Spatial price comparisons in poverty measurement. An example from Cambodia.

(pre-conference version)

Jörgen Dalén, April 22. 2006[1]

Abstract: A central objective of poverty measurement is to estimate the number of persons falling below an estimated poverty line (the so called headcount). The poverty line is followed over time on the basis of fixed food and nonfood reference bundles which need to be price updated based on relevant and accurate price indexes, which have both a temporal and a spatial dimension.

Using data from a household budget surveys in Cambodia the report focuses on procedures for estimating:

- The food poverty line for other regions and time periods based on the same food reference bundle.
- The housing component of the nonfood poverty line using hedonic regression based on imputed rent and house characteristics reported by households.
- The ranking of household poverty using household-specific price indexes

.

Acknowledgement: This report describes work done for the International Consulting Office (ICO) of Statistics Sweden. It uses data from the Cambodia Socio-Economic Survey 2004 carried out by the National Institute of Statistics, Cambodia. This survey is funded by UNDP and in specific areas sponsored by The World Bank and Statistics Sweden. Sten Johansson introduced this topic to me and made a number of useful suggestions to earlier versions of the report.

[1] Statistical Consultant for Statistics Sweden, ICO. E-mail: jorgendalen@hotmail.com

1. Introduction

There are two main approaches to poverty measurement. The relative approach defines the poverty line in relation to the average standard of living, enjoyed by society. For example, in this approach one could define households with income/consumption less than X % of the median income/consumption of a society as beeing poor.

Another approach is the absolute approach. This approach tries to set a poverty line, which has the same real value (purchasing power) over time and space (unless explicitly changed) and thus enables comparisons of poverty levels between years, countries and regions. The absolute approach is the topic for this report.

The economics of setting poverty lines was developed in a number of papers in the 1980's and 1990's. Greer and Thorbecke (1986) and Ravaillon (1998) appear to provide in-depth theoretical treatments of various approaches. Kakwani (2002) summarises earlier approaches.

Index number problems in this area arise from the need to separate price and quantity effects in consumption changes over time and space. Poverty line estimation normally uses data from a Household Expenditure Survey to estimate the consumption levels of various households. In such surveys both value and quantity of various food and non-food items are collected. Price estimates could be taken from three sources: CPI data, special price collections carried out in conjunction with an HES or unit values obtained by dividing individual values and quantities of items in the survey itself.

This paper will discuss these issues based on a particular case in which I was involved – poverty estimation in Cambodia based on the Cambodian Socio-Economic Surveys (CSES) of 1993/94 and 2004. My involvement in the project took the form of comments and improvements to an earlier report by James C. Knowles (2005)[2]. The comments concerned the price updating of a 1993/94 poverty line as well as price calculations for estimating a new poverty line based on 2004 expenditure data. The discussion in this note will also make references to Knowles' report.

This report puts together notes which were provided in brief reports prepared for Statistics Sweden during late 2005.

2. Poverty line estimation

There are two key concepts in poverty line estimation.

The *food poverty line* is based on the estimated cost of a single national reference food bundle providing an average subsistence diet of (in the Cambodian case) 2,100 calories per day per capita (i.e., averaged over persons of all ages and sexes)[3]. For estimating this, calory tables showing the nutritional content of basic foodstuffs are used. The food poverty line is a value amount expressed in the national currency, which exactly suffices to buy this food reference bundle.

[2] James C. Knowles: *A New Set of Poverty Estimates for Cambodia*, 1993/94 to 2004. Draft 26.07.05.

[3] An alternative approach is to use an equivalence scale for consumption of individual household members, in which a men, women and children of various ages need different amounts of calories.

In addition to the food poverty line there is a baseline nonfood allowance, which is defined as the estimated value of nonfood consumption per capita of households whose total consumption is just equal to the food poverty line.

The *total poverty line* is the sum of the food poverty line and the nonfood allowance.

The two types of poverty line are used to calculate poverty *headcounts*, i.e., the number of households whose total consumption per capita falls below the poverty line. It immediately follows that the headcount will be higher with respect to the total proverty line than to the food poverty line.

The basic procedure is to first estimate a poverty line for a well-defined time period and geographic area within which a constant price level is assumed. We call this the *reference* period/area. The index number problem then arises when comparing poverty lines between time periods, regions in a country or between countries. The last case, between countries, will not be discussed here.

We will focus on index procedures involved in estimating:

- The food poverty line for other regions and time periods based on the same food reference bundle.
- The housing component of the nonfood poverty line. Here we use hedonic regression based on imputed rent and house characteristics reported by households.
- The ranking of household poverty using household-specific price indexes

The data at hand are from two versions of the Cambodian Socio-Economic Survey (CSES), from 1993/94 and 2004.

The 2004 survey, which is the main data source used, includes a total of 15,000 households interviewed in 900 villages during a 15-month period from November 2003 through January 2005. By comparison, only 5,578 households from 498 villages were interviewed in the 1993/94 survey.

In the 2004 survey the households reported their food and nonfood consumption during a calendar month. In addition they reported the characteristics of the house they lived in as well as its monthly rent (actual rent if the house was rented, but when owned what they thought it would rent for).

3. Estimation of the food poverty line over time and space

In the Cambodian case there are three possible sources for estimating price variation. These are the i) *Consumer Price Index*, ii) *unit values* based on reported values and quantities of food items in the CSES itself and iii) *Village prices*, which are special price quotations collected within the CSES exercise.

CPI. Knowles (2005) took the Phnom Penh CPI over the whole period 1993/94-2004 (divided into two links) as the backbone for temporal price change. This is the only reliable source for temporal price change in Cambodia but covers only the Phnom Penh area. CPI has the advantage of being an established indicator based on an internationally accepted methodology. Very important is that CPI specifications for goods and services are narrow enough to prevent quality changes from entering the price index.

But, since the CPI only covers Phnom Penh for the major part of the period, Knowles also uses CSES data from the survey rounds in 1993/94, 1997 and 2004 for estimating the spatial price index (i.e., between the three regions: Phnom Penh, other urban and rural areas). The CSES data he uses are of two types.

Unit values are prices that can be calculated from values and quantities reported by the households themselves. There appears to be a tradition in the poverty measurement literature to use unit values as price estimation tools. In the 1993/94 SESC the recall method was used, where households were asked about their consumption, with regard to the value and quantity of consumed products in a previous period. In the CSES 2004 values and quantities were instead obtained from a diary, where expenditures were noted during a calendar month. In both cases, the price of a product could be obtained by dividing the value with the quantity.

Unit values are generally not accepted as proper tools for price comparisons by economists and price statisticians. The reason is that two transactions are normally different with respect to the quality of the purchased good or service. If average qualities of a product in two regions in the same time period or in two time periods in the same region are different a bias results in the estimated price index. On the other hand, if the product is perfectly homogeneous there will be no bias and if it is nearly homogeneous, bias will be small. Another situation that would result in small bias is if there is some quality variation but the variation is similar between the two points of comparison.

In the area of food in a primitive economy like Cambodia's, there are many major products that are homogeneous or nearly so. Rice,already distinguished into three products by quality and fruits and vegetables, generically identified, are probably nearly homogeneous. Whether quality variation for some other food products (like bread or meat) is small or similar over regions and time is an important but unexplored issue.

Unit values are used by Knowles to form *median unit values* for a certain item over a whole region, i.e., simultaneously over several transactions made by a certain household in a month, all households in a village and all villages in a region. The median is used for its robustness against extreme values at both ends of the distribution caused by measurement errors. If there is some quality variation but still the same variety with respect to quality dominates the consumption in both ends of the comparison, then there will be no bias in the unit value comparison. But if the quality spread is more scattered and the market moves faster, then the median will not offer any bias protection.

A strong aspect of unit values is that they are based on real transactions and thus in principle show exactly what has been paid and for how large a transaction.

Village prices. In 1997, 1999 and 2004, CSES interviewers collected village prices for food and nonfood items (in 2004 there were 53 food prices of which Knowles used 46 to construct a spatial price index). Items in the village price list are normally specified narrowly (e.g. for peanuts "raw seed, medium size, good quality") in order to minimize quality variation within the item. This fact makes them more suitable for price comparison than unit values.

However, in many villages village prices are missing (9 % of the villages did not report any prices and the average number of prices was 53 out of a theoretical maximum of 3x51=153).

Also, the village prices suffer from not being actual transaction prices (which may be the result of negotiation) but are more like list prices.

Knowles' procedure is to estimate price change for Phnom Penh from 1993/94 to 2004 by the Phnom Penh CPI. Two sets of spatial indexes – one in 1993/94 based on unit values in recall data and one in 2004 based on village prices – are used to estimate spatial price indexes in these two years separately. Estimates for intermediate years (1997 and 1999) do not influence the comparison between 1993/94 and 2004.

By linking the temporal CPI to the two sets of spatial indexes temporal price indexes for the other two regions are obtained.

For food prices there is an option to use unit values from the CSES 2004 diary for spatial indexes instead of village prices. For the temporal benchmark (the Phnom Penh CPI) and the 1993/94 spatial index (based on recall unit values), Knowles' method cannot be improved upon.

On balance we consider that unit values be preferred over village prices, where quality variation within the product is judged to be small or else similar between the points of comparison. Where this is not the case, village prices are probably more suitable for comparison purposes.

For food, the diary unit values are preferred to village prices for the following reasons:

- They are more representative, since all transactions for the sampled household are covered.
- The actual amount paid for an actually purchased quantity is used.
- Many village prices are missing and there are none at all in many villages.
- The 74 food items in the unit value list of food items are mostly homogenous or nearly homogenous leading to small quality variation between regions. (However, some food items are excluded from this list.)
- Since unit values (albeit based on the recall method) were used in 1993/94, unit values also for 2004 result in greater methodological coherence.

The methodology for computing the new spatial index, based on diary unit values, involved the following three steps:

1. Median unit values are calculated over all cash transactions for own household consumption, for each item, in each month and region[4]. Villages not in the sampling frame in 1993/94 were excluded for the sake of comparability with the spatial index for that year.[5]

2. Aggregation over items. Only the 74 items in the 1993/94 food reference bundle were included in this stage. Aggregation followed the Fisher index. Table 2 is the result of this aggregation step.

3. Aggregation over months. Equal weights for each month were used. Table 1 is the result of this final step.

Table 1 presents the spatial indexes according to the two methods[6]. Both for other urban areas and for rural areas, we obtain a higher estimate of the price level and thereby a higher food poverty line, implying a higher poverty headcount. For other urban areas, the difference is 5.2 percent. For rural areas the difference is even larger, we find a 12.6 percent higher price level

---

[4] The same limitation to cash transactions for own household consumption was applied by Prescott and Pradhan (1997) in their report on the 1993/94 survey. Excluding non-cash transactions makes the price observations more reliable and also tends to limit the comparisons to those varieties of a product that are traded in the market, hence tends to reduce quality variation.

[5] 592 villages (out of 900) were included. In addition to the 560 villages that matched the sampling frame, all villages in the Phnom Penh area were also included.

[6] It should be noted that our estimates use finally edited data, whereas Knowles, due to time constraints had to make do with preliminary data.

and food poverty line than Knowles, which implies a much higher poverty headcount for the rural areas. The food poverty line in the Phnom Penh urban area is not affected.

Table 1: Spatial price indexes and updated food poverty lines in 2004 (Phnom Penh =100)

Region	Spatial price index based on		Updated food poverty lines based on	
	diary unit values	village prices[a]	diary unit values	village prices[a]
Phnom Penh	100.0	100.0	1,782	1,782
Other urban areas	92.6	88.0	1,650	1,568
Rural areas	87.8	78.0	1,565	1,389

[a] (Knowles 2005, tables 2 and 62)

Table 2, which is the result of second aggregation step above, provides some further information on the variation over months in the spatial indexes. This variation may be caused by several factors. There could be a real seasonal variation but there is also the distinct possibility that the variation is due to random fluctuation caused by the sampling of villages and households.

Table 2: Monthly spatial price indexes 2004 based on diary unit values (Phnom Penh =100 each month)

Month	Phnom Penh	Other urban areas	Rural areas
1	100.0	87.6	84.9
2	100.0	85.7	83.1
3	100.0	91.1	86.1
4	100.0	100.8	86.2
5	100.0	97.4	96.0
6	100.0	92.6	81.8
7	100.0	101.3	95.7
8	100.0	94.9	86.1
9	100.0	88.0	88.4
10	100.0	92.6	85.4
11	100.0	92.1	96.7
12	100.0	87.3	83.6

4. Estimating the housing component of the nonfood poverty line

For housing a large number of characteristics are collected, which makes hedonic regression, for price measurement possible. In this method the (log of) rent is used as the dependent variable in a regression and the characteristics as the right-hand side variables. The estimated coefficients determine the "price" of each characteristic. Knowles' uses hedonic estimates both for 1993/94 and 2004 to estimate spatial rent indexes.

Housing consumption is, in the CSES, estimated from information on rent. If the dwelling is rented then the actual rent paid is collected and else an estimate of what the dwelling would rent for if rented out is asked for. When estimating housing inflation it is therefore logical to focus on a deflator for precisely this type of information – actual or imputed rent. The Cambodian CPI uses imputed rent in its housing component and is therefore suitable for use as a temporal index for the Phnom Penh area[7].

Estimating price variation for housing over time and space is more difficult than for other products, since no two houses are exactly identical. A methodology is needed for separating out quality differences from price differences. Houses can be expected to be of higher quality in urban areas and especially Phnom Penh than in rural areas and can also be expected to increase in quality over time in all regions. Not taking account of quality differences would, under these circumstances create a serious bias in price index estimates.

The method used for this purposes by Knowles and also in this note is hedonic regression. Hedonic regression is appropriate for estimating price change, where data on a number of quality-relevant characteristics are available along with the prices (rents in this case). This is the case in the Cambodian CSES (SESC), where data are collected on roof material, wall material, floor material, light, fuel, and toilet and water facilities. These characteristics are obviously of great relevance for the user perception of the quality of the house. Characteristics were collected in all survey rounds, i.e., in 1993/94, 1997 and 2004. Definitions used are sufficiently similar for use in temporal comparisons[8], although temporal comparisons are not the focus of this report.

There are several variants of the hedonic method, which can give different results. The variant used by Knowles is usually called the "*time dummy method*" (or perhaps "*space dummy method*" in the case of spatial comparisons). In this method the price index is estimated directly through the coefficient for time/space in the regression itself. Knowles calculates several hedonic price indexes for housing, spatial as well as temporal, using rental values (imputed, in some cases actual rents) as dependent variables. However, in the end only the spatial indexes for 1993/94 and 2004, presented in his Table 68, are used. Only the 2004 spatial index is used for updating the nonfood allowance.

Another method is the *hedonic quality adjustment method*[9]. In this method, the coefficients of characteristics are explicitly used for quality adjustment along with the changes in the rates of the characteristics between the points of comparison. We have three reasons to prefer this variant of the hedonic method, especially for the present purposes.

Firstly, the dummy method forces the characteristics coefficients for the two points of comparison to become equal, which in reality they are not. This may cause differences in these coefficients to enter into the coefficient for time and thus create a bias in the estimate of price change. In the context of comparisons over time, this may be a smaller problem, since usually time periods of equal length are compared on the basis of roughly equal number of observations. But in this case, where spatial comparisons between areas of widely different size are concerned, the problem could be much greater. In 2004 some 80 % of the sampled households are from the rural area, which creates a major imbalance when estimating the coefficients and the time dummy.

[7] The housing index in the CPI follows prices of a subsample of houses from the SESC 1993/94 sample of households. Households are asked how much the house would rent for in the present time period

[8] A few new characteristics were introduced in 2004. With few exceptions they are easily related to the 1993/94 definitions. Unclear at this moment are a) whether "asbestos" (used in 1993/94) is equivalent to "fibrous cement" (used in 2004) b) whether concrete roof material was included in the "other" category in 1993/94. These problems do not affect the spatial comparisons presented in this note.

[9] Triplett, Jack E., (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products,* Brookings Institution. The present application to groupwise comparisons is not, however, explicitly treated in Triplett's handbook.

Secondly, the quality adjustment method, as we will see, allows a decomposition of the quality adjustment into factors related to the changes in all the characteristics separately. This increases the transparency of the hedonic method for interpretation purposes.

Thirdly, the two points of comparisons (regions in this case) are treated symmetrically and given equal weight by our method.

The index formula applied is

$$\boldsymbol{I} = \frac{\overline{\boldsymbol{P}}_b}{\overline{\boldsymbol{P}}_a} \Big/ \mathbf{exp}\left[\sum \overline{\beta}_j \left(\overline{\boldsymbol{x}}_j^b - \overline{\boldsymbol{x}}_j^a\right)\right], \qquad (1)$$

where the logged price (rent) is used as the dependent variable in the regression, $\overline{\boldsymbol{x}}_j^b$ and $\overline{\boldsymbol{x}}_j^a$ are averages of characteristic j in region b and a, respectively, $\overline{\boldsymbol{\beta}}_j$ is the average regression coefficient for characteristic j over separate regressions made for regions a and b, and $\overline{\boldsymbol{P}}_b$ and $\overline{\boldsymbol{P}}_a$ are the geometric means of rents in region b and a, respectively. Note that the points of comparison (a and b) are treated symmetrically in this formula.

An interpretational advantage of (1) is that one can decompose the estimated price change into two components:

$$\text{Raw price index} = \frac{\overline{\boldsymbol{P}}_b}{\overline{\boldsymbol{P}}_a}, \text{ and} \qquad (2)$$

$$\text{Quality index} = \mathbf{exp}\left[\sum \overline{\beta}_j \left(\overline{\boldsymbol{x}}_j^b - \overline{\boldsymbol{x}}_j^a\right)\right] \qquad (3)$$

We thus have the relation: Quality adjusted price index = Raw price index / Quality index.

A further aspect of our method is that we use a chained procedure. The estimated price indexes (each according to formula 1) are based on chained pairwise comparisons in which Phnom Penh is compared to other urban areas and other urban areas to rural areas. Each pairwise comparison is based on the maximum set of characteristics in common to the two areas. In this way it is possible to retain a maximum number of observations in each of the pairwise comparisons. Chaining indexes through intermediate situations is the best practice, when two points of comparison (e.g., two countries, two years) are initially far apart[10].

For example, some characteristics for houses in Cambodia do not exist in the Phnom Penh urban area. Examples in 2004 are salvaged materials for roofs and walls or dug wells for water. There is thus no basis for including them in a comparison between Phnom Penh prices and prices in other areas. On the other hand they do exist in both other urban areas and in rural areas, although to a smaller extent in the former, and can and should be included in a price comparison between those two areas. Since they are not so common in other urban areas, their exclusion from the comparison between Phnom Penh and other urban areas is not a big concern.

[10] One can compare this practice with the ICP comparisons between two distant countries (e.g. Sweden and Portugal), where a third country (e.g. France) is taken as an intermediate point for pairwise comparisons with both countries.

Spatial housing comparisons 2004

In Table 3 we have used the hedonic quality adjustment method, as described above for spatial comparisons between the three regions based on CSES 2004. We obtain much higher relative rent levels in both other urban and in rural areas compared to Knowles.

It is not surprising that two different hedonic methods could give so different results. In our judgment they are the result of the chained and symmetric treatment in our calculation formula of the three regions, compared with a comparison where the rural observations are allowed to dominate the estimates of the regression coefficients.

Table 3: Spatial price indexes for housing, 2004

Region	Estimated price index	Price index in Knowles (2005)
Phnom Penh	100.0	100.0
Other urban areas	75.8	64.7
Rural areas	71.4	50.8

Tables 5-8 provide details on the calculations. Tables 5-6 give results of the comparison between Phnom Penh and Other urban areas and Tables 7-8 give the results of the comparison between other urban areas and the rural areas. The key regression characteristics are in Table 4[11].

Table 4: Regression diagnostics

	Phnom Penh to other urban areas		Other urban to rural areas	
	Other urban areas	Phnom Penh	Other urban areas	Rural areas
R^2	*62.8*	*43.1*	*65.1*	*46.3*
# observations	815	591	1592	8838
# variables	33	33	45	45

For the comparison between Phnom Penh and other urban areas we estimate a raw price index of 395.5 and a quality index of 299.8 according to Table 5. This means that the unadjusted prices are almost four times higher in Phnom Penh compared to other urban areas. At the same time quality is three times higher. As a result the quality adjusted price index can be estimated to 131.9, i.e., prices for identical houses are 31.9 % higher in Phnom Penh than in other urban areas. Furthermore we are also able to discern the factors which contribute most to the quality index. These are, from top to down, toilet facilities (contributing 24 % - exp 0.216 – to the quality index), light (20 %), wall material (16 %), floor material (16 %), size (m^2, 14 %), fuel (12 %), roof material (9 %) and source of water (8 %).

Table 6 provides more detail as to the means and coefficients of all the characteristics in the two regions. For example we can see that the single most important factor contributing to the high quality index (the highest value in the right-most column) is that 80 % of the Phnom Penh houses have concrete walls, whereas only 19 % of those in the other urban areas have that characteristic. This factor alone contributes to a 54 % (exp 0.43) higher quality index for Phnom Penh.

[11] Full information on regression diagnostics can be obtained from the author on request.

Table 5: Price index and aggregate quality adjustments (log scale) per category, Phnom Penh to other urban areas according to formula (1)

Category	Aggregate quality adjustment, logs
Floor	0.146
Fuel	0.111
Light	0.179
Month	-0.002
Roof	0.090
size	0.127
toilet	0.216
Wall	0.151
Water	0.080
Sum of logs	1.098
Quality index	299.8
Raw price index	395.5
Quality adjusted price index	131.9

Table 6: Means and coefficients of characteristics in the price index comparing Phnom Penh to other urban areas

Characteristic	Coefficient Other urban Areas	Coefficient Phnom Penh	Mean, Other Urban areas	Mean, Phnom Penh	Effective adjustment
logtime	1.08	-0.09	2.67	2.67	0.00
logfloorarea	0.64	0.42	3.84	4.02	0.10
logroomnum	0.24	0.44	0.45	0.54	0.03
floor_cement	0.43	-0.17	0.10	0.11	0.00
floor_ceramictiles	0.51	-0.05	0.17	0.70	0.12
floor_clay	0.51	-0.63	0.04	0.01	0.00
floor_parquet	0.56	0.05	0.10	0.11	0.00
floor_polstone	1.00	1.39	0.00	0.00	0.00
floor_wood	0.31	-0.38	0.58	0.04	0.02
fuel_charcoal	0.04	-0.52	0.28	0.24	0.01
fuel_firewood	-0.39	-0.27	0.47	0.04	0.14
fuel_firewoodcharcoal	-0.30	-0.54	0.05	0.06	0.00
fuel_gas	0.24	-0.36	0.14	0.57	-0.03
fuel_gaselectr	-0.16	-0.32	0.02	0.08	-0.01
fuel_publicelectric	0.38	-0.19	0.00	0.01	0.00
light_private	0.20	0.76	0.17	0.08	-0.04
light_public	0.47	0.85	0.58	0.92	0.22
roof_concrete	0.24	-0.05	0.06	0.50	0.04
roof_fibrouscement	0.13	0.11	0.08	0.10	0.00
roof_gia	-0.15	-0.09	0.53	0.33	0.02
roof_mixed6	-0.43	-0.36	0.03	0.00	0.01
roof_thatched	-0.44	-0.32	0.11	0.01	0.04
roof_tiles	0.15	0.21	0.20	0.06	-0.02
toilet_connectedtosewerage	0.41	0.59	0.08	0.90	0.41
toilet_pitlatrine	-0.15	-0.09	0.03	0.00	0.00
toilet_septictank	0.13	0.64	0.58	0.08	-0.19
wall_concrete	0.37	1.05	0.19	0.80	0.43
wall_gia	0.05	1.56	0.03	0.01	-0.02
wall_plywood	-0.04	0.61	0.21	0.12	-0.02
wall_woodlogs	0.15	1.12	0.44	0.06	-0.24
water_bought	-0.12	-0.13	0.18	0.02	0.02
water_pipedindwelling	0.19	0.12	0.41	0.96	0.08
water_tubed	0.30	-0.12	0.28	0.01	-0.02

For the comparison between other urban areas and rural areas in Table 7 we estimate a raw price index of 215.3 and a quality index of 202.8, resulting in a quality adjusted price index of 106.2, i.e., prices for identical houses are 6.2 % higher in other urban areas than in rural areas. In this case the most important class of characteristics is light, contributing 25 % to the quality index, followed by toilet facilities (13 %), size (12 %) and fuel (11 %). We also see, in the rightmost column of Table 8, that the single most important characteristic is public lighting, contributing 23 % (exp 0.21) to the quality index.

Table 7: Price index and aggregate quality adjustments (log scale) per category, other urban to rural areas according to formula (1)

Category	Aggregate quality adjustment, logs
floor	0.022
fuel	0.101
light	0.223
month	0.001
roof	0.022
size	0.110
toilet	0.118
wall	0.072
water	0.039
Sum of logs	0.707
Quality index	202.8
Raw price index	215.3
Quality adjusted price index	106.2

Table 8: Means and coefficients of characteristics in the price index comparing other urban areas to rural areas

Characteristic	Coeffi-cient, Other urban	Coeffi-cient, rural	Mean, Other urban	Mean, rural	Effec-tive Adjust-ment
Logtime	0.36	1.53	2.67	2.67	0.00
Logfloorarea	0.57	0.44	3.65	3.50	0.08
Logroomnum	0.25	0.13	0.33	0.16	0.03
floor_cement	0.15	0.35	0.08	0.03	0.01
floor_ceramictiles	0.34	0.55	0.10	0.02	0.04
floor_clay	0.22	0.14	0.07	0.08	0.00
floor_parquet	0.28	0.38	0.11	0.08	0.01
floor_polstone	0.76	1.23	0.00	0.00	0.00
floor_wood	0.18	0.29	0.63	0.78	-0.04
fuel_charcoal	0.11	0.56	0.19	0.03	0.05
fuel_firewood	-0.26	0.16	0.65	0.92	0.01
fuel_firewoodcharcoal	-0.09	0.24	0.05	0.01	0.00
fuel_gas	0.27	0.62	0.08	0.02	0.02
fuel_gaselectr	0.03	0.99	0.01	0.00	0.01
fuel_privateelectric	0.49	0.53	0.00	0.00	0.00
fuel_publicelectric	0.74	0.58	0.00	0.00	0.00
light_battery	-0.07	0.14	0.12	0.28	-0.01
light_private	0.39	0.43	0.12	0.06	0.02
light_public	0.62	0.66	0.37	0.05	0.21
roof_concrete	0.38	0.34	0.03	0.01	0.01
roof_fibrouscement	0.50	0.40	0.07	0.04	0.01
roof_gia	0.16	0.22	0.51	0.32	0.04
roof_mixed6	-0.03	0.25	0.02	0.01	0.00
roof_mixed7	0.32	-0.04	0.00	0.01	0.00
roof_salvaged	0.12	-0.54	0.00	0.00	0.00
roof_thatched	-0.06	-0.16	0.18	0.23	0.01
roof_tiles	0.33	0.32	0.18	0.30	-0.04
toilet_connectedtosewerage	0.46	0.78	0.05	0.01	0.02
toilet_openland	-0.16	-0.19	0.23	0.47	0.04
toilet_otherwithoutseptictank	0.22	0.01	0.02	0.02	0.00
toilet_pitlatrine	0.01	0.16	0.03	0.02	0.00
toilet_septictank	0.18	0.24	0.38	0.12	0.05
wall_concrete	0.50	0.32	0.12	0.03	0.03
wall_fibrouscement	0.93	0.35	0.00	0.00	0.00
wall_gia	0.01	0.21	0.04	0.02	0.00
wall_other	-0.11	-0.11	0.08	0.18	0.01
wall_plywood	0.03	0.09	0.19	0.18	0.00
wall_salvaged	-0.19	-0.24	0.01	0.01	0.00
wall_woodlogs	0.16	0.23	0.37	0.27	0.02
water_bought	-0.10	0.01	0.13	0.07	0.00
water_pipedindwelling	0.08	0.30	0.22	0.04	0.04
water_protecteddugwell	-0.05	-0.17	0.16	0.19	0.00
water_publictap	0.49	-0.24	0.00	0.00	0.00
water_tubed	0.16	-0.08	0.25	0.29	0.00
water_unprotecteddugwell	0.05	-0.18	0.06	0.13	0.00

Spatial housing comparisons 1993/94

Parallel computations leading to spatial hedonic estimates for housing for 1993/94 have also been carried out. The aggregate results are in Table 9. Differences are only minor.

Table 9: Spatial price indexes for housing and all nonfood, 1993/94

Region	Estimated housing price index	Housing price index in Knowles (2005)	Estimated total nonfood price index	Nonfood price index in Knowles (2005)	Nonfood price index in Prescott and Pradhan (1997)
Phnom Penh	100.0	100.0	100.0	100.0	100.0
Other urban	72.8	69.0	90.7	90.4	68.4
Rural areas	54.6	51.7	78.2	78.0	60.1

However, these estimates do not determine the baseline nonfood allowances, which are instead based on a regression model applied by Prescott and Pradhan (1997). We would like to call attention to the fact that their spatial index is very different from ours and Knowles'. They estimate the prices in other urban and rural areas to be much lower, which in turn leads to lower nonfood allowances and lower headcounts with respect to the total poverty line. In Table 10 we provide the alternative nonfood allowances that would follow from Knowles and my price index estimates. They are remarkably higher than Prescott and Pradhan's.

This issue is therefore of great importance in determining the true picture of poverty development in Cambodia during the period 1993/94-2004. The alternative estimates given in Table 10 imply much higher nonfood poverty lines in 1993/94 and thus also a larger poverty headcount in that period.

Table 10: Nonfood allowances 1993/94 according to different methods

Region	Prescott and Pradhan (1997)	Implicit in Knowles	Implicit in Dalén
Phnom Penh	393	393	393
Other urban areas	269	355	356
Rural areas	236	307	307

Price indexes at the household level

When establishing the 2004 poverty baseline estimates for Cambodia, the households are ranked according to their actual consumption, as measured by the Cambodian Socio-Economic Survey (CSES). In this survey consumption was measured in both value and quantity terms, during each month from November 2003 to January 2005 in monthly representative samples of villages. The ranking of the households, from poorest to richest, must be made in real (price-adjusted) terms to be meaningful for poverty measurement.

In his work plan for the Cambodian Poverty Report[12] Dr. Knowles proposed to use household level price indexes for this purpose. In these indexes the price level of each household is to be compared to the average price level in Phnom Penh (PP) in 2004. In this way a comparison of the level of real consumption can be over all households in Cambodia and a ranking of their level of real consumption can be made. This is possible only because households have reported all their expenditures for daily recording by the interviewers.

[12] Jim Knowles: Workplan for the Preparation of the 2004 Poverty Estimates. Updated May 28, 2005.

Here we describe how these indexes were obtained, following Knowles' proposal. I have computed two different price indexes: one for food and one for housing. (For other nonfood than housing Knowles' estimates based on village prices was used.)

Household price indexes for food

The computation of household price indexes for food involved the following steps:

- A file with all diary transactions in the 2004 Cambodian CSES, and in the calendar year of 2004, was used. The file was edited for mistakes and the quantity data were based on standardised units. Only transactions coded to be for own household consumption or other consumption purposes were used and transactions coded for production were thus excluded. Own production used for consumption as well as all other forms of acquisition were included. The inclusion of own production is crucial, since its price level is lower than for market production and consumption values from own production thus represent larger consumed quantities.
- For each item all transactions made by a single household were accumulated. This means that values and quantities were summed for each item code. Prices (unit values) were calculated by dividing the monthly household consumption in value terms by the consumption in quantity terms.
- Only item codes which are among the 75 with highest aggregate consumption values and are at the same time deemed to be homogeneous enough for price comparisons are included. Table 5 below lists included and excluded items among the 75 largest. The included food item covered 63 % of all food consumption.
- In Phnom Penh a unit value for each item code was calculated over all households, weighted by the sample household weights, multiplied by number of persons in the household.
- For each consumption item found in the monthly diary of each sample household (including those in PP), a price index was calculated by dividing the PP unit value with the household unit value.
- The household price index was finally computed as a weighted average of price indexes of all consumption items found in the monthly diary of the household. The weights were the actual diary consumption values of that specific household.[13]

The final price index is to be interpreted as the amount that the consumption of each household is to be multiplied by in order to reach the price level of an average PP household. By doing this

---

[13] Here we departed from Knowles recommendation to use weights based on aggregate consumption values. We find aggregate values to be inappropriate since the aim is to price adjust the *actual* consumption of each household to a common basis. Formally, what is desired is a ranking of a household j according to its consumption in PP prices, which can be written as $\sum_k x_k^j p_k^{PP} = \sum_k x_k^j p_k^j * \sum_k \frac{x_k^j p_k^j}{\sum_k x_k^j p_k^j} \frac{p_k^{PP}}{p_k^j}$. The resulting formula demonstrates that the nominal consumption of a household should be multiplied with a weighted index with the weights of household j. (x_k^j denotes the quantity of item k consumed by household j and p_k^j the price of item k paid by household j. p_k^{PP} is the Phnom Penh price of item k.

the real consumption of all households in Cambodia is obtained and can be compared to all other households for poverty comparison purposes.

In order to illustrate the household indexes for food, we provide some basic statistics on the calculations. We see that on average[14] food consumption values were increased by 3 % in Phnom Penh, 25 % in Other urban areas and with 125 % in Rural areas in order to make them comparable in quantity terms. In Phnom Penh the distribution of the indexes was concentrated with a small standard deviation whereas, partly due to some outliers, the standard deviations were much larger in Other urban and in Rural areas.

Table 11: Food indexes, basic statistics

	Phnom Penh urban	Other urban	Rural
Geometric mean	1.032	1.246	2.252
Standard deviation	0.15	2.22	2.85
Maximum	2.68	93.24	268.82
Minimum	0.57	0.67	0.80

Household price indexes for housing

For housing the household price index is based on a regression model similar to the one used by Knowles (2005)[15] and above. A difference in this case is that the models are applied to single dwellings rather than to all dwellings in a region. For this reason the index number procedures above are not relevant here.

The following regression model was used:

$$\log P_j = \alpha + \sum_k \beta_k x_{jk} + \sum_l \gamma_l \log y_{jl} + \varepsilon_j \qquad (1)$$

P_j, the dependent variable is the rent value of the dwelling. Two kinds of explanatory variables were used:

- Quantitative variables, y_{jl}, reflecting the size of the dwelling. Logged versions of two such variables were used: number of rooms (logroomnum) and floor area (logfloorarea). The month of the year (1-12) was also used as a quantitative variable (not logged).
- Dummy variables, x_{jk}, describing properties of the house construction and facilities in the house. Variable names are basically self-explanatory. These variables reflect responses to questions regarding
  - The floor of the house (6 variables, clay floor was set as base category)
  - The roof of the house (6 variables, simple roofs, either thatched, salvaged, mixed or with plastic sheets were set as base category)

---

[14] The geometric mean is more appropriate than the arithmetic mean to use for averaging when changes or index numbers are concerned.

[15] James C. Knowles: A New Set of Poverty Estimates for Cambodia, 1993/94 to 2004. Draft, July 26, 2005

- The walls of the house (4 variables, simple walls, either thatched, by bamboo, salvaged or made by other material were set as base category)
- Water availability in and arround the house (4 variables, no availability to clean water was set as the base category)
- Toilet facilities in and around the house (4 variables, no own toilet was set as the basic category)
- Fuel use (7 variables, simple fuel solutions such as by kerosene or private electricity was set as base category)
- Lighting in the house (2 variables, simple light solutions such as by kerosene, battery or others was set as base category)

The coefficients obtained in the regressions were used to estimate a "model rent", which is what the dwelling should rent for according to the regression model. This model rent could be interpreted as some kind of objective estimate of what the rented dwelling is worth on the market. It could be written as

$$\hat{P}_j = \exp(\hat{\alpha} + \sum_k \hat{\beta}_k x_{jk} + \sum_l \hat{\gamma}_l \log y_{jl}) \ , \qquad (2)$$

where the hat over the coefficients now signifies that they are estimated in the regression.

Since three regressions were run, one for each of the regions, we can now calculate three price levels for a given house defined by a given set of characteristics. In practice only the Phnom Penh price level was used so that a price index for the dwelling of household j was calculated as

$$Rentindex_j = \frac{\hat{P}_j^{PP}}{P_j} \qquad (3)$$

The rent index thus shows by what factor the rent value stated by the household itself would have to be multiplied in order to obtain the average price level of a house with the same characteristics in the Phnom Penh urban area rented in the same month.

Some statistics on the rent indexes are given in Table 12. The average rent indexes are close to one in Phnom Penh, as expected, but above 2 in the other areas. This could be interpreted in two ways: Either the households in other areas underestimate the rent value of their houses or, perhaps more likely, the same house in Phnom Penh has a higher rent value simply due to its more central location.

Table 12: Rent indexes, basic statistics

Geometric mean	0.979	2.119	2.218
Standard deviation	2.25	5.09	6.33
Maximum	35.80	119.69	237.93
Minimum	0.04	0.02	0.005

In Table 13 we compare the levels and distributions of rent, before and after rent adjustment according to formula (3) above. We see that the effect of the adjustment is to narrow down the differences in rent between the three regions, and also between households within regions as can be seen from the smaller standard deviations for the adjusted rents. Households who stated very high rental values for their dwellings tend to have rent indexes smaller than one and vice versa.

Table 13: Rents and adjusted rents

	Phnom Penh urban areas		Other urban area		Rural areas	
	Nominal rent	Rent in average Phnom Penh price level	Nominal rent	Rent in average Phnom Penh price level	Nominal rent	Rent in average Phnom Penh price level
Average	500420	332697	116479	173592	41392	73522
Coefficient of variation	1.54	0.67	2.90	1.46	3.60	1.47

Table 14 provides the regression coefficients for the estimated regressions in the three regions. As expected, the size variables (floor area and number of rooms) are strongly significant in all regions. Various characteristics relating to the construction of the house (roof, floor, wall) and utilities (water, light, toilet) are very strongly significant in rural areas but less so in urban areas. Fuel use is not significant in any area.

Table 14: Regression coefficients in household index regressions (t-values in brackets)

Category of variables	Variable	Phnom Penh urban	Other urban	Rural
	Intercept	6.964 (5.6)	8.149 (13.4)	7.758 (31.2)
Quantitative variables	Month	0.001 (0.1)	0.005 (0.7)	0.012 (4.5)
	Logfloorarea	0.412 (7.5)	0.574 (12.5)	0.457 (22.6)
	Logroomnum	0.441 (6.1)	0.264 (4.4)	0.138 (4.5)
Dummy variables, roof	roof_tiles	0.535 (1.2)	0.323 (3.8)	0.477 (15.7)
	roof_fibrouscement	0.421 (1.0)	0.503 (4.7)	0.544 (11.0)
	roof_gia	0.243 (0.6)	0.181 (2.7)	0.379 (14.4)
	roof_mixed6	-0.203 (-0.3)	0.006 (0.0)	0.414 (3.8)
	roof_concrete	0.273 (0.6)	0.396 (2.7)	0.455 (4.6)
	roof_other	0.313 (0.5)	-0.074 (-0.4)	0.065 (1.8)
Dummy variables, wall	wall_woodlogs	1.124 (2.8)	0.208 (3.3)	0.283 (11.4)
	wall_plywood	0.601 (1.6)	0.075 (1.1)	0.137 (4.7)
	wall_concrete	1.043 (2.8)	0.554 (5.0)	0.375 (5.1)
	wall_gia	1.533 (3.0)	0.059 (0.5)	0.243 (3.7)
Dummy variables, floor	floor_wood	0.384 (0.9)	-0.019 (-0.2)	0.162 (4.8)
	floor_cement	0.596 (1.5)	-0.049 (-0.4)	0.236 (3.7)
	floor_parquet	0.815 (2.0)	0.069 (0.7)	0.222 (4.7)
	floor_polishedstone	2.161 (3.1)	0.491 (1.0)	1.041 (3.0)
	floor_ceramictiles	0.720 (1.8)	0.124 (1.0)	0.410 (4.1)
	floor_other	0.683 (1.5)	-0.204 (-1.1)	-0.136 (-1.6)
Dummy variables, water	water_pipedindwelling	0.133 (0.4)	0.089 (1.1)	0.324 (5.5)
	water_public	0.239 (0.4)	0.189 (2.8)	-0.100 (-4.2)
	water_dugwell	0.785 (1.0)	-0.013 (-0.2)	-0.183 (-7.8)
	water_bought	-0.091 (-0.2)	-0.082 (-1.0)	0.043 (1.1)
Dummy variables, toilet	toilet_connectedtosewerage	0.250 (0.5)	0.505 (4.3)	0.917 (10.1)
	toilet_septictank	0.315 (0.6)	0.238 (3.7)	0.357 (10.7)
	toilet_pitlatrine	-0.417 (-0.4)	0.055 (0.4)	0.288 (4.9)
	toilet_publicshared	-0.421 (-0.7)	0.202 (1.6)	0.294 (2.9)
Dummy variables, light	light_public	0.873 (2.2)	0.654 (9.7)	0.558 (10.2)
	light_private	0.806 (1.9)	0.412 (5.0)	0.364 (8.3)
Dummy variables, fuel	fuel_firewood	0.514 (0.6)	-0.775 (-1.3)	-0.323 (-1.4)
	fuel_charcoal	0.292 (0.4)	-0.416 (-0.7)	0.085 (0.4)
	fuel_firewoodcharcoal	0.284 (0.4)	-0.619 (-1.0)	-0.232 (-0.9)
	fuel_gas	0.463 (0.6)	-0.250 (-0.4)	0.156 (0.6)
	fuel_publicelectric	0.624 (0.7)	0.202 (0.3)	0.112 (0.3)
	fuel_gaselectr	0.492 (0.6)	-0.496 (-0.8)	0.502 (1.6)
	fuel_other	0.798 (0.9)	-0.514 (-0.9)	-0.496 (-2.0)
	R2	**0.431**	**0.648**	**0.456**