	0	0.09
8	0.18	0.15	0.22	0	0	0.10
8	0.08	0.15	0.22	0.02	0	0.09
10	0.22	0.22	0.26	0.04	0	0.15
Mean	0.11	0.15	0.17	–0.01	0	0.09

Notes: Deciles are defined using the sampling weights, so that each decile contains an equal share of the All India population. The numbers shown are the differences in median log unit values between urban and rural within each decile. The averages in the bottom row are simple averages over the deciles. The final column uses all 147 foods, with the difference in median unit values weighted by the average of the urban band rural budget shares.

Table 6

Törnqvist GEKS multilateral food price indexes by state and sector

	Without correction			With correction		
	Rural	Urban	Ratio	Rural	Urban	Ratio
Jammu & Kashmir	1.00	1.09	1.09	1.00	1.05	1.05
Himachal Pradesh	1.00	1.14	1.15	1.00	1.09	1.09
Punjab	0.94	1.08	1.12	0.94	1.04	1.09
Uttaranchal	0.92	1.02	1.13	0.94	1.04	1.09
Haryana	0.96	1.05	1.09	0.96	1.05	1.07
Delhi	..	1.19	..	..	1.13	..
Rajasthan	0.92	1.04	1.11	0.93	1.03	1.07
Uttar Pradesh	0.82	0.99	1.17	0.84	0.99	1.12
Bihar	0.82	0.96	1.12	0.86	0.98	1.07
Assam	0.96	1.18	1.16	0.99	1.15	1.09
West Bengal	0.86	1.08	1.20	0.89	1.05	1.12
Jharkhand	0.82	1.04	1.19	0.87	1.02	1.09
Orissa	0.80	0.95	1.14	0.85	0.96	1.06
Chhatisgarh	0.83	0.99	1.16	0.88	0.98	1.06
Madya Pradesh	0.82	0.98	1.18	0.86	0.98	1.12
Gujarat	1.00	1.17	1.17	1.02	1.14	1.11
Maharashtra	0.95	1.18	1.24	0.97	1.15	1.17
Andhra Pradesh	0.90	1.03	1.13	0.92	1.02	1.07
Karmataka	0.88	1.02	1.24	0.92	1.00	1.17
Kerala	0.99	1.06	1.06	0.98	1.02	1.04
Tamil Nadu	0.97	1.08	1.17	1.00	1.05	1.10

Notes: Small states and rural Delhi excluded. The first panel, without correction, shows multilateral indexes calculated from the means of the logs of reported unit values, averaged by state and sector. The second panel is based on a set of regressions in which the logarithms of the unit values for each good are regressed on the logarithm of per capita total expenditure and a fully interacted set of state and sectoral dummies. These regressions are then used to calculate hypothetical mean log unit values for each sector and state assuming, counterfactually, that each has the same mean log per capita total expenditure. This correction, in the second panel, tends to reduce both the sectoral and spatial differentials. Within each panel the column labeled ratio shows the urban to rural differential for each state from the PPPs in the first and second columns.

Table 7: Budget shares and log unit values by sector, Brazil 2002

	Urban share	Rural share	Urban premium	Adj. Premium	Cash premium	Adj. cash premium
Polished rice	1.72	0.97	3.8	0.6	1.1	–1.8
Soybean oil	1.25	0.58	–7.0	–6.4	–6.9	–6.4
French bread	0.52	1.25	5.9	0.3	5.9	0.2
Crystal sugar	1.28	0.44	–4.4	–3.1	–4.2	–2.8
Fresh milk	1.40	0.26	23.4	23.0	14.0	12.3
Cassava flour	1.13	0.22	18.4	11.8	12.6	7.0
Ground coffee	0.86	0.42	–3.1	–4.9	–4.2	–6.1
Choice beef	0.83	0.38	5.2	4.3	3.2	1.9
Carioca beans	0.81	0.33	8.2	5.0	4.1	1.1
Frozen chicken	0.61	0.51	–4.7	–5.3	–5.1	–5.9
Prime beef	0.61	0.35	9.6	7.0	5.2	1.6
Index over 38 largest share foods	15.96	9.31	7.09	4.54	4.15	1.40

Notes: All figures are percentages. The first two columns show the average shares of each item in total household expenditure. The third column shows the differences in the means of the logs of unit values, while the fourth column shows the same numbers adjusted for household per capita expenditure and household size. The final two columns replicate these calculations using unit values only for cash transactions. The bottom row of the table shows the Törnqviist indexes for 38 of the most important foods in the survey.