DISCUSSION PAPER

PRICES, LOCAL MEASUREMENT UNITS AND
SUBSISTENCE CONSUMPTION IN RURAL
SURVEYS: AN ECONOMETRIC APPROACH WITH
AN APPLICATION TO ETHIOPIA

by

Bart CAPÉAU
Stefan DERCON

Development Economics

Katholieke Universiteit Leuven
Departement Economie
Naamsestraat 69
B-3000 Leuven

Prices, unit values and local measurement units in rural surveys: an econometric approach with an application to poverty measurement in Ethiopia

Bart Capéau* and Stefan Dercon[+]

*Center for Economic Studies
Katholieke Universiteit Leuven
and
[+]Department of Economics
Oxford University

corresponding author:
Stefan Dercon
Department of Economics
Manor Road
Oxford OX1 3UQ
United Kingdom
e-mail: stefan.dercon@economics.ox.ac.uk

Revised Version September 2004

Abstract

For many research problems in developing countries, some information on prices faced by households is required for the analysis, for example if subsistence consumption is a substantial part of consumption. These prices are not readily available from household surveys, nor is it straightforward to observe them. Furthermore, quantities consumed and produced are often in local units presenting further problems for the analysis. 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 exogenous 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. We find that mean unit values result in the overestimation of prices due to outliers and other sources of measurement error. Exogenous consumer price sources, often collected at larger markets outside the village, tend to be slightly lower than our estimates.

1.   Introduction

' "The village down below is not very rich, is it?"
"Yes and no," Remigio answered. "Some prebenders live there, abbey dependents, and they share our wealth in the good years. For example, on Saint John's Day they received twelve bushels of malt, a horse, seven oxen, a bull, four heifers, five calves, twenty sheep, fifteen pigs, fifty chickens, and seventeen hives. Also twenty smoked pigs, twenty-seven tubs of lard, half a measure of honey, three measures of soap, a fishnet ..."
"I understand, I understand," William interrupted him. "But you must admit that this tells me nothing of the situation of the village, how many among its inhabitants have prebends, and how much land those who are not prebendaries possess to cultivate on their own. ..."
"Oh, as far as that goes," Remigio said, "a normal family down there has as much as fifty tablets of land."
"How much is a tablet?"
"Four square trabucchi, of course."
"Square trabucchi? How much are they?"
"Thirty-six square feet is a square trabucco. Or, if you prefer, eight hundred linear trabucchi make a Piedmont mile. And calculate that a family -in the lands to the north- can cultivate olives for at least half a sack of oil."
"Half a sack?"
"Yes, one sack makes five emine, and one emina makes eight cups."
"I see" my master said, disheartened. "Every locality has its own measures. Do you measure wine, for example, by the tankard?"
"Or by the rubbio. Six rubbie make one brenta, and eight brente, a keg. If you like, one rubbio is six pints from two tankards."
"I believe my ideas are clear now," William said resigned.
"Do you wish to know anything else?" Remigio asked, with a tone that to me seemed defiant.'

Umberto Eco, *The name of the Rose*, 1984 (1983-1980), Picador (Secker & Warburg Ltd.-Gruppo Editoriale Fabri-Pompiani), London, p.269.

There is an increased availability of large data sets collected in developing countries and large data collection programs, such as the Living Standards Measurement Study (LSMS) and similar surveys. They have provided an impetus to micro-econometric research on rural and urban households, using state of the art techniques (Deaton, 1997). Nevertheless, the first problem the Williams of today are confronted with is how to convert locally measured information into standardized measures. In rich societies, standardization has long been introduced, so that this may not seem a relevant issue. However, in many developing countries, not least in rural Africa, the use of standardized units is still far from universal. Similarly, data available to economic historians are often in localized units. Economic analysis typically requires measurements in monetary units, and so it does not suffice to convert local units into metric ones. The researcher also has to find the 'right' price. The

assumption of geographically perfectly integrated markets with a well defined 'one-price' is a strong one and is often not valid in developing countries (Barrett, 1996).

The problem may seem to have an easy solution: one should collect local prices and conversion factors, and use these to convert quantities recorded in local units into metric units and value quantities if required. Indeed, nothing in this article will fundamentally reject this approach. Such an approach allows for the construction of a data set with one observation for each price, at a certain point in time, on a local, neighbouring market (e.g. Van de Walle, 1988 for Indonesia; Glewwe, 1990). Still, the choice of the relevant market or issues of timing of data collection create their own problems with measurement.[1] With respect to the conversion factors, questionnaires often ask to convert the data to metric units on the spot. This puts the burden of conversion on the enumerator or even the household, which may present problems if these are not units usually used by these agents.[2] Alternatively, a survey at the community or market level about these local units of measurement could result in the appropriate conversion factors.[3]

Overall this would seem to be the best solution, even though it is not unproblematic.. For example, at least with respect to prices, the World Bank's Living Standard Measurement Survey website reports that most surveys have separate community price questionnaires (see also Glewwe and Grosh, 1995). But closer inspection of the available LSMS data reveals many surveys that do not have them, including some of the early rounds of the Côte d'Ivoire or Ghana surveys. The problems with prices in the Côte d'Ivoire survey are well documented and gave rise to conflicting views about what was happening in this period (see Grootaert and Kanbur, 1994; Deaton, 1997). Similarly, a survey in 1989 of six Ethiopian villages (see Webb et al., 1994) did not include a separate price survey. When the Ethiopian Rural (panel) Household Survey, was launched in 1994, the sample from these six villages was included in the panel, but the comparison over time was at first made difficult since there was no

---

[1]For example, Deaton, (1997, p.37), discusses the problems experienced in the LSMS to collect price data at the community level. This further points to the relevance of alternative approaches, such as the econometric approach proposed in this paper.

[2]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.

[3]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.

comparable price survey in 1989 to the one collected in 1994 (Dercon and Krishnan, 2003). The choice typically presented to the researcher is either to find a way of constructing prices from the existing data or ignore the data set. However careful the household data may have been collected, many otherwise useful data sets cannot be used unless a way of valuing quantities is found.

Most researchers use ad-hoc methods to construct prices from registered expenditures and quantities (if necessary the latter are first converted into metric units) in order to convert non-purchased consumption into monetary values. In this paper, we propose a *statistical* method to obtain market prices and conversion factors into metric units from observations on household expenditures in monetary units and in physical, though locally measured units. We apply the method to household consumption panel data, collected in Ethiopia in 1994-95. This statistical approach might overcome some difficulties to alternative ways to obtain price information and conversion factors (for a discussion of such alternatives and their problems, see Levin, 1991). The present approach allows for a joint determination of prices and conversion factors, using only data available within a household survey, in a context where units are not standardized. The approach is useful in contexts were there is no or questionable data on conversion units and prices from other sources, either as a means of checking this information or to allow further analysis to take place. The main advantage over alternative methods, such as calculating medians or means, is that we use more than one moment of the distribution while being *explicit* about the nature of measurement errors.

In this data set, the problems are real: the quantities were expressed in 70 different units, of which only two were metric quantity or volume measures (kilogrammes and litres). Our application to Ethiopian data shows that by using this approach, we find that exogenous price surveys coinciding with the household survey appear to result in slightly lower price estimates, probably related to collecting consumer prices at the nearest urban centre with high availability, rather than at the local retail markets. Mean unit values (dividing registered expenditures by registered quantities and averaging) would have resulted in an overestimation of prices, linked to outliers and other sources of measurement error. This could lead to biased estimates of poverty and consumption levels, and wrong identification of the poor.

Ideally, our approach should take into account two major criticisms of earlier attempts to retrieve prices from information in the data. An easy way to 'estimate' prices from the data, in the absence of the conversion problem, is simply to divide expenditures by quantities –

unit values.[4] 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 that purchase particular goods using different measurements units. Mean (or median) prices per unit are compared across the different units to obtain the relative quantity conversion factor. Ultimately, only limited information on the distribution of unit values and conversion factors are used in this case.

Using unit values has been strongly criticized in the literature. In his seminal work, Deaton (1987, 1988, 1990, 1997) has criticized the previous approaches to obtain prices for two reasons. First, 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. Secondly, measurement error in the unit values creates another source of potential bias. 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.

Our approach is partly inspired by this work – but has to face up to a particular difficulty: not only are prices unknown in many datasets, 'true' localized conversion units are also not known. In a cross-section, this means that there are too many unknowns for Deaton's approach to be applicable. For this reason, we ignore the issue of endogenous quality and focus on the measurement error problem. In fact, a careful look at his empirical results for Côte d'Ivoire shows that the measurement error problem dominates the bias caused by quality choice, especially in rural areas.[5] Since we are dealing with an African rural data set as well, this may justify for us to concentrate in our approach mainly on the measurement error problem. We then confine ourselves to the question how to obtain more reliable prices and appropriate conversion factors for estimating monetary valuations of non-purchased consumption in Ethiopia, and do not undertake in the present article an attempt to estimate price elasticities. For much applied welfare analysis on developing countries, information on prices and conversion factors *per se* remains valuable, even if, within the survey, subsistence

---

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

[5]For Côte d'Ivoire, for example, he finds that the quality elasticities are not large, and in the rural sector they are not significantly different from zero. He argues that there is a lower availability of different quality varieties of the same good to choose from in rural markets than in cities. (Deaton, 1987, p.23).

consumption was registered.[6] For example, prices are needed to construct cost of living deflators for welfare comparisons over time or space. 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.[7] In fact, these disturbance terms are a hybrid of the inherent stochastic nature of the data and measurement errors. In this paper we will only consider a simple error structure for measurement error and we will only focus on 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.

Even if we assume that the problem of endogenous quality choice is absent on empirical grounds (see note 5), our estimates might be biased due to exogenous differences in quality choice. Panel data can provide evidence on the presence of this problem. Access to panel data over relatively short periods of time, where incomes do not change very much, but in which we have several observations on purchases and consumed quantities, allows to control for inter-individual differences in quality choice by a time-invariant household-specific effect. Fixed effect estimation of a specification with unknown quantity conversions and prices can then be considered to control for such quality differences. Comparing the cross-section and panel data estimations provide then a test of the importance of exogenous quality differences.

To illustrate the strength of the approach, we will use the information obtained 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. While this is a strong assumption in a rural context with thin or missing markets (Singh et al., 1986, Low, 1986), the results are quite telling. Basic poverty analysis is used to illustrate the consequences of different approaches.

---

[6]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 elasiticties without price data needed price data to implement the procedure.

[7]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.

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. We will refer to the latter as the 'community survey', which contains a 'price' and 'conversion survey'. The data allow therefore to compare 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 appendices.

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 expenditures on this commodity. The data are clustered by communities and prices are assumed to be constant per community. We start from a simple accounting identity: price per unit times the purchased quantity equals monetary expenditures. The possibility of quality differences is initially ignored. The commodity under consideration is therefore assumed to be homogeneous for all households. Local units are 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 kilogrammes depending on the actual commodity.[8] 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 kilogramme of that good in community j; $V^*(i)$ equals the amount spent on the good by household i and a_{jk} is the conversion rate or factor of local unit k into kilogrammes for the commodity considered in community j.[9] As far as respondents report values and quantities without error, the following identity should hold, assuming constant prices and conversion factors per community:

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

If quantities are measured in kilogrammes, 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. However, 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, Lambert and Magnac (1997) – one of the few articles that explicitly deals with the issue of estimating both prices and conversions - define $p(i)_{jk}$ as the commodity's price paid by household i in community j, if the quantity purchased is measured in unit k, so that for all k and j, equation (1) can be rewritten as:

---

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

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

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

One can then replace $V^*(i)$ and $q^*(i)_k$ by their *observed counterparts $V(i)$ and $q(i)_k$* , in order to obtain an estimate of $p(i)_{jk}$. 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:

$$a^*_{jk} = \frac{p^*_{jk}}{p^*_j}. \tag{3}$$

While easy to implement, such two step procedure involves inefficiency, since only the mean of the distribution is used for the estimation. 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). For this, we assume that the reported expenditures $V(i)$ and quantities $q(i)_k$ are random and independent guesses by respondents.[10] Therefore, the empirical counterpart to identity (1) is a stochastic equation of the following type:

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

To illustrate our interpretation: if the observed values are $V(i) = V^*(i) + v(i)$, *and $q(i)_k = q^*(i)_k +$ w(i)*, where *both v(i)* and *w(i)* are independently and normally distributed variables with zero

---

[10]$V(i)$ 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.

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 it is (j,k)-cluster specific. This reveals a form of *natural tendency towards heteroskedasticity* with this type of data. There may be other reasons to transform identity (1) into its stochastic counterpart (4), such as the lack of very exact local measurement units (some of these are simply cups or cans). Quality differences could break down identity (1) as well: as far as these quality differences are truly random phenomena, not correlated with other sources of departure from identity (1), they could suggest specification (4). We will stick however to the measurement error interpretation in the sequel.[11]

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

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

in which $u(i)$ (from now on) denotes a normally distributed random variable with zero mean and identical variance across observations. It can easily be seen that the mean of the ratio of values and quantities per community is an unbiased linear estimator for p_j. However, this is not an efficient estimator. The proposed estimator is an OLS-estimator of p_j for the model $\dfrac{V(i)}{q(i)} = p_j + \dfrac{u(i)}{q(i)}$, which is still heteroskedastic. 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). Since we want to allow for measurement error in reported quantities too, and since quantities are also measured in other units than kilogrammes, this simple example has no practical relevance for our purpose.

---

[11] Notice that an empirical counterpart to identity (1) could be maintained if respondents made first a guess about the quantities, $q(i)_k$, and then calculate $V(i)$ as $p_j\, a_{jk}\, q(i)_k$. No specific estimation techniques would be needed as the price and conversion factor could be calculated from the identity: $V(i) = p_j\, a_{jk}\, q(i)_k$. However, it could be easily verified that the latter model is not satisfied by the data. If respondents do not remember prices either and make a guess,

By using an econometric approach to prices and conversion factors, the problem caused by heteroskedasticity and the inefficiency of estimation can be overcome. 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 larger, when the amount bought is larger. 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:

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

where $e^{u(i)}$ is a composite of independent lognormally distributed measurement errors on quantities and expenditures.

The variance of $V(i)$ conditional on $q(i)_k$ is now dependent upon the amount consumed. In Capéau (1995) the disturbance hypotheses in (4) and (6) are tested against each other. It turns out that the specification in (6) performs 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:

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

different procedures than the ones proposed here might be necessary.

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:

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

Not all units are specific to the locality. Some observations are expressed in kilogrammes or litres. Also, in a few cases local units are expressed in standard units, common across a larger area: for example, a particular type of a small coffee cup, produced by a local factory, is commonly used to value items such as sugar of coffee beans. Exceptionally, there are common standards to *all* communities (as for example *quintal*, about 100kg). In estimating (8) we used this information by imposing restrictions on some units across communities. These restrictions make the econometric approach less self-evident. Suppose that all units were locally specific, then the approach described can be shown to be exactly equivalent to estimating a geometric mean and simple calculations could be done to obtain conversion factors. Once restrictions for some units exist across communities, this is not anymore the case since empirical information about covariances of the data across sites can be taken into account.

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. It would therefore be desirable to control for a quality difference in the price-unit value relationship. Let us define the price in community *j* at time *t* for a specific commodity of quality *v* as p^v_{jt}. Let us define v_i as a quality index of the commodity bought by household *i*. In case the interval between two registrations is not too large, the choice of quality by a household *i* can assumed to be fixed (preferences are assumed to be given and incomes are not expected to change drastically in short time intervals). Consequently p^v_{jt} is equal to $v_i p_{jt}$. The short run intertemporal counterpart to identity (1) is then equal to:

$$V^*(i)_t \equiv p_{jt} a_{jk} q^*(i)_{kt} v_i . \qquad (9)$$

In a cross-section data set, it then becomes impossible to identify the price by the approach

12

described above. If richer households buy higher quality goods, then the earlier estimation procedures are biased. If, for example, average wealth across communities differs, estimated price differentials can reflect differences in quality choices in stead of true price differentials. Further assumptions are needed to proceed. For example, one could consider that some households typically 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. It is plausible to assume that over relatively short periods of time, households do not change the quality of the commodities bought, for example, because wealth or permanent income does not change rapidly and if substitution effects are small for particular commodities, such as staples. On the basis of the same assumptions concerning measurement error, we can derive a short run intertemporal version of (7), on the base of (9), for those households for whom we observe market transactions in subsequent time periods t:

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

Here 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 over (i,t), with zero mean and constant variance. If we assume that a particular household buys a commodity of a fixed quality over short periods of 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.[12]

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

---

[12]The data requirements are however much higher: 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. The random effects model may be used, but this comes at a cost. Quality is a constant for each household in the fixed effects formulation. 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. Both approaches will be used in the next section.

incomes. Here we assume that households do not change the quality choice of the goods consumed over short periods of time, 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 then we could expect that the panel estimators for the prices to be close to the pooled estimates.[13] The proposed panel estimators allow for verifying whether individual heterogeneity might cause our resulting estimates to be biased because of systematic preference or wealth differences across communities. This might be the case if a model without individual specific effects is rejected in favour of a random or fixed effects model. If, as in Deaton's approach, quality is endogenous, then our test might be erroneously specified and can give wrong indications. In such a case an additional equation has to be added to the model, which requires a model of choice of quality. Such an undertaking is beyond the scope of the present article.[14]

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 two estimates of the variance for all three estimators, the first being simply a first order approximation, while the latter is based on the *exact* variance formula of the uncorrected parameter estimator. Since we work with non-normal distributions, these variances cannot be

---

[13]If the panel estimates are statistically different from zero, then this does not necessarily imply that quality matters. It is also possible that the household level fixed or random effect is actually measuring some systematic, household specific measurement error. The panel estimates would control for this.

[14] Disney, Kedir and McKay (2001) provide a methodology to correct simultaneously for measurement error and endogeneity bias while estimating price and income effects, on the base of model of endogenous choice of measurement unit.

used in a simple fashion to construct confidence intervals. In appendix 1, we explain how confidence intervals can be constructed.

3. Data and estimations

We use data 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.[15] Three rounds were used, and data were collected with about 6 months in between, resulting in three observations on consumption and other variables. 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 concerning about 80 different food items. Respondents had to report on purchases (total expenditures during the last week), the quantity purchased and the unit in which the latter was expressed. Then they were asked for the same commodity about consumption from own harvest or stock and again quantities and units were recorded. Finally, the same information was recorded about consumption from gifts, wages in kind, barter or loans in kind they received from 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 (kilogrammes and litres). Table 1 gives the distributions across some of the main units recorded in the survey with the frequencies involved.

---

[15]The approach remains valid for clustered data (which is the usual practice for collecting budget surveys in developing countries, because of budgetary reasons) as long as sufficiently large groups using similar factor conversions and/or facing the same price, are available. Using mean unit values in samples with many small clusters might also result in imprecise estimates, but the standard errors of this approach are usually not reported.

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

Unit	all	purchases	own stock	gifts
Kilogrammes	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 in table 1 are transcriptions of the Amharic names. Given the century-long dominance of Amharas in Ethiopia, Amharic is the lingua franca in much of Ethiopia.[16] 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: it can be the cup in which coffee is served in public houses. 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, at least for some units and if the data permit it.

In order to be able to estimate equation (4), more observations than coefficients to be estimated

---

[16]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, they were put under the same code. Since the conversion codes estimated for these

have to be available, of course. In addition to this, in every village there should be at least one household that registers purchased consumption in terms of the *numéraire* (usually kilogrammes or litres) or there should be at least one cross restriction to another village (a measurement unit for which the same conversion factor is imposed as in the other village and in which unit there are observations in both villages) in which a household registering purchased consumption in terms of the *numéraire*. For 58 commodities, sufficient data points and restrictions were available to estimate regressions to determine conversion rates and prices. In the end only a small fraction of all the quantities that needed to be converted and valued could not be estimated. The lacking conversion factors or prices applied to goods which were not abundantly consumed.[17] In that case, information from the community survey will be used in the poverty analysis below. To illustrate the procedure, we report just one regression, for 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 registered quantities bought by the household. 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 kilogrammes prevented us to use such interaction dummies. Sometimes, community specific conversion factors are not desirable and a single dummy across communities for these units suffices. Table 2 presents this regression.

---

units are usually cluster-specific, the translation problem would not affect the analysis.

[17]In 19 cases, no observations on kilogrammes 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 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 kilogrammes.

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 kilogramme 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 exponents of estimated parameters results in an *over*estimation of both prices and conversion rates. Therefore we report corrected estimates for teff, maize and wheat on the basis of similar regressions as the previous one - pooling the data of different rounds (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 appendix 1) and Golberger's 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 appendix 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. The latter are obtained by first converting local units into metric units using the conversion information from the community survey, and then calculating the mean unit value (reported expenditures divided by converted quantity).

Table 3 **Conversion factors in kilogrammes 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)
community survey		-	-	7.00	5.00	-	1.00	-	-	0.55	0.75*

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

villages		one kunna of wheat					one tassa of wheat				
		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 3 rounds 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 appendix 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 inconsistent with the estimates from the regressions, since they lie outside 95% confidence interval implied by the standard errors of the regressions.

Table 4 provides some data on prices, obtained from the same regressions as the estimates in table 3. These are compared with the estimates from the community level price survey, and mean unit values (values divided by quantities converted into metric units on the bases of the community survey information).

Table 4   **Estimated prices, calculated unit values and price survey data (per kilogramme), for teff, wheat and maize for selected villages.** Pooled regressions; reported 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 3 rounds of the ERHS. ML estimates are the coefficients 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.
*=community survey estimate falls outside 95 percent confidence interval

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. The data of only four of the fifteen villages were used. These four villages are situated in the Southern part of the country, where maize is an important crop both in production and in consumption.[18] Equation (9) was estimated as a pooled cross-section regression, as a 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.[19] 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 a direct comparison with the pooled cross-section and the random effects model, the latter had to be re-specified. The explanatory variables included to capture prices are now a constant, a base year price dummy for three of the four sites maintained in this analysis (the site Adele Keke serves as reference category), and price dummies for all the four sites for the two subsequent periods of observation. Consequently, the price dummies in the second and in the third round could be interpreted as changes in (the logarithm of) prices just as in the panel regression equations. The dummies for the first round are differences in the logs of prices relative to the reference site; the constant is the base year log of the price for Adele Keke. The reported coefficients are therefore not prices and conversion factors, but they could be used to calculate them.

---

[18]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 purchase the same commodities in each period, limiting the scope for panel data estimation techniques.

[19]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.

The results suggest that neither the fixed effects model nor the random effects model contribute significantly to the explanatory power of the pooled regression: 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 signficant 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 difference between the qualities of commodities purchased by different households over short periods of time.[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 exogenous quality difference problems.

---

[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.

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 conversion factors, 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.[21] 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. The approach allows us to see the problems inherent with using unit values and also provides a consistency check on using some other sources of information for prices.

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 (8) 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.

First, note the low mean consumption levels: only about $10-13 per month. Next, note that the regression approach suggests higher consumption than implied by either of the alternative approaches. 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

---

[21]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.

in the first round. However, since these estimates are in nominal terms, 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.[22] 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 Authority.

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 value in birr for 1994 was 36 birr.
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.
In applying these poverty lines for poverty measures, site specific poverty lines were used to reflect site-specific price differences. The overall poverty lines are sample weighted means of the regional lines.

The table shows that the poverty lines over time are sensitive to the source of price data used. Using unit values, the poverty lines show a small decline in the second half of 1994 and an increase afterwards. The poverty lines using the price survey and the regression-based prices

---

[22]The approach has also been used for Ethiopia in Dercon and Krishnan (1996). Essentially the approach for the minimum costs of basic food needs is followed (see also Ravallion and Bidani, 1995), 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 focus in this paper is on poverty incidence and intensity, the poverty lines are appropriate deflators for nominal consumption.

both suggest an inflationary trend from early 1994 until the middle of 1995. However, this pattern is different for the Food CPI and for the poverty line index on the base of rural prices registered by the Central Statistical Authority (CSA). These inflation estimates illustrate one of the typical puzzles when comparing household data over time. Official inflation figures show very low rural inflation. When applied to the mean consumption levels (and as we will show the poverty measures), this would have implied quite important increases in the standard of living in rural Ethiopia between 1994 and 1995. However, the community survey conducted simultaneously in the different sites included in 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 (column 5), suggest that this difference cannot be ascribed to higher inflation figures in rural regions as compared to urban price evolution: using only the rural prices as registered by the CSA resulted in an increase of the poverty line of only 7 percent during the same period, which is again less than the 13 percent price increase perceived when using information on prices in the community surveys. The econometric approach to obtain prices provides an alternative means of assessing the changes in prices over the period. In the present case, it shows that the intertemporal pattern implied by the price information from the community survey roughly coincides with the survey estimations. It indicates that linking secondary price data from national sources to household surveys with specific sample design might be problematic.

Table 8 illustrates the consequences of using the different sources of price data for poverty measures. We report the head count and the poverty gap index for each of the three rounds using four possible poverty lines constructed respectively on the base of the community survey data on prices, the unit values, the regression-based estimates and the implied price changes from the food CPI. We also give mean consumption per capita in 1994 prices. T-tests for the difference in the estimates are in brackets: these compare each estimate with the estimated value for 1994. 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	using	using mean unit

			survey data		regressions		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.

What can we conclude from this analysis? First, the weakness of using mean unit values is illustrated by this example. Poverty would be estimated to be higher in each period than using the community survey approach and its fluctuations are higher as well. This is entirely due to the exacerbating impact of measurement error in the unit values. The unit value price data are typically higher than the prices from the community survey, especially for the products included in the poverty line food basket. For example, the 1994 value of the food basket used in the poverty line is 45 birr per month, compared to 37 birr when using prices from the community survey. This stems from the occurrence of true outliers in the right tail of the distribution of registered expenditures. Overall, some unit values must be lower than the community survey data on prices, since 1994 mean consumption is roughly the same using both methods. The estimated value of mean consumption in 1994 using the regression technique is higher than for all the other methods. Apparently, especially non-poor respondents consume products with higher estimated prices than the unit value or with higher estimated conversion factors than those registered in the community survey.

Secondly, exogenous price data sources may not give very different poverty estimates in this example, compared to those based on the regressions, but mean levels of real consumption are systematically lower than those implied by using only information from within the survey, when controlling for measurement error. Prices appear to be higher on average when based on

information from within the survey data, as compared to the exogenous data. This would be consistent with exogenous price information to be often measured *in practice* at nearby, maybe urban, markets with plenty of availability and regularity, making price collection easier. The actual retail price paid in the villages of the sample may be higher. Using 'exogenous' price surveys would then result in an underestimation of the value of subsistence consumption in some communities. That mean consumption is higher and that the poverty gap is lowest when using within survey regression results is related to this, though this might also be partly due to differences between estimated and community survey conversion factors. These are not the only implications: despite the fact that poverty levels using the head count may not seem to be affected by using the exogenous prices relative to the within-survey prices (with more or less the same poverty levels), different households are identified to be poor across the sample. About 7 percent of households in the survey would be 'differently' classified as poor or non-poor when using the exogenous local price survey relative to the within-survey estimation of prices.

5.   Conclusions

This paper addresses one of the important, yet in practice often ignored problems faced by applied economists when analyzing demand and consumption, namely that of constructing money metric values and prices from information on expenditure and quantities. In this paper, we focus on contexts where quantity metric conversion units are not standard. The paper develops a systematic approach to obtain prices and conversion factors, for example to value non-marketed production in semi-subsistence economies. Focusing on the measurement problem, we derive a simple approach to estimate prices and conversion factors from regressions, using the observations on expenditures on consumption and the amounts purchased in local units. Using hypotheses on the measurement error distribution, it is possible to correct for the bias involved in the usual ad-hoc procedures to obtain prices and to convert non-marketed consumption in values and quantities. The proposed estimation approach allows us to exploit information on cross-community restrictions; without these restrictions, the estimator reduces to a geometric sample mean. By using panel data over a relatively short time period, we find that the cross-section price variability cannot be ascribed to exogenous differences in quality. Although the paper focused on the consumption data, the approach could also be used for 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 shows some of the advantages of the approach described in this paper. Unit values may suffer from outliers and other forms of measurement error. Exogenous price surveys may also not be error free, because prices may be collected at markets not directly relevant for the households in the survey. The results also show that standard unit value approaches appear to have overestimated poverty by overestimating prices.

What does this suggest for the practice of data collection? First, the common sense approach remains valid: to collect prices relevant for the sampled households in local markets, while trying to find conversion factors for any non-standard local measurement units in the field. It is bound to be most cost effective. Secondly, if for some reason the data available to the researcher do not contain these conversion factors and prices, or there are reasons to suspect the quality of them, our approach can provide either a useful check or an alternative, superior to using the simple mean or median unit value for prices and quantities. Thirdly, each method gives quite different results, so the dangers of mixing different approaches while handling the data are substantial. For example, using our approach to fill in prices for, say, one region where prices had not been collected, while using price surveys for the rest of the country, may result in a systematic bias, so the advice would be to use the alternative approach for the entire country, at least as a check.

References

Barrett, C. B. (1996), "Market Analysis Methods: Are Our Enriched Toolkits Well-Suited to Enlivened Markets?", *American Journal of Agricultural Economics*, 78, pp. 825-829, August.

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 Universtity 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.

Dercon, S. and P.Krishnan (2003) "Changes in poverty in villages in rural Ethiopia: 1989-95", in A.Booth and P.Mosley, (eds.) 2003, *The New Poverty Strategies* , Palgrave MacMillan, London.

Disney, R., A.M.Kedir and A.McKay, (2001), "Local measures, quality effects and the estimation of demand elasticities in urban Ethiopia", University of Nottingham, mimeo.

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.

Glewwe, P. and M.Grosh (1995), "A Guide to Living Standards Surveys and Their Data Sets", LSMS Working Paper #120, The World Bank, 1995

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

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

Grootaert, C., and R.Kanbur, (1994), "A New Regional Price Index for Côte d'Ivoire Using Data from the International Comparisons Project*", Journal of African Economies*, vol.3, no.1.

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, 3th 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.

Strauss, J., (1982), "Determinants of food consumption in Sierra Leone. Application of the quadratic expenditure system to the consumption-leisure component of a household-firm model", *Journal of Development Economics*, 11, 327-353.

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

Webb, P., J. Von Braun and Y.Yohannes, (1992), *Famine in Ethiopia: Policy Implications of Coping Failure at National and Household Levels*, International Food Policy Research Institute

Appendix 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:

Δ = the matrix of dummies serving as explaining variables in equation (8) of the main text; β = the vector of unknown parameters to be estimated in equation (8). It are the logarithms of prices and conversion rates; γ = 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 $\hat{\beta}_{OLS}$ be the OLS-estimator for β in equation (8). Then:

$$\hat{\gamma} = exp(\hat{\beta}_{OLS}). \qquad (A.1)$$

It is a well known result that maximum likelihood estimators are invariant for reparameterisation 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 $\hat{\beta}_{OLS}$ is known to be normally distributed according to $N(\beta, \Sigma(\hat{\beta}_{OLS}))$, 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 $LN\left(\beta, \Sigma\left(\hat{\beta}_{OLS}\right)\right)$.[23] According to a well-known property of the lognormal distribution it follows that:

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

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

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

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

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

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:

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

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.

---

[23] 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.

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:

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

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

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

where $diag \ a$ converts a vector a into a diagonal matrix with the elements of a on the diagonal; $D \equiv \left(vecdiag(\Delta'\Delta)^{-1}\right) \cdot \left(vecdiag(\Delta'\Delta)^{-1}\right)'$; $\hat{\sigma}^4$ is the square of the usual estimator for the variance of the OLS-regression; $\hat{\Omega} = \left[\frac{\partial F}{\partial vecdiag(\hat{\Sigma})}\right] \cdot \left[\frac{2\hat{\sigma}^4}{v} D\right] \cdot \left[\frac{\partial F}{\partial vecdiag(\hat{\Sigma})}\right]$. [24]

Since $\hat{\gamma}$ is lognormally distributed according to $LN\left(\beta, \Sigma(\hat{\beta}_{OLS})\right)$, 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:

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

---

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

in which J is a square matrix with all elements equal to 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 $\underline{\gamma}_1$:

$$\frac{\alpha}{2} = \int_{\underline{\gamma}_l}^{\infty} f_{\gamma_l}\left(\hat{\gamma}_l|\hat{\gamma}_{l'\neq l}(X,Y);\hat{\beta}_{OLS}(X,Y);\hat{\Sigma}\left(\hat{\beta}_{OLS},X,Y\right)\right)d\hat{\gamma}_l, \qquad (A.10)$$

$$\frac{\alpha}{2} = \int_{0}^{\overline{\gamma}_l} f_{\gamma_l}\left(\hat{\gamma}_l|\hat{\gamma}_{l'\neq l}(X,Y);\hat{\beta}_{OLS}(X,Y);\hat{\Sigma}\left(\hat{\beta}_{OLS},X,Y\right)\right)d\hat{\gamma}_l, \qquad (A.11)$$

where an estimator which is made dependent upon (X,Y) indicate the point estimates of our estimators for the given data set (X,Y); $f_{\hat{\gamma}_l}$ is the conditional density of a lognormal distribution. $\left[\overline{\gamma}_l,\underline{\gamma}_l\right]$ 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:

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

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 in stead 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.

Appendix 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	1.1899127	1.1899117	0.0081271	0.0079976	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	0.6796012	0.679601	0.0764383	0.0704159	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	0.3930201	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	5.3998169	5.3996723	0.0186624	0.0166661	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	0.0060527	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	0.0376238	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	2.0789674	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	1.2263431	1.2263418	0.0103067	.0099352	.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	0.0145032	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	6.0261031	6.0260136	0.0673836	0.0620427	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	0.3840458	0.3840458	0.0227391	0.0222222	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	0.6343607	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	6.7793181	6.7792778	0.0141116	0.0135638	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	3.0946519	3.0946405	0.0303676	0.0288996	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	(0.09208)	(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)
	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	(0.17469)	(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*

	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	(0.28178)	(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)
	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.