Prices, local measurement units and subsistence consumption in rural surveys: an econometric approach with an application to Ethiopia

Bart Capéau and Stefan Dercon

WPS/98-10

March 1998

Centre for the Study of African Economies
Institute of Economics and Statistics
St Cross Building
Manor Road
Oxford OX1 3UL

Bart Capéau is at the Center for Economic Studies, Katholieke Universiteit Leuven, Belgium.

Stefan Dercon is at the Centre for the Study of African Economies, Oxford University and at the Center for Economic Studies, Katholieke Universiteit Leuven, Belgium.

Keywords: household surveys, unit values, subsistence consumption, local measurement units
JEL classification codes: D4, I3, R2

e-mail: stefan.dercon@economics.ox.ac.uk

Acknowledgements: Etalem Tjirongo and Pieter Serneels provided excellent research assistance. Useful comments were received from Geert Dhaene, Erik Schokkaert and Pramila Krishnan. All errors are our own.

Abstract: For many research problems in developing countries, some information on prices faced by households is required for the analysis, but these prices are not readily available from household surveys, nor is it straightforward to observe them, especially if subsistence consumption is a substantial part of consumption. Furthermore, quantities consumed and produced are often in local units presenting further problems for the analysis. Building on Deaton's (1987) seminal work, we provide an econometric approach to estimate prices and quantity conversion factors from household expenditure data.. We use panel data from rural Ethiopia to illustrate the approach and to investigate the potential quality bias in the estimation of the prices. In an application we show that the conclusions about poverty changes over time are significantly affected by using less appropriate strategies to convert local units and to value subsistence consumption.

1. Introduction

In recent years there has been a rapid increase in the number of large data sets collected in developing countries. Large data collection programs, such as the Living Standards Measurement Study (LSMS) and similar surveys, have yielded data for and an impetus to microeconometric research on rural and urban households, using state of the art techniques (Deaton (1997)). For many research problems, some information on prices faced by households is required for the analysis, but these prices are not readily available from the surveys, nor is it straightforward to observe them. For example, in many of the countries involved, not least in Africa, most of the rural households still derive a substantial part of their income and consumption from own production ('subsistence consumption' or 'autoconsommation'). The problem of the appropriate price to value subsistence consumption is rarely resolved in a satisfactory way. This involves methodological problems such as the endogeneity of the price in household decision making with imperfect markets, the role of quality or whether the consumer or the producer price is most appropriate to value own production (Singh et al., (1986), Low, (1986), Deaton, (1987)). It also includes related measurement problems such as deficiencies in price surveys and measurement error in the household data (Deaton, (1997)). In this paper, we propose a method to obtain market prices from household data that explicitly deals with the measurement problems inherent to the type of data used. We will estimate prices from household consumption panel data, collected in Ethiopia in 1994-95. To show the possible strength of the approach, we will use the information obtained to value subsistence consumption to impute a total consumption measure for the households involved. When doing this, we will abstract from some of the methodological problems and assume that the cluster-level consumption price is the appropriate price to value subsistence consumption. We will also use the prices obtained to calculate a cost-of-basic-needs food poverty line (Ravallion and Bidani (1994)). This poverty line will then be used to assess changes in food poverty over the survey period, in comparison with the results which use other standard methods of valuation. We find that the methods used affect the results significantly.

Prices are not the only problem in the calculation of values for consumption and income. In most developing countries standardisation of measures and units has not yet been consolidated. For most large transactions in wholesale markets, metric or imperial units are increasingly used. Nevertheless, in small markets and especially for commodities not traded by a household, a large numbers of local, traditional units are used. Our econometric method allows for a joint determination of prices and conversion factors from household data.

Few researchers concern themselves with any of these problems. Indeed, it has proved very hard to find much trace of the procedures and practices used in published empirical work. Part of the problem is that by the time researchers obtain the data, conversions and valuations have been completed. Furthermore, the time needed to check conversions and valuations is usually too large given the time constraints of the researcher. Nevertheless, for any researcher by choice or by command involved in working from raw data with the purpose of obtaining income and consumption estimates for analytical purposes, the number of apparently arbitrary choices to be made during the valuation exercise is at least a common source of frustration and doubt.

There are a few obvious solutions to these problems (for a discussion, see Levin (1991)).

Even in some of the poorest countries, most commodities consumed are traded on the local markets, even if most households would consume it from their own stocks or fields. During the data collection phase, prices could then be gathered from local markets and used in the valuations of consumption and income (e.g. see Van de Walle (1988) for Indonesia; Glewwe (1990)). Similarly, in most localities, units of measurement can be assumed to be similar across households, even if there may be differences between areas. A survey at the community or market level about these local units of measurement could then result in the appropriate conversion factors[1] . However, in many of the large scale data collection exercises, these data appear not to be collected or considered of poor quality. Logistical problems such as the very large number of clusters or the lack of a market on exactly the day the community was visited contribute to these problems[2] . Moreover, such community level survey contains mostly only one observation, while there are many observations implicitly available from the household survey itself.

The basic question is how to extract this information from a household survey in a satisfactory way. Usually, the consumption data available are expenditures on purchased commodities and quantities on both purchased and subsistence consumption. Depending on the questionnaire design, quantities may already be recorded in metric or imperial units, putting the burden of conversion on the enumerator or the household, which may present problems if these are not units usually used by these agents[3] . A similar approach for the valuation of subsistence consumption appears to be increasingly used in which households are asked to value *themselves* their subsistence consumption or their own harvest. Again, the measurement problem involved may be large[4] . Even if these shortcuts have been used, for many purposes the problems are not necessarily solved. For example, researchers may want to investigate calorie-income relationships or other problems related to calorie-intake. Prices and/or quantities in standardised units may then still be required.

The most common approach if monetary expenditures (or incomes) and quantities are available is to derive unit values for the commodities, by dividing expenditures (or incomes) by

[1] In the 1988 SUSENAS survey, this approach was used. Enumerators were carrying containers and were asked to obtain conversion factors per village. Although the data appear generally of high quality, this left even within the SUSENAS survey, for a substantial number of problem cases in the data set which needed to be addressed during cleaning.

[2] For example, Deaton, (1997, p.37), discusses the problems experienced in the LSMS to collect price data at the community level. These problems make the case for an econometric approach stronger.

[3] An appropriate procedure would be to measure during the interviews the commodities involved, avoiding the need for a conversion. The quality of the underlying data is then likely to be very high. Usually, this is too time-consuming and costly, and rarely this approach seems to be taken for some of the large scale multi-purpose surveys. Enumerators or households then have to use their judgement, which in general can hardly be considered an appropriate strategy. Measurement error problems can then not be addressed within the data.

[4] For example, some of the World Bank Priority Surveys use this approach (e.g. the Burkina Faso 1994 survey). Another example is the HRDS survey in Tanzania from 1993.

the quantities[5]. Using additional assumptions, such as a constant price per cluster, the mean or the median unit value can then be obtained as an estimate of the local price. This estimate can be used for those households in the cluster for whom only non-traded quantities are available. Similarly, conversion factors are obtained by identifying households active on the market using different measurements units. Mean (or median) prices per unit are compared across the different units to obtain the relative quantity conversion factor.

In his seminal work, Deaton (1987,1988,1990, 1997) has criticised the previous approaches to obtain prices, since prices thus obtained hide the heterogeneous quality of the commodities involved (see also Singh et al., 1986). Since quality is endogenous in demand relationships - quality is a choice - estimations of for example price elasticities in demand equations are biased. Measurement error in the unit values bias the results further. He derives an approach in which price elasticities can be derived without price data, taking into account measurement error as well as the price-unit value difference. Versions of the model are estimated on Côte d'Ivoire and Indonesia data. A careful look at his empirical results learns however that the measurement error problem dominates the bias caused by quality choice, especially in rural areas[6].

In this paper, we build on his approach in a less ambitious project: how to obtain more reliable prices and conversion factors when constructing income or consumption values in Ethiopia. Indeed, in much applied welfare analysis on developing countries, such information remains crucial, even if within the survey values for subsistence consumption were obtained[7]. For example, to derive absolute poverty lines price information remains necessary if corrections for different cost of living over space or time are to be implemented for poverty comparisons. Similarly, yield information remains a critical value in much agricultural production analysis.
We use a simple regression equation in which quantity conversions and prices per cluster are simultaneously estimated in a cross-section, assuming a specific functional pattern for the disturbance terms[8]. In fact, these disturbance terms are a hybrid of the inherent stochastic nature of the data and measurement errors. Since this approach cannot correct for heterogeneity in quality of the commodities, we extend the analysis to account for random or fixed quality differences across households. Using a panel data set we obtain estimates of prices and conversion factors, corrected for quality and measurement error. In this paper we will only consider a simple error structure for measurement error and we will only focus on

[5] Strauss (1982, p.335) used this approach in an equivalent way to obtain sales prices by using values and quantities sold.

[6] On Côte d'Ivoire for example he finds that "[t]he quality elasticities are not large, and in the rural sector they are not significantly different from zero. Presumably there is a good deal less choice than there is in the cities" (Deaton (1987, p.23)).

[7] In fact, even in the Côte d'Ivoire survey used in Deaton (1987), for his estimates of price elasticities without price data, subsistence consumption needed to be valued by some imputation, since its value was not directly recorded in the survey, in order to obtain a measure of total annual consumption (as a proxy for income) (p.20). Unfortunately, although these imputed expenditures were used in his analysis, it is nowhere mentioned how they were obtained. In short, even his price elasticities without price data needed price data to implement the procedure.

[8] By using regressions, we can use the variance-covariance structure of the data. Using only the mean or median of a particular variable implies using only limited information about a distribution.

consumer prices. In Capéau (1995) the analysis is extended to incorporate different assumptions on the error structure and he nests the derivation of consumer and producer prices in one specification.

The data used in this paper are derived from a rural household survey conducted in Ethiopia. It covers 1477 households in 15 villages which were interviewed three times during 1994 and 1995. In each round detailed information is available on the value of consumption purchased in the last week and the quantity in local units, and information on the quantity in local units of the consumption obtained from own stocks or production, and from gifts. In each village, a detailed price and conversion factors questionnaire was implemented during each round. The data allow therefore to implement three different ways of obtaining price and conversion information: using the community level questionnaire, using means (or medians) of unit values and using the econometric approach suggested in this paper. In the next section, we describe the model used for the estimations. In section 3 the econometric analysis is presented while in section 4 we evaluate some of the consequences of this and other approaches. Section 5 concludes. Technical details concerning the econometric analysis and detailed results are contained in the annexes.

2. An econometric specification

Suppose we have for some households data on the quantities purchased in the market, expressed in a local unit and the monetary value of the expenditure on this commodity. The data are clustered by communities and prices are assumed to be constant per community. We start from a simple bookkeeping identity: price per unit times the purchased quantity equals monetary expenditures. Basically, the possibility of quality differences is initially ignored. The commodity under consideration is therefore assumed to be homogenous for all households. Local units are also assumed to be fixed per community, although they may vary across communities. We allow for the possibility that a local unit has a different weight in kilograms depending on the actual commodity[9]. We introduce the following basic notation: $q(i)^*_k$ equals the quantity of the good under consideration that is purchased by household i, measured in unit k; p_j is the price per kilogram of that good in community j; $V^*(i)$ equals the amount spent on the good by household i and the a_{jk} is the conversion rate or factor of local unit k into kilograms for commodity i in community j[10]. If values and quantities were measured without error, then they would be related according to following identity:

$$V^*(i) \equiv p_j\ a_{jk}.\ q^*(i)_k. \tag{1}$$

If the unit is measured in kilograms, then a_{jk} equals one. In that case the price p_j can be identified and used for all other observations to obtain the correct conversion factor. We do not observe the true expenditures and quantities but only the (random) guesses of the respondents of these true values. Other procedures then need to be used. For example, two of the few authors who explicitly deal with the issue of both prices and conversions, Lambert and Magnac (1997) define $p(i)_{jk}$ as the commodity's price paid by household i in community j, if

[9] As we will see below, many units used in rural areas are volume units, not weights. As a consequence, the specific gravity of commodities will be relevant for the weight conversion.

[10] Since our analysis is commodity specific, we do not need to index goods.

the quantity purchased is measured in unit k, so that for all k and j, equation (1) can be written as:

$$p(i)_{jk} \equiv a_{jk}.p_j = V^*(i)/q^*(i)_k \qquad (2)$$

One can then replace $V^*(i)$ and $q^*(i)_k$ their *observed counterparts* $V(i)$ and $q(i)_k$, in order to obtain an estimate of $p(i)_{jk}$ [11] . This procedure forms the first step in the estimations proposed by Lambert and Magnac (1997). Next they correct for outliers: estimates of $p(i)_{jk}$ outside the 25-75 percent interval in a jk-cluster are set at the first or third quartile value. Then they calculate means over households of the estimates for $p(i)_{jk}$, denoted by p^*_{jk} (p^*_j for the *numéraire*). Their estimates for conversion rates, denoted as a^*_{jk} , are then equal to:

$$a^*_{jk} = p^*_{jk} / p^*_j \qquad (3)$$

While easy to implement, such two step procedure involves inefficiency, since only means of distributions are used for estimations. Nevertheless, it is more careful than most practices, since it explicitly considers the problems of conversions. In most studies it appears to be assumed away.

In the present study we opt for a full econometric specification of the basic identity (1). A first step in that direction is assuming that the observed $V(i)$ and $q(i)_k$ [12] are random variables, drawn from a particular distribution. Therefore identity (1) is transformed into a stochastic equation:

$$V(i) = p_j\, a_{jk}.\, q(i)_k + u(i) \qquad (4)$$

To illustrate our interpretation: if the observed values are = $V^*(i) + v(i)$, and $q(i)_k = q^*(i)_k + w(i)$, where *both* $v(i)$ and $w(i)$ are independently and normally distributed variables with zero mean and variances $\sigma(v)^2$ and $\sigma(w)^2$ for each observation i, then, by a well known convolution result (Kendall and Stuart, 1969, p. 249-250), it follows that $u(i) = v(i) - p_j\, a_{jk}\, w(i)$, is normally distributed as well with a zero mean. But the variance of u(i) is equal to $(p_j\, a_{jk})^2\, \sigma(w)^2 + \sigma(v)^2$ and hence (j,k)-cluster specific. This reveals some kind of *natural tendency towards heteroskedasticity* with this type of data. The lack of very exact local measurement units (some of these are simply cups or cans) can be another reason to switch from identity (1) towards the stochastic specification in (4). Finally quality differences could break down identity (1) as well. As far as these quality differences are truly random phenomena they could suggest specification (4).

A simple illustration of this approach is the case where there would be no doubt about units of measurement. All quantities are measured in kilograms say. Furthermore, let us assume away the heteroskedasticity problem (people do recall physical units exact, but may doubt about how much they have paid). Equation (4) simplifies to:

---

[11] Below, we drop the asterisks to indicate that observed data on expenditures and quantities are stochastic.

[12] $V(i)$ and $q(i)_k$ are the answers in a survey to the questions "how much did you spend on that commodity in the reference period?" and "how much did you buy of that good?". The intended interpretation is that respondents might not perfectly recall how much they bought or spent and make a guess.

$$V(i) = q(i).p_j + u(i) \tag{5}$$

in which u(i) (from now on) denotes a normally distributed random variable with zero mean and identical variance across observation. It can easily be seen that the mean of the ratio of values and quantities per community is an unbiased linear estimator for $p(i)_j$. However, this is not an efficient estimator: using Gauss-Markov, the variance of this estimator is at least as large as that of least squares estimator (and in most cases strictly larger).

By using an econometric approach to prices and conversion factors, these problems can be avoided. Nevertheless, the disturbance structure of (5) has some drawbacks. Assuming standard normality of the disturbances implies that there is a non-zero probability that an individual would have a sufficiently poor memory that she thinks she received some money when buying the good, or that she bought a negative amount, while paying. Secondly, though satisfying the *natural tendency towards heteroskedasticity*, it is not obvious to have community and to a lesser extent measurement unit dependent variances. Rather, we would expect that errors tend to be greater, when the amount bought is greater. Finally, we have a measurement error model by this specification for which OLS-estimation would be biased. These problems are generally resolved by assuming a multiplicative disturbance term which is lognormally distributed. This gives our basic econometric specification for subsequent analysis:

$$V(i) = p_j\, a_{jk}\, q(i)_k\, e^{u(i)} \tag{6}$$

The variance of V(i) conditional on $q(i)_k$ is now dependent upon the amount consumed. This is as it should be according to our earlier suggestions. In Capéau (1995) the disturbance hypotheses in (5) and (6) are tested against each other. As it turns out the specification in (6) does work much better. In fact -under certain regularity conditions - a transformation of the conversion rates and prices in (6) can be estimated using OLS. In particular, rewrite (6) as:

$$\ln V(i) - \ln q(i)_k = \ln a_{jk} + \ln p_j + u(i) \tag{7}$$

then conversion factors can be obtained from estimated coefficients on dummies for each unit k that appears in each community j; prices for each community can be obtained from dummies per community. To allow estimation, one unit will have to be chosen as the *numéraire* in which prices and conversions will be expressed in (7) and dropped from the regression. Since this is a linear model and given the assumptions on the disturbance term, the least squares estimates will be unbiased and efficient. Define dum_{jk} as a dummy which is one if the unit is k in village j, let unit m be the *numéraire* and define dum_j to be a dummy which is one for an observation in community j. Then for K units and J communities, the equation to be estimated is:

$$\ln V(i) - \ln q(i)_k = \Sigma^{K}_{k=1(k\neq m)}\Sigma^{J}_{j=1} \ln a_{jk}\, dum_{jk} + \Sigma^{J}_{j=1} \ln p_j\, dum_j + u(i) \tag{8}$$

Equation (1) and the other equations were derived under the assumption that there was a unique price per community for a particular observed transaction related to one commodity. This may not be true if there is heterogeneous quality of the commodity purchased by the households. As Deaton (1987) argued, it would be desirable to control for a quality difference in the price-unit value relationship. Let us define the price in community j for commodity i of quality v as p^v_j. Let us define v^i as a quality index of the commodity bought by household i so that p^v_j is equal to $v^i.p_j$. Equation (1) can then be written as:

$$V^*(i) = a_{jk}.q^*(i)_k.v^i.p_j \qquad (9)$$

In a cross-section data set, it then becomes impossible to identify p_j by the approach described above, unless further assumptions are made. For example, one could consider that some households buy a higher quality of the commodity and estimate different prices for different groups using interaction terms. A panel data set could allow more appropriate estimation procedures. Using t to denote different time periods, for those households for whom we observe market transactions over time we can define an intertemporal version of (7) using (9) as:

$$\ln V(i)_t - \ln q(i)_{kt} = \ln a_{jk} + \ln p_{jt} + \ln v^i + u(i,t) \qquad (10)$$

in which we assume that while prices can change over time, conversion factors and quality are constant over time. The random error u(i,t) is assumed to be independently and identically distributed with zero mean and constant variance. If we assume that a particular household buys a commodity of a fixed quality over time, then a fixed effects estimator can be used to obtain estimates for prices for i in each cluster, purged of the quality effect, given appropriately defined dummies for prices in each period and conversion factors for each unit. If in the data, we find that households systematically report amounts of the good in the same unit in each period then the conversion factors cannot be obtained directly from the fixed effects model. Quality is in this formulation a constant for each household. Alternatively, we could assume that the quality effects are drawn randomly from a distribution across the households (but constant over time). A random effects model could then be estimated and quality corrected prices for each period and conversion factors could then be derived from appropriately defined dummy variables[13]. Both approaches will be used in the next section. Note that this remains short of the approach by Deaton (1987) where quality becomes an endogenous variable which is allowed to change for each household in response to prices and incomes: quality is a choice variable. Here we assume that households do not change the quality of the goods consumed even if prices or incomes change across rounds. This may be a strong assumption. However, if measurement error is the main cause of differences in unit values - as most of Deaton's estimations suggest - then we could expect that the panel estimators for the prices will be close to the pooled estimates.[14]

[13] Note that this procedure does not solve all the problems of valuing subsistence consumption in a very satisfactory way. While we could have a normalised price for a commodity, we cannot judge by this the quality of the commodity as it is consumed from own stocks by all households. Indeed, all we have is some price estimate and fixed or random quality effects across households. For those who participate in the market, retrieving the fixed effects could provide us with a household level quality dependent price, which can be used for valuing this household's consumption from stocks. Imputed expenditures for households not participating in the market for this commodity will need to be assumed to be from a particular standard quality.

[14] If the panel estimates are statistically different, then this does not imply necessarily quality to matter. It is also possible that the household level fixed or random effect is actually

An important issue remains as yet unresolved. Estimation of (8) will give point estimates for the logarithms of prices and conversion factors. However, as Goldberger (1968) has shown, simply taking exponents of these coefficients will not give unbiased estimates of the conversions and prices (see also Kennedy (1981)). In fact this procedure would overestimate both prices and conversion rates. Two corrections were proposed by Goldberger, both of which will be reported. The first correction still overestimates the 'true' values on average. The second correction leads to an unbiased estimator. The formulae for these corrections and a discussion can be found in annex 1. Recall that all three estimators (simply taking exponents *and* Goldberger's two corrections) are consistent. In fact simply taking exponents is a maximum likelihood estimator.

But these point estimates will give us no indication of the accuracy of the estimations. To provide a reliable approach, variances and confidence intervals need to be obtained. We report for all three estimators two estimates of the variance, the first being simply a first order approximation, while the latter is based upon the *exact* variance formula of the uncorrected parameter estimator.. Since we work with non-normal distribution, these variances cannot be used in a simple fashion to construct confidence intervals. In the annex we indicate how confidence intervals still could be constructed.

3. Data and estimations

The data used in the application are derived from the Ethiopian Rural Household Survey, conducted by the Economics Department of Addis Ababa University and the Centre for the Study of African Economies at Oxford University. This survey collected data on a panel of 1477 households in 15 villages across the country. Thus far, three rounds have been implemented, resulting in three observations on consumption, incomes, assets, health and anthropometric data. We will only use the food consumption data in this paper. They were collected on a one-week recall basis. Households were prompted for consumption and transactions related to a 40 different food items. Questions were asked on purchases, for which total expenditures in the last week, the quantity and the unit of the purchase. Then they were asked for the same commodity about consumption from their own harvest or stock and quantities and units were recorded. Finally, the same was recorded about consumption from gifts, wages in kind, barter or loans in kind they received from both friends, relatives, NGOs etc.

Across the 1477 households, we recorded 12815 different consumption entries in the last week. Of these, 74 percent were purchases and 22 percent were subsistence consumption. The rest were gifts and loans. The quantities were expressed in 70 different units, of which only two were metric quantity or volume measures (kilograms and litres). Table 1 gives the distributions across some of the main units recorded in the survey with the frequencies involved.

---

measuring some systematic, household specific measurement error. The panel estimates would control for this.

Table 1 Frequencies (in percent) of units in consumption data.

unit	all	purchases	own stock	gifts
kilograms	19.8	20.6	16.5	20.7
kunna	5.4	1.4	17.7	9.2
medeb	9.3	11.5	4.3	1.2
esir	4.5	4.2	6.5	1.2
bobo	2.7	2.2	4.8	0.3
pieces	4.1	2.9	8.5	1.5
litres	3.2	3.2	2.7	4.5
tassa	7.5	7.9	4.5	13.5
kubaya	2.5	2.5	2.9	1.4
birchiko	10.5	13.3	3.3	4.4
sini	12.0	15.0	4.1	4.0
bottles	1.7	1.9	0.6	3.4
guchiye	4.2	2.5	3.6	27.1
sahen	1.4	0.8	3.2	1.5
weket	1.2	1.5	0.0	0.3
other	10.0	8.6	16.8	5.8

Source: First round of Ethiopian Rural Household Survey (1994)

Most units described above are transcriptions of the Amharic names used. Given the century-long dominance of Amharas in Ethiopia, Amharic is largely the lingua franca in much of Ethiopia[15]. Some of these units are in fact volume measures, but used for commodities for which weights would be more relevant for analysis. For example, tassa is a unit for liquids (a big serving can), but commonly used for cereals and pulses as well. Birchiko is a glass. Sini is a cup. For example, it can be the cup in which coffee is served in public houses. But it is also commonly used as a measure to buy oil seeds, coffee beans, spices, etc. in. Medeb is a little pile of for example vegetables, as they are put for sale on the market. It is well known that many of these units are different across communities. Some units are only relevant in particular communities. Factors such as the standard types of pottery or baskets used in different parts of the country are likely to determine some of the volume measures. For example, a tassa tends to be substantially smaller in the South than in the North of the country. To account for this possibility, cluster-specific conversion rates will need to be estimated, if the data permit it.

For 58 commodities, sufficient data points were available to estimate regressions to determine conversion rates and prices[16]. To illustrate the procedure, we report just one regression, for

[15] In some non-Amhara areas the same units may be used, but would be known by other names. Only if there was certainty about the correspondence between units in different languages were they put under the same code. Since the conversion codes estimated for these units are usually cluster-specific, the translation problem would not affect the analysis.

[16] In 19 cases, no observations on kilograms or litres were available so that only conversion rates and prices relative to other units could be estimated. However, in some cases, *numéraires* such as standard soft drink bottles, etc. were available, so that appropriate measures

teff (a cereal which is one of the main staple foods in the country). Equation (8) was estimated using differences in the log of reported expenditures and the log of the quantity reportedly bought. Right hand side variables were dummies for the villages in which transactions were observed and dummies interacting the unit in which the quantity was bought with the village. In a few cases, the lack of observations measured in kilograms prevented us to use interaction dummies between measurement units and villages, every time when this was desirable. Table 2 presents this regression.

could be constructed. Also, for conversions of subsistence consumption, this was not a problem, since what is needed is a price for a local unit, not necessarily in kilograms.

Table 2 Regression for determination of prices and conversion rates teff. Data from ERHS (1994)

variable	coefficient	standard error
village 1	0.928	.2441
village 3	0.789	.2441
village 4	0.355	.1114
village 5	0.784	.2502
village 6	1.483	.3594
village 7	0.548	.1342
village 9	0.784	.2441
village 10	1.099	.3453
village 12	0.406	.1685
village 13	0.375	.1726
village 14	0.693	.3453
village 15	0.000	.3453
village 16	0.604	.1801
kunna*village 3	1.953	.3152
kunna*village 4	2.640	.3628
kunna*village 7	1.847	.1872
kunna*village 9	1.988	.4229
kunna*village 13	2.957	.3860
quintal	4.481	.2229
bobo	0.338	.3628
kubaya	-0.978	.2119
birchiko	-0.724	.3782
other	0.310	.4229
guchiye	0.348	.4345
sahen	-0.280	.2270
tassa*village 3	-0.096	.2990
tassa*village 4	0.057	.3628
tassa*village 9	-0.902	.4229
tassa*village 10	-1.792	.4883
tassa*village 14	-0.462	.3987

n=84 joint significance F=58.088

Since kilogram is used as the *numéraire*, the coefficients can be interpreted as the logarithm of the prices in kg in each village and the conversion rates in kg for each unit. The regression is jointly significant, but this is hardly a surprise for an analysis which starts from an identity, so that the non-stochastic part of our model is beyond any doubt. The standard errors give some indication on how well prices and conversion rates may be estimated. All (logs of) prices and most (logs of) conversion rates are significantly different from zero. Some conversion rates are not significant and closer inspection revealed that this is for those units appearing very infrequent in the data.

Since we want to use the estimated values for further analysis, we are mainly interested in the point estimates. We also want to know how much confidence we can have in these estimates and whether prices and conversion factors from other sources are consistent with these results. Obviously, the point estimates of interest are the conversion rates and prices themselves, not their logarithms. As has been mentioned in the previous section, simply taking an exponents of estimated parameters results in an *over*estimation of both prices and conversion rates.

Therefore we report on the basis of similar regressions as the previous one, pooling the data of different rounds - corrected estimates for teff, maize and wheat (table 3). These are three important commodities both in agriculture and in consumption. We only give the results for two of the most common non-standard units (*tassa* and *kunna*) for a few villages in which these units are important. We calculated three point estimates: the ML-estimator (simply taking exponents, corresponding to (A.1) in annex 1) and Golbergers' two corrections (respectively (A.3) and (A.4)). We computed also two estimates for the variance, based on (A.6) to (A.8) and on (A.9). Finally 95 percent unidimensional confidence intervals were computed (using (A.12)). In the main text we only give the ML-estimator and the first of the two Goldberger corrections, the second variance estimator (using (A.9)) of the corrected estimator and the implied 95 percent confidence interval. In annex 2, the full results are given. For comparison, conversion rates and prices from the survey conducted at the community level are given, as are the mean unit values.

Table 3 **Conversion factors in kilograms per unit for selected units (kunna and tassa) and villages; pooled sample.**

		one kunna of teff					one tassa of teff				
villages		Dinki	Debre Berhan	Sirbana Godeti	Korode-gaga	Imdibir	Dinki	Debre Berhan	Sirbana Godeti	Korode-gaga	Imdibir
estimates	ML estimate	7.797	7.419	6.186	7.303	-	0.711	0.964	-	0.406	0.416
	corrected estimate (1)	7.719	7.373	6.153	6.900	-	0.705	0.925	-	0.383	0.408
	variance (2)	1.226	0.692	0.399	6.403	-	0.008	0.080	-	0.020	0.006
	95% confidence interval	(6.380, 9.339)	(6.014, 9.038)	(5.267, 7.189)	(4.026, 11.830)	-	(0.616, 0.807)	(0.540, 1.586)	-	(0.224, 0.657)	(0.328, 0.509)
community survey		5.00*	5.00*	7.00	5.00	-	1.00*	1.00	-	0.60	0.75*

		one kunna of maize					one tassa of maize				
villages		Dinki	Debre Berhan	Sirbana Godeti	Korode-gaga	Adado	Dinki	Debre Berhan	Sirbana Godeti	Korode-gaga	Adado
estimates	ML estimate	-	-	6.395	5.670	-	0.914	-		0.530	0.387
	corrected estimate (1)	-	-	6.226	5.559	-	0.893	-	-	0.520	0.366
	variance (2)	-	-	2.253	1.302	-	0.040	-	-	0.010	0.017
	95% confidence interval	-	-	(4.274, 9.068)	(4.606, 6.709)	-	(0.684, 1.165)	-	-	(0.450, 0.601)	(0.340, 0.395)

Table 3 (cont.) Conversion factors in kilograms per unit for selected units (kunna and tassa) and villages; pooled sample.

		one kunna of wheat					one tassa of wheat				
villages		Dinki	Debre Berhan	Sirbana Godeti	Korode-gaga	Adado	Dinki	Debre Berhan	Sirbana Godeti	Korode-gaga	Imdibir Adado
estimates	ML estimate	7.212	6.065	5.178	9.384	7.071	0.903	0.994	-	0.597	0.783
	corrected estimate (1)	6.750	6.033	5.115	8.999	6.841	0.866	0.983	-	0.585	0.763
	variance (2)	7.353	0.390	0.658	7.688	3.418	0.070	0.023	-	0.014	0.031
	95% confidence interval	(4.727. 9.640)	(4.987, 7.299)	(4.164, 6.284)	(5.437, 14.900)	(4.790, 9.770)	(0.673, 1.115)	(0.735, 1.314)	-	(0.503, 0.681)	(0.651, 0.895)
community survey		5.00	5.00	7.00*	5.00	6.00	1.00	1.00	-	0.55	0.75

Notes:
Results based on sample regressions pooled over the different villages using data from the first round of the ERHS. ML estimates are the coefficient s from the regression (8) using (A.1), i.e. simply taking exponents. The corrected estimate (1) is the first Goldberger correction, using (A.3). The variance (2) is the estimate of the variance of the latter corrected estimator, using (A.9). The 95 percent confidence interval (3) is calculated using (A.12). Full results are in annex 2. A discussion of the estimators and corrections can be found in Annex 1.
*=community survey estimate falls outside 95 percent confidence interval

The results suggest that most conversion factors are estimated with relatively small standard errors. The values obtained are mostly in line with those obtained from the community level surveys collected in each village. The Goldberger corrections are uniformly lower than the ML-estimates. This is as it should be since the ML-estimates are *biased* upwards. Because the first correction is still biased upwards, it is no surprise to see that the unbiased estimator (the second correction) gives still (slightly) lower values than the first one, though this is not necessarily an analytical result (see the results in annex 2). It is however much more important to see that the first correction, which is computationally much easier turns out be a very good approximation for the unbiased estimator. In a few cases the community estimates appear to be quite inconsistent with the estimates from the regressions, since they lie outside five percent interval implied by the standard errors of the regressions. If there is no reason to assume that the answers of our respondents are biased, this means the observation(s) drawn to construct the community level estimate (based on data from a separate price survey) come from a different distribution. Using the community level estimate may then be misleading in the analysis of the data.

Table 4 provides some data on prices, obtained from the same regressions as the estimates in tables 3. These are compared with the estimates from the community level price survey, and mean unit values (values divided by quantities).

Table 4 Estimated prices, calculated unit values and price survey data (per kilogram), for teff, wheat and maize for selected villages. Pooled regressions; values for first round only (1994).

Prices for teff	Dinki	Debre Berhan	Sirbana Godeti	Korodegaga	Adado
ML estimate	2.374	1.566	1.775	2.191	2.731
corrected estimate (1)	2.354	1.561	1.768	2.150	2.677
variance (2)	0.094	0.016	0.026	0.185	0.299
95% confidence interval	(1.99, 2.79)	(1.34, 1.81)	(1.54, 2.03)	(1.64, 2.81)	(2.05, 3.51)
community survey	2.00	1.50	1.45*	2.20	-
mean unit values	2.48	1.81	1.90	2.27	1.83*
Prices for maize	Dinki	Debre Berhan	Sirbana Godeti	Korodegaga	Adado
ML estimate	1.088	1.226	0.979	1.450	1.476
corrected estimate (1)	1.068	1.207	0.970	1.428	1.403
variance (2)	0.045	0.049	0.019	0.063	0.233
95% confidence interval	(0.83, 1.37)	(0.89, 1.64)	(0.75, 1.25)	(1.27, 1.61)	(1.11, 1.78)
community survey	1.00	-	0.98	1.60	-
mean unit values	1.17	1.18	1.25	2.15*	0.93*
Prices for wheat	Dinki	Debre Berhan	Sirbana Godeti	Korodegaga	Adado
ML estimate	1.733	1.407	1.101	1.279	1.414
corrected estimate (1)	1.677	1.405	1.095	1.268	1.391
variance (2)	0.205	0.005	0.013	0.029	0.067
95% confidence interval	(1.37, 2.06)	(1.30, 1.53)	(0.94, 1.27)	(1.06, 1.51)	(1.16, 1.66)
community survey	1.76	1.50	1.42*	1.33	-
mean unit values	1.87	1.58*	1.09	1.73*	1.54

Notes:
Results based on sample regressions pooled over the different villages using data from the first round of the ERHS. ML estimates are the coefficient s from the regression (8) using (A.1), i.e. simply taking exponents. The corrected estimate (1) is the first Goldberger correction, using (A.3). The variance (2) is the estimator of the variance of the latter corrected estimator, using (A.9). The 95 percent confidence interval is calculated using (A.12). Full results are in annex 2. A discussion of these estimators and the corrections can be found in Annex 1.

The corrected estimators are again uniformly lower than the ML-estimator and the variance is relatively small. The confidence intervals may still appear quite large for some prices in some areas. Nevertheless, in a significant number of cases, the estimates from the price survey and especially the mean unit values lie outside this confidence interval. Several values for prices of these cereals were also missing in the community level survey. For those prices, the household survey data provide the only way to obtain prices.

Panel data estimates using fixed and random effects are provided in table 5 for maize for four villages in the Southern part of the country, where this is an important crop both in production and in consumption[17]. Equation (9) was estimated as a pooled cross-section regression, as a

[17] Only for few commodities enough observations over time for a particular commodity could be found. In line with seasonal patterns in harvests and prices, many households did not

household-level random effects model and as a household-level fixed effects model. Since many households who purchased maize in more than one period did so using different units, it was possible to estimate conversion factors for the main units, even in the fixed effects model[18]. However, given the way dummies are used to allow the estimation of prices in each round, perfect collinearity between the price dummies and the household level fixed effects required to drop one of the time dummies for each village. The remaining time dummies measure therefore the difference relative to the dropped time dummy (included in the fixed effects). To allow direct comparison with the pooled cross-section and the random effects model, the latter were respecified so that the price dummies in the second and in the third round could be interpreted as changes in (the logarithm of) prices as well[19]. We report the actual coefficients; prices and conversions can be obtained taking exponents.

The results suggest that the restrictions imposed by a random effects or a fixed effects model are not significantly different from zero suggesting that the pooled regression outperforms the other regressions: the LM-test on the random effects versus the pooled regression is not significant, while the F-statistic testing the fixed effects model relative to the pooled regression is also (just) not significant at the 5 percent level. (The fact that the Hausman-test argues in favour of the fixed effects model relative to the random effects model does not change this conclusion.) In line with this finding, looking closer at the estimated coefficients, we observe only marginal differences. All this suggests that it is unlikely that there is a systematic

---

purchase the same commodities in each period limiting the scope for panel data estimation techniques.

[18] In other words, the unit in which households bought maize in each round often changed over time, so that the dummy reflecting the unit was not cancelled out by differencing.

[19] We included therefore a constant and a time invariant dummy for three villages. Prices in the first round can then be found using the constant for the village without a dummy (Adele Keke), while the coefficients on the time invariant dummy measure the difference in the price in round 1 between this base village and the respective villages.

difference between the qualities of commodities purchased by different households[20]. As a consequence, it appears to be correct to interpret the estimated coefficients of the pooled regressions as market prices, which are not too much affected by quality difference problems.[21]

[20] An alternative interpretation of these results is that there is unlikely to be a systematic measurement error per household in recording quantities and values, or a systematic difference in the conversion factors of the local units used by households.

[21] Of course, we have to acknowledge again that the specification used does not allow quality to become a choice variable for the household, as in Deaton (1987), although as was argued before, given the thin markets in rural Ethiopia and in line with Deaton's own findings on rural areas, this may not be an important shortcoming.

Table 5 **Pooled cross-section, random household effects and fixed household effects estimation for maize. Three rounds of ERHS.** Sample restricted to four villages and to panel observations for households. Unbalanced panel estimates; 307 observations from 145 households. Standard errors in brackets.

	Pooled				Random effects				Fixed effects			
variable	Adele Keke	Korode-gaga	Aze Deboa	Garagodo	Adele Keke	Korode-gaga	Aze Deboa	Garagodo	Adele Keke	Korode-gaga	Aze Deboa	Garagodo
kunna	1.934	2.076	2.207	3.510	1.947	1.983	2.164	3.500	2.021	1.298	1.938	3.432
	(0.287)	(0.348)	(0.152)	(0.186)	(0.281)	(0.355)	(0.145)	(0.183)	(0.398)	(0.694)	(0.190)	(0.247)
tassa	-0.286	-0.457		-0.439	-0.305	-0.539		-0.466	-0.397	-0.940		-0.625
	(0.104)	(0.318)		(0.234)	(0.109)	(0.309)		(0.232)	(0.190)	(0.439)		(0.320)
birchiko			-1.452	-0.768			-1.482	-0.750			-1.568	-0.746
			(0.104)	(0.172)			(0.101)	(0.167)			(0.143)	(0.219)
price round 2[2]	0.308		-0.533	-0.344	0.304		-0.498	-0.336	0.293		-0.422	-0.330
	(0.085)		(0.121)	(0.090)	(0.077)		(0.115)	(0.085)	(0.085)		(0.139)	(0.103)
price round 3[2]	0.135	0.091	0.053	0.006	0.136	0.110	0.052	0.003	0.144	0.247	0.046	-0.001
	(0.105)	(0.191)	(0.054)	(0.054)	(0.098)	(0.174)	(0.049)	(0.049)	(0.120)	(0.219)	(0.054)	(0.055)
round 1 price[1]	0.621	-0.333	-0.212	-0.951	0.624	-0.269	-0.185	-0.969				
	(0.064)	(0.288)	(0.120)	(0.187)	(0.064)	(0.284)	(0.118)	(0.183)				
	joint significance: F(19,287)=261.06**				LM-test of random effects versus pooled: $\chi^2(1)= 0.40$ Hausman-test fixed versus random: $\chi^2(16) =32.95$**				joint significance: F(162,144)=34.46** fixed effects model versus pooled model: F(143,144)=1.18			

**= significant at 1%

*= significant at 5%

[1]= coefficient on constant for Adele Keke; coefficients on site dummies over the entire sample for other villages; these coefficients measure difference from Adele Keke.

[2]=differences relative to first round price coefficients.

4. Application: poverty changes in rural Ethiopia 1994-1995

To show a potential use of the econometric approach to derive price and quantity conversions, we will focus on changes in food poverty lines and food poverty levels over the first three rounds of the Ethiopian rural household survey collected in 1994 and 1995[22]. Using the data, we calculated food consumption levels per month per capita, using different methods of converting and valuing non-purchased consumption. Mean consumption levels per capita across the three rounds in nominal terms are given in table 6.

Table 6 Mean consumption levels per capita in three rounds ERHS, using different conversion and price data (in birr; exchange rate in 1994/5: 6 birr per $)

	food per capita using price survey and conversion survey		food per capita using mean unit values and conversion survey		food per capita using regression approach	
1994	59.6	(72.1)	60.1	(73.8)	65.0	(95.0)
1994a	75.5	(73.2)**	75.7	(73.5)**	78.7	(78.6)**
1995	69.4	(99.7)**	69.5	(99.6)**	71.8	(103.4)*

Survey data were collected during first half of 1994, second half of 1994 (1994a) and the first half of 1995. The timing within the seasonal calendar of the first and third round coincide. Price survey and conversion survey were collected in nearest market and within the village. Mean unit values are mean values per community of the expenditure on a purchased food item divided by the amount in kg bought by the household. Amounts in kg obtained using conversions from conversion survey. The regression approach involves the results from the cross-section regressions as in (7) for each commodity. Missing prices and conversions were in all cases taken from community surveys. In brackets, the standard deviation is given. We also implemented a test on differences between the means assuming independent samples. The test-statistic for 1994a compares its value with 1994; the test-statistic for 1995 compares 1995 with 1994. **=significant at 1 percent; *=significant at 5 percent.

Note the low mean consumption levels: only about $10-13 per month. The regression approach results in higher food consumption estimates, although the pattern in nominal consumption per capita is similar, with an increase in the second half of 1994 and a decrease in 1995, although the level remains considerably higher than in the first round. However, since these estimates are in nominal terms, appropriate deflators are needed. Appropriately defined poverty lines can provide these deflators.

To construct a poverty line, we established a basket of food commodities which provide 2300 Kcal per person, taking into account a typical diet of the poorer half of the sample[23]. The

[22] Data on non-food consumption are also available and have been used for poverty calculations elsewhere (Dercon and Krishnan (1996)). Since the focus in this paper is on food prices and conversions, we only focus on food consumption and food poverty.

[23] The approach has also been used for Ethiopia in Dercon and Krishnan (1996). Essentially the approach for the minimum costs of basic food needs as in Ravallion and Bidani (1995) is followed, using the diet of the poorer half of the sample as a base. The result is a basket of 18 different commodities, including five cereals, three pulses, milk, coffee, salt, sugar, spices, potatoes, enset, onions and cabbage. By using the poorer half of the sample for the weights in the poverty line, one may argue that the poverty lines are not good deflators for the entire sample. Since the

value of this basket, using unit values, estimated values and prices from the price survey, are given in table 7. The table also shows the food CPI for Ethiopia and the value of the poverty line using an alternative regional level price source collected by the Ethiopian Central Statistical Office.

Table 7 Index of poverty line values using different price sources (1994=100).

	poverty line using unit values	poverty line using price survey	poverty line using regressions	poverty line using CSA rural prices	Food CPI
1994	100	100	100	100	100
1994a	98	106	107		106
1995	111	113	113	107	104

'Food CPI' is the national rural consumer price index collected by the Central Statistical Office.
The 'poverty line using CSA rural prices' is the average index value of the basket of commodities valued using prices for the respective regions obtained from the Central Statistical Authority (CSA).
The 'poverty line using unit values' is the average index value of the basket of commodities valued using the mean unit values from the survey data. The value in birr for 1994 was 45 birr per month.
The 'poverty line using price survey' is the average index value of the basket of commodities valued using the price survey collected in each survey site. The value in birr for 1994 was 37 birr per month.
The 'poverty line using regressions' is the average index value of the basket using the estimated prices from the regressions in each round and for each commodity. The value in birr for 1994 was 38 birr per month.
All these poverty lines are population weighted means for the sample. In applying these poverty lines for poverty measures, site specific poverty lines were used to reflect site-specific price differences.

The table shows that that the poverty lines over time will be sensitive to the source of price data used. Unit values show a small decline in the second half of 1994 and an increase afterwards. Both the price survey as the estimated prices using the regressions show an increase in both the later part of 1994 and in 1995. This pattern is however quite different from the Food CPI or from other rural prices obtained from the Central Statistical Authority (CSA). Indeed, these inflation estimates illustrate well one of the typical puzzles when comparing household data over time. Official inflation figures show that prices between the different rounds did not increase very much: rural inflation is suggested to have been very low. When applied to the mean consumption levels (and as we will show the poverty measures) then this would have implied quite important increases in the standard of living in rural Ethiopia between 1994 and 1995. However, the price survey implemented in the survey sites of the ERHS suggested important price increases of about 13 percent between roughly the same periods in 1994 and 1995. The poverty line calculations using the CSA rural prices, shown in the table, suggest that this difference is not due to different weights to commodities or to regions in the rural food CPI: calculated in this way, the poverty line rose only by 7 percent in the same period. Since measurement error in the price survey may be at the root of these differences, the econometric approach to obtain prices provides an alternative means of assessing the changes in prices over the period. As can be seen, the results using the econometric estimates of the prices is very close to those using the price survey data, suggesting that the price increases are genuine. This also illustrates that the approach may form an alternative for assessing changes in the cost of living over time in different cross-

focus in this paper is quite different, we will use the poverty lines also as the deflators for nominal consumption.

sections or in panel data, when price survey data in the clusters are missing. This alternative is likely to be more reliable than to rely on price data collected in a very different context.

Table 8 illustrates the potential consequences on poverty measures of using the different sources of price data. We report the head count and the poverty gap index for each of the three rounds using four possible poverty lines: using the price survey data, using the unit values, using the regressions and using the implied price changes from the food CPI. We also give mean consumption per capita in 1994 prices. t-tests of differences in the estimates, comparing each estimate with the value for 1994 are given in brackets as well. Consistent standard errors for the poverty measures and for testing differences in poverty estimates assuming independent samples are obtained using Kakwani (1986).

Table 8 Poverty and consumption using different price sources 1994-1995.

		using food CPI		using price survey data		using regressions		using mean unit values	
Head Count P_0	1994	0.46		0.46		0.44		0.54	
	1994a	0.37	(-4.85)	0.36	(-5.43)	0.35	(-4.90)	0.38	(-8.68)
	1995	0.42	(-2.15)	0.44	(-0.89)	0.44	(-0.02)	0.52	(-1.24)
Poverty Gap P_1	1994	0.19		0.19		0.17		0.24	
	1994a	0.13	(-6.48)	0.13	(-7.47)	0.12	(-5.85)	0.14	(-9.76)
	1995	0.18	(-1.56)	0.19	(-1.56)	0.19	(1.28)	0.22	(-1.64)
Real consumption	1994	60		60		65		60	
per capita in birr	1994a	71	(4.45)	70	(4.08)	75	(2.98)	77	(6.11)
(1994 prices)	1995	67	(2.26)	61	(0.43)	66	(0.26)	63	(0.79)

Definitions as in tables 6 and 7.
In brackets, t-tests for difference in mean, assuming independent samples. Standard errors for poverty measures and differences between poverty measures are obtained using Kakwani (1986).
Differences tested are 1994a relative to 1994 and 1995 relative to 1994.

All the results show a large reduction in poverty and a large increase in real consumption in the second half of 1994. For all estimates and for all measures, the reduction between 1994 and 1994a is significant. This is likely to be a seasonal effect. Since the first and third round were collected at roughly the same period, their comparison is likely to be more relevant for a measurement in changes in the standard of living across the panel households. The results illustrate the problems with using the CPI estimates to update the poverty line. In that case, there is a significant reduction in the head count index between 1994 and 1995, while there is a significant growth in per capita consumption. However, when using the price survey or the regressions, we find no significant change in poverty or per capita real consumption in this period. Using the CPI data would have meant quite misleading results in the change in food consumption and poverty in this period. Note finally that using the mean unit values would have yielded the same pattern over time as the price survey or regression approach. The poverty measure itself is however more than a fifth higher and the fluctuation over time is larger when using the unit values compared to the other price data.

5. Conclusions

The paper developed a systematic approach to obtain prices and conversion factors to value non-marketed production in semi-subsistence economies. Building on Deaton's (1987) work, but focusing more on the measurement problem, we derive a simple approach to estimate unit values (prices) and conversion factors from regressions using the observations on the values of purchased consumption and the amounts purchased in local units. Since it considers explicitely the measurement error distribution, it is shown that it is superior to the usual ad-hoc procedures used to obtain prices and to convert non-marketed consumption in values and quantities. Random and fixed effects estimation procedures suggest that the unit values obtained from the standard regressions may well be interpreted as prices, and that the cross-section variability is mainly linked to non-household specific measurement error, rather than household level quality differences. Although the paper focused on the consumption data, the approach could also be used on production data, as long as the commodities are sold by at least some members of the cluster.

The application to changes in poverty in Ethiopia clearly shows some of the advantages of the approach described in this paper. Since in each cluster prices were collected, the econometric approach to determine prices allowed us to check the validity of the results obtained via the price and conversion survey data. The price survey showed a much larger price increase in the sample than what was implied by the national data from the Central Statistical Authority. Since the price data proved to give consistent results with those obtained by the regression approach, the CSA data should be questioned, at least relative to our survey villages. The results is therefore that consumption and poverty changed very little between 1994 and 1995, contrary to the results using the official inflation figures.

The results on Ethiopia provide therefore a warning for research using price data from a source unconnected to the underlying surveys to obtain price-corrected poverty lines over time. If no price data were collected, it would be useful and maybe necessary to use the econometric approach described in this paper to provide reliable results.

References

Capéau, B., (1995), “Measurement error and functional form: a proposal to estimate prices and conversion rates from the ERHS1994”, mimeo.

Davidson, R.. and J.G. MacKinnon, (1993), *Estimation and Inference in Econometrics*, Oxford, Oxford University Press

Deaton, A., (1987), “Estimation of own- and cross-price elasticities from household survey data”, *Journal of Econometrics*, vol.36, pp.7-30.

Deaton, A., (1988), “Quality, quantity and spatial variation of prices”, *American Economic Review*, vol.78, pp.418-430.

Deaton, A., (1990), “Price elasticities from survey data. Extensions and Indonesian results”, *Journal of Econometrics*, vol.44, pp.281-309.

Deaton, A., (1997), *The Analysis of Household Surveys: a Microeconometric Approach to Development Policy*, Washington D.C. and Baltimore: The World Bank and Johns Hopkins University Press.

Dercon, S. and P. Krishnan, (1996), “A consumption-based measure of poverty for rural Ethiopia in 1989-1994”, in: B. Kebede and M. Taddesse, *Poverty and Economic Reform*, Proceedings of the fifth Annual Conference of the Ethiopian Economic Association, Debre Zeit.

Fuller, W.A., (1987), *Measurement Error Models*, New York, John Wiley and Sons.

Glewwe, P. (1990), “The Measurement of Income Inequality under Inflation”, *Journal of Development Economics*, vol.32, pp.43-67.

Goldberger, A., (1968), “The Interpretation and Estimation of Cobb-Douglas Functions”, *Econometrica*, 35, 464-472.

Greene, W.H., (1993), *Econometric Analysis*, 2nd ed., New York, Macmillan

Johnson, N.L. and S. Kotz, (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, New York, John Wiley and Sons,

Kakwani, N., (1986), “Measuring Poverty: Definitions and Significance Tests with Application to Côte d’Ivoire”, *Review of Economics and Statistics*, 43-66.

Kendall, M.G. and A. Stuart, (1969), *The Advanced Theory of Statistics*, vol. I, 3rd ed., London, Charles Griffin and company ltd.

Kennedy, P.E., (1981), "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations", *American Economic Review*, 71, 801

Lambert, S. and T. Magnac, (1997), "Implicit prices and recursivity of agricultural households' decisions", INRA and CREST, mimeo.

Levin, C., (1991), "Rural household data collection in developing countries: designing instruments and methods for collecting consumption and expenditure data", Working Papers in Agricultural Economics, Department of Agricultural Economics and Cornell Food and Nutrition Policy Program, October.

Low, A., (1986), *Agricultural Development in Southern Africa: Farm-household Economics and the Food Crisis,* Portsmouth, NH: Heinemann.

Ravallion, M. and B. Bidani, (1994), "How Robust is a Poverty Profile", *World Bank Economic Review,* vol.8, no.1, pp.75-105.

Singh, I., L. Squire and J. Strauss, (1986), *Agricultural Household Models: Extensions, Applications and Policy*, Baltimore: Johns Hopkins University Press.

Van de Walle, D., (1988), "On the use of the SUSENAS for Modelling Consumer Behavior", *Bulletin of Indonesian Economic Studies*, vol.24, no.2, pp.107-122.

Annex 1 : estimators and variances for prices and conversion rates

In the main text we report three different estimates for prices and conversion rates. For each of them we also give two different estimates of their variance. Finally we constructed unidimensional confidence intervals. The present annex contains some information on how these numbers were calculated. We omitted a large amount of intermediate steps. All of them could be obtained upon request from the authors. In order to simplify notation we introduce the following variables:

$\Delta =$ the matrix of dummies serving as explaining variables in equation (8) of the main text; $\beta =$ the vector of unknown parameters to be estimated in equation (8). It are the logarithms of prices and conversion rates; $\gamma =$ the vector of prices and conversion rates, i.e. the vector of unknown parameters we want to estimate. By definition we have: $\gamma \equiv \exp(\beta)$, where the exponent operator, applied on a vector, is executed pointwise. This convention will be maintained throughout this annex for all functions which are usually defined on scalars. Finally $\Sigma(\hat{\beta})$ will denote the variance-covariance matrix of an estimator $\hat{\beta}$ for β.

A.1 Point estimators

Our first estimator is the *usual* one for this type of problems. Let $\hat{\beta}_{\mathrm{OLS}}$ be the OLS-estimator for β in equation (8). Then:

$$\hat{\gamma} = \exp\left(\hat{\beta}_{\mathrm{OLS}}\right) \tag{A.1}$$

It is a well-known result that maximum likelihood estimators are invariant for reparametrisation of a model (see Davidson and MacKinnon (1993), p. 253-255). Therefore (A.1) is rightly called a maximum likelihood estimator in the main text.

However, such a maximum likelihood estimator is often biased. This criticism also applies to the present case. Since $\hat{\beta}_{\mathrm{OLS}}$ is known to be normally distributed according to $\mathrm{N}\left(\beta, \Sigma\left(\hat{\beta}_{\mathrm{OLS}}\right)\right)$, and $\hat{\gamma}$ is simply an exponent of a vector of normally distributed random variables, it can be shown by change of variables that $\hat{\gamma}$ is lognormally distributed according to $\mathrm{LN}\left(\beta, \Sigma\left(\hat{\beta}_{\mathrm{OLS}}\right)\right)$[24]. According to a well-known property of the lognormal distribution it follows that:

$$\mathrm{E}(\hat{\gamma}) = \exp\left(\beta + \frac{1}{2}\mathrm{vecdiag}\Sigma\left(\hat{\beta}_{\mathrm{OLS}}\right)\right) \tag{A.2}$$

where $\mathrm{vecdiag}\,\mathrm{A}$ stores the diagonal of a square matrix A in a column vector.

[24] As usual the parameters of a lognormal distribution are the mean vector and the variance-covariance matrix of the logarithm of these variables, not to be confused with the mean and the variance-covariance matrix of the variables themselves.

Hence the maximum likelihood estimator $\hat{\gamma}$ is biased. Therefore Goldberger (1968) (see also Kennedy (1981)) proposed two corrections:

$$\tilde{\gamma} = \exp(\hat{\beta}_{OLS} - \frac{1}{2}\mathrm{vecdiag}\left(\hat{\Sigma}\left(\hat{\beta}_{OLS}\right)\right)) \tag{A.3}$$

$$\tilde{\tilde{\gamma}} = \exp(\hat{\beta}_{OLS}) \bullet F\left(\hat{\Sigma}\left(\hat{\beta}_{OLS}\right), v\right) \tag{A.4}$$

where $\hat{\Sigma}\left(\hat{\beta}_{OLS}\right)$ is the usual estimator of the variance-covariance matrix of a vector of parameters in an OLS-regression, v is the number of degrees of freedom in the OLS-regression and:

$$F\left(\hat{\Sigma}\left(\hat{\beta}_{OLS}\right), v\right) = \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}v\right)^j \Gamma\left(\frac{1}{2}v\right)}{\Gamma\left(\frac{1}{2}v + j\right)} \cdot \frac{\left(-\frac{1}{2}\mathrm{vecdiag}\left(\hat{\Sigma}\left(\hat{\beta}_{OLS}\right)\right)\right)^j}{j!} \tag{A.5}$$

with $\Gamma(\cdot)$ equal to the Gamma-function en $\bullet$ is pointwise multiplication of two vectors.

The first correction is easy to calculate but still biased. It is still an overestimation of the 'true' values. The second one gives an unbiased estimator but requires some approximation to be calculated in practice. In fact the first correction turns out to be a good approximation in the regressions we performed (see table 3 and 4 in the main text). It should be noted that all estimators for prices and conversion rates proposed here, are consistent.

A.2 Variances

For estimating the variance covariance matrix of the estimators proposed in the previous section we used two methods. First we applied the delta-method, which is a first order approach to calculate moments of continuous and invertible functions of random variables (see Fuller, 1987, p.85-87 and Greene, 1993, p.75). This results in:

$$\hat{V} = \left[\mathrm{diag}\ \exp(\hat{\beta}_{OLS})\right]\left[\hat{\Sigma}(\hat{\beta}_{OLS})\right]\left[\mathrm{diag}\ \exp(\hat{\beta}_{OLS})\right] \tag{A.6}$$

$$\tilde{V} = \left[\mathrm{diag}\ \tilde{\gamma}\right]\left[\hat{\Sigma}\left(\hat{\beta}_{OLS}\right) + \frac{2\hat{\sigma}^4}{v} D\right]\left[\mathrm{diag}\ \tilde{\gamma}\right] \tag{A.7}$$

$$\tilde{\tilde{V}} = \left[\mathrm{diag}\ \tilde{\tilde{\gamma}}\right]\left[\hat{\Sigma}\left(\hat{\beta}_{OLS}\right)\right]\left[\mathrm{diag}\ \tilde{\tilde{\gamma}}\right] + \left[\mathrm{diag}\ \tilde{\gamma}\right]\left[\hat{\Omega}\right]\left[\mathrm{diag}\ \tilde{\gamma}\right] \tag{A.8}$$

where $\mathrm{diag}\, a$ converts a vector a into a diagonal matrix with the elements of a on the diagonal; $D \equiv \left(\mathrm{vecdiag}(\Delta'\Delta)^{-1}\right) \cdot \left(\mathrm{vecdiag}(\Delta'\Delta)^{-1}\right)'$; $\hat{\sigma}^4$ is the square of the usual estimator for the variance of the OLS-regression $\hat{\Omega} = \left[\frac{\partial F}{\partial \hat{\Sigma}}\right]\left[\frac{2\hat{\sigma}^4}{v} D\right]\left[\frac{\partial F}{\partial \hat{\Sigma}}\right]$. [25]

[25] We thank Geert Dhaene for removing some errors in previous formulations of (A.7) and (A.8).

Since we know that $\hat{\gamma}$ is lognormally distributed according to $LN\left(\beta, \Sigma\left(\hat{\beta}_{OLS}\right)\right)$, using the formula for the variance-covariance matrix of a multivariate lognormal distribution, which is known (see Johnson and Kotz, 1972, p.20), could form the basis for another estimator of the variance-covariance matrix. For the ML-estimator this results in:

$$
\begin{aligned}
V(\hat{\gamma}) = \operatorname{diag}&\left[\exp(\hat{\beta}_{OLS} + \frac{1}{2}\operatorname{vecdiag}\left(\hat{\Sigma}\left(\hat{\beta}_{OLS}\right)\right)\right] \cdot \\
&\left[\exp\left(\hat{\Sigma}\left(\hat{\beta}_{OLS}\right)\right) - J\right] \cdot \operatorname{diag}\left[\exp(\hat{\beta}_{OLS} + \frac{1}{2}\operatorname{vecdiag}\left(\hat{\Sigma}\left(\hat{\beta}_{OLS}\right)\right)\right]
\end{aligned}
\tag{A.9}
$$

in which J is a square matrix in which each element equals one. It should be noted that this estimator is based on an exact expression for the variance-covariance matrix of $\hat{\gamma}$. Nevertheless we do not corrected for the fact that the variance-covariance matrix of the OLS estimator was replaced by an estimator, hence by a random variable. Replacing in (A.9) $\hat{\beta}_{OLS}$ by $\ln\tilde{\gamma}$ or $\ln\tilde{\tilde{\gamma}}$ provides our second estimator for the variance-covariance matrices of the corrected estimators. In this case, these do not depart from an exact expression for the variance-covariance matrices since neither $\tilde{\gamma}$ nor $\tilde{\tilde{\gamma}}$ are lognormally distributed.

A.3 Confidence intervals

We opted for calculating unidimensional $(1-\alpha)\%$ confidence intervals conditional upon the other parameters being evaluated at the point-estimates. This gives smaller intervals than the unconditional ones based on for example t-statistics in OLS-regressions. For the purpose upon hand a smaller - though theoretically sound - interval is better since it signals much faster when community survey estimates differ significantly from the estimated value. Again, we started from the fact that $\hat{\gamma}$ is known to be lognormally distributed for solving out the following integrals to $\overline{\gamma}_1$ and $\underline{\gamma}_1$:

$$
\frac{\alpha}{2} = \int_{\underline{\gamma}_1}^{\infty} f_{\gamma_1}\left(\hat{\gamma}_1 \middle| \hat{\gamma}_{1'\neq 1}(X,Y); \hat{\beta}_{OLS}(X,Y); \hat{\Sigma}\left(\hat{\beta}_{OLS}, X, Y\right)\right) d\hat{\gamma}_1 \tag{A.10}
$$

$$
\frac{\alpha}{2} = \int_{0}^{\overline{\gamma}_1} f_{\gamma_1}\left(\hat{\gamma}_1 \middle| \hat{\gamma}_{1'\neq 1}(X,Y); \hat{\beta}_{OLS}(X,Y); \hat{\Sigma}\left(\hat{\beta}_{OLS}, X, Y\right)\right) d\hat{\gamma}_1 \tag{A.11}
$$

where an estimator which is made dependent upon (X,Y) indicate the point estimates of our estimators for the given data set (X,Y); $f_{\hat{\gamma}_1}$ is the conditional density of a lognormal distribution. $\left[\overline{\gamma}_1, \underline{\gamma}_1\right]$ is then the $(1-\alpha)\%$ confidence interval. Again, we neglected the fact that the distribution is changed by the fact that we replaced the true variance-covariance matrix by a point estimate. It can be shown that this is identical to:

$$\left[\exp\left(\hat{\beta}_{OLS} - z_{\frac{\alpha}{2}} \cdot \text{vecdiag}\,\hat{\hat{\Sigma}}\left(\hat{\beta}_{OLS}\right)\right), \exp\left(\hat{\beta}_{OLS} + z_{\frac{\alpha}{2}} \cdot \text{vecdiag}\,\hat{\hat{\Sigma}}\left(\hat{\beta}_{OLS}\right)\right)\right] \qquad \text{(A.12)}$$

where $z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}\%$ critical value of the standard normal distribution and $\hat{\hat{\Sigma}}$ is the (estimated) conditional variance of the OLS-estimator. The use of the standard normal variate instead of the t-statistic indicates the error we make by acting as if the variance is known. For the other estimators we use a similar approach by replacing $\hat{\beta}_{OLS}$ in the previous formula by $\ln\ \tilde{\gamma}$ or $\ln \tilde{\tilde{\gamma}}$. The error we make is again more severe, since beside the error we make by not correcting for the use of estimates for the variance, we depart from the 'wrong' distribution since neither $\tilde{\gamma}$ nor $\tilde{\tilde{\gamma}}$ are lognormally distributed.

Comparison of the estimates of the two approaches for estimating the variances of the corrected estimators reveals that the second approach gives 'reasonable' approaches (the values are not totally different from the delta-method which is consistent and their difference with the delta method lies in the same direction as the difference between the two approaches for the ML estimator). This indicates that also the construction of confidence intervals for the corrected estimators are not unreasonable. A study of the exact properties of our *variance* estimators is postponed for future research.

Annex 2
Detailed estimates for conversion factors and prices of pooled data set

Table A.1 Conversion Factors for teff

village	estimates						community survey	
	Kunna			Tassa			Kunna	Tassa
	ML	corr1	corr2	ML	corr1	corr2		
Dinki	7.7966814	7.7192239	7.7192207	0.7105709	0.7048751	0.7048749	5*	1
variance 1	1.2138616	1.1900625	1.1899117	0.0081271	0.0079984	0.0079976		
variance 2	1.2507913	1.2260623	1.2260613	0.0083258	0.0081928	0.0081928		
confidence interval	(6.444,9.433)	(6.380,9.339)	(6.380,9.339)	(.621,.813)	(.616,.807)	(.616,.807)		
D. Birhan	7.4190437	7.3728123	7.3728111	0.9640987	0.9252603	0.9252538	5*	1
variance 1	0.6881327	0.6796547	0.679601	0.0764383	0.0704524	0.0704149		
variance 2	0.7011635	0.6924522	0.692452	0.086498	0.0796693	0.0796682		
confidence interval	(6.052,9.095)	(6.014,9.038)	(6.014,9.038)	(.562,1.652)	(.540,1.586)	(.540,1.586)		
S. Godeti	6.185506	6.153489	6.1534883	pooled with	Dinki		7	
variance 1	0.3971119	0.3930458	0.39302					
variance 2	0.4033446	0.3991799	0.3991798					
confidence interval	(5.294,7.227)	(5.267,7.189)	(5.267,7.189)					
Korodegaga	7.3029674	6.9004981	6.9004057	0.4057204	0.383361	0.3833559	5	0.6
variance 1	6.0466308	5.4036744	5.3996723	0.0186624	0.016678	0.0166657		
variance 2	7.1713809	6.4027264	6.4025548	0.0221339	0.0197615	0.019761		
confidence interval	(4.261,12.52)	(4.026,11.83)	(4.026,11.83)	(.237,.695)	(.224,.657)	(.224,.657)		
Imdibir and Adado	--	--	--	0.4159153	0.4084387	0.4084381		0.75
variance 1	--	--	--	0.0062759	0.0060541	0.0060527		
variance 2	--	--	--	0.0066272	0.0063911	0.0063911		
confidence interval	--	--	--	(.334,.518)	(.328,.509)	(.328,.509)		

Note:
ML is Maximum Likelihood estimate, corresponding to (A.1);
corr1 is Goldberger correction 1, i.e. using (A.3);
corr2 is Goldberger correction 2, i.e. using (A.4);
variance 1 is based on (A.6) to (A.8);
variance 2 is based on (A.9);
confidence interval is 95 percent unidimensional confidence interval (using (A.12))

Table A.2 Conversion Factors for maize

village	estimates						community survey	
	Kunna			Tassa			Kunna	Tassa
	ML	corr1	corr2	ML	corr1	corr2		
Dinki	--	--	--	0.9140968	.8927756	0.8927749		1
variance 1	--	--	--	0.0394411	0.0376275	0.0376238		
variance 2	--	--	--	0.0423388	0.0403867	0.0403867		
confidence interval	--	--	--	(.701,1.193)	(.684,1.165)	(.684,1.165)		
Sirbana Godeti	6.3949549	6.2257367	6.2257305	--	--	--	7	.5-1
variance 1	2.1934358	2.0792004	2.0789633	--	--	--		
variance 2	2.3774818	2.2533245	2.25332	--	--	--		
confidence interval	(4.390,9.315)	(4.274,9.068)	(4.274,9.068)	--	--	--		
Kordegaga	5.669962	5.5585519	5.5585488	.5296905	.5200503	.5200501	5	.55
variance 1	1.2759597	1.2264448	1.2263418	0.0103067	.0099359	.0099352		
variance 2	1.3543189	1.3016192	1.3016178	0.0108911	.0104983	.0104983		
confidence interval	(4.698,6.843)	(4.606,6.709)	(4.606,6.709)	(.458,.612)	(.450,.601)	(.450,.601)		
Adado	--	--	--	0.386732	0.3663977	0.3663963		.75
variance 1	--	--	--	0.0161564	0.0145064	0.014503		
variance 2	--	--	--	0.0190076	0.0170613	0.0170612		
confidence interval	--	--	--	(.359,.416)	(.340,.395)	(.340,.395)		

Note:
ML is Maximum Likelihood estimate, corresponding to (A.1);
corr1 is Goldberger correction 1, i.e. using (A.3);
corr2 is Goldberger correction 2, i.e. using (A.4);
variance 1 is based on (A.6) to (A.8);
variance 2 is based on (A.9);
confidence interval is 95 percent unidimensional confidence interval (using (A.12))

Table A.3 Conversion Factors for wheat

village	estimates						community survey	
	Kunna			Tassa			Kunna	Tassa
	ML	corr1	corr2	ML	corr1	corr2	5	1
Dinki	7.2115385	6.7501336	6.75000835	0.9029334	0.8663802	0.8663777		
variance 1	6.87773123	6.0281391	6.0260136	0.0673836	0.0620558	0.0620423		
variance 2	8.3923156	7.3527652	7.3526561	0.0762991	0.0702466	0.0702462		
confidence interval	(5.050,10.30)	(4.727,9.640)	(4.726,9.640)	(.702,1.162)	(.673,1.115)	(.673,1.115)		
D. Berhan	6.0648194	6.0329063	6.032906	0.993957	0.9825839	0.9825837	5	1
variance 1	0.3881162	0.3840562	0.3840458	0.0227391	0.0222235	0.0222222		
variance 2	0.3943099	0.3901711	0.390171	0.0235384	0.0230028	0.0230028		
confidence interval	(5.013,7.337)	(4.987,7.299)	(4.987,7.299)	(.743,1.330)	(.735,1.314)	(0.735,1.314)		
Sirbena Godeti	5.1775701	5.1151865	5.1151852	--	--	--	7*	
variance 1	0.6499146	0.6344	0.6343603	--	--	--		
variance 2	0.6740009	0.6578569	0.657857	--	--	--		
confidence interval	(4.215,6.360)	(4.164,6.284)	(4.164,6.284)	--	--	--		
Korodegaga	9.3837823	8.9991522	8.9991254	0.5967638	0.5850568	0.5850564	5*	.55
variance 1	7.3706826	6.7807681	6.7792778	0.0141116	0.0135652	0.0135638		
variance 2	8.3591761	7.6879547	7.687909	0.0149767	0.0143949	0.0143949		
confidence interval	(5.669,15.53)	(5.437,14.90)	(5.437,14.90)	(.513,.695)	(.503,.681)	(.503,.681)		
Adado	7.0710678	6.8411203	6.8411076	0.7825423	0.7633777	0.7633769	6	.75
variance 1	3.3060004	3.0951748	3.0946405	0.0303676	0.0289032	0.0288995		
variance 2	3.6513658	3.4177461	3.4177334	0.0327159	0.0311331	0.0311331		
confidence interval	(4.951,10.10)	(4.790,9.770)	(4.790,9.770)	(.667,.918)	(.651,.895)	(.651,.895)		

Note:
ML is Maximum Likelihood estimate, corresponding to (A.1);
corr1 is Goldberger correction 1, i.e. using (A.3);
corr2 is Goldberger correction 2, i.e. using (A.4);
variance 1 is based on (A.6) to (A.8);
variance 2 is based on (A.9);
confidence interval is 95 percent unidimensional confidence interval (using (A.12))

Table A.4 Prices for teff, maize and wheat

	Teff	maize	wheat		teff	maize	wheat
village	maximum likelihood			village	maximum likelihood		
Dinki	2.3739472	1.0883611	1.7333333	D. Berhan	1.5662296	1.2260843	1.4070374
variance 1	(0.09362)	(0.04424)	(0.19865)	variance 1	(0.01632)	(0.04795)	(0.00456)
variance 2	(0.09598)	(0.04679)	(0.21941)	variance 2	(0.01649)	(0.05030)	(0.00457)
interval	(2.00,2.81)	(0.85,1.40)	(1.41,2.13)	interval	(1.35,1.82)	(0.90,1.67)	(1.30,1.53)
	correction 1				correction 1		
	2.354311	1.0682268	1.6769662		1.5610273	1.2066845	1.4054183
variance 1	(0.09209)	(0.04262)	(0.18599)	variance 1	(0.01622)	(0.04645)	(0.00455)
variance 2	(0.09440)	(0.04507)	(0.20537)	variance 2	(0.01638)	(0.04872)	(0.00456)
interval	(1.99,2.79)	(0.83,1.37)	(1.37,2.06)	interval	(1.34,1.81)	(0.89,1.64)	(1.30,1.53)
	correction 2				correction 2		
	2.354310	1.0682263	1.6769631		1.5610272	1.206684	1.4054183
variance 1	(0.09208)	(0.04262)	(0.18595)	variance 1	(0.01622)	(0.04645)	(0.00455)
variance 2	(0.09440)	(0.04507)	(0.20537)	variance 2	(0.01638)	(0.04872)	(0.00456)
interval	(1.99,2.79)	(0.83,1.37)	(1.37,2.06)	interval	(1.34,1.81)	(0.89,1.64)	(1.30,1.53)
	community survey				community survey		
	2.00	1.00	1.76		1.50	--	1.50
	mean unit values				mean unit values		
	2.48	1.17	1.87		1.81	1.18	1.58
village	maximum likelihood			village	maximum likelihood		
S. Godeti	1.7749717	0.9791123	1.1010777	Korodegaga	2.1908902	1.4495593	1.278802
variance 1	(0.02545)	(0.01866)	(0.01336)	variance 1	(0.18140)	(0.06190)	(0.02876)
variance 2	(0.02576)	(0.01921)	(0.01358)	variance 2	(0.19199)	(0.06470)	(0.02953)
interval	(1.54,2.04)	(.76,1.26)	(0.95,1.28)	interval	(1.67,2.87)	(1.29,1.64)	(1.07,1.53)
	correction 1				correction 1		
	1.7678184	0.9696293	1.0950274		2.1498805	1.4283636	1.2676074
variance 1	(0.02524)	(0.01830)	(0.01321)	variance 1	(0.17473)	(0.06011)	(0.02826)
variance 2	(0.02555)	(0.01884)	(0.01343)	variance 2	(0.18487)	(0.06282)	(0.02901)
interval	(1.54,2.03)	(.75,1.25)	(0.94,1.27)	interval	(1.64,2.81)	(1.27,1.61)	(1.06,1.51)
	correction 2				correction 2		
	1.7678183	0.9696292	1.0950273		2.1498773	1.4283631	1.2676072
variance 1	(0.02524)	(0.01830)	(0.01321)	variance 1	(0.17468)	(0.06011)	(0.02826)
variance 2	(0.02555)	(0.01884)	(0.01343)	variance 2	(0.18487)	(0.06282)	(0.02901)
interval	(1.54,2.03)	(0.75,1.25)	(0.94,1.27)	interval	(1.64,2.81)	(1.27,1.61)	(1.06,1.51)
	community survey				community survey		
	1.45*	0.98	1.42*		2.20	1.60	1.33
	mean unit values				mean unit values		
	1.90	1.25	1.09		2.27	2.15*	1.73*

village	maximum likelihood		
Adado	2.7306203	1.4762985	1.4142136
variance 1	(0.29306)	(0.22169)	(0.06612)
variance 2	(0.31087)	(0.25835)	(0.06948)
interval	(2.09,3.57)	(1.16,1.87)	(1.18,1.69)
	correction 1		
	2.6774832	1.4030914	1.3910288
variance 1	(0.28185)	(0.20031)	(0.0640)
variance 2	(0.29889)	(0.23336)	(0.06723)
interval	(2.05,3.51)	(1.11,1.78)	(1.16,1.66)
	correction 2		
	2.6774789	1.4030864	1.3910281
variance 1	(0.28178)	(0.20027)	(0.06397)
variance 2	(0.29889)	(0.23336)	(0.06723)
interval	(2.05,3.51)	(1.11,1.78)	(1.16,1.66)
	community survey		
	--	--	--
	mean unit values		
	1.83*	0.93*	1.54

Note: Maximum likelihood is the Maximum Likelihood estimate, corresponding to (A.1);
correction 1 is Goldberger correction 1, i.e. using (A.3);
correction 2 is Goldberger correction 2, i.e. using (A.4);
variance 1 is based on (A.6) to (A.8);
variance 2 is based on (A.9);
interval is 95 percent unidimensional confidence interval (using (A.12));
community survey is estimate based on community survey
mean unit values are calculated mean unit values in the survey data, using community survey conversion rates;
*= value lies outside 95 percent confidence interval.