Measurement error and functional form
a proposal to estimate prices and conversion rates from the ERHS 1994

Bart Capéau[*]

---

[*] Thanks to Stefan Dercon, Geert Dhaene, Pramila Krishnan, Erik Schokkaert and Tesfayi G. Selassie.

1 Prices, quality, quantity, measurement error?

The raw data of the Ethiopian Rural Household Survey (ERHS) contain information on quantities of purchased consumption goods and sold producer goods as well as on the amount of money received or expended on these goods. The quantities are expressed in local measures. The same measurement unit can differ over regions and goods at once. Others remain quite constant for the whole country or for a large number of goods. Beside, there are at best some rough guesses about how to convert these local measures into kilograms available. The present note contains a proposal to retrieve alternative conversion rates from the data. They need to be estimated jointly with the (unknown) prices of the goods. Afterwards it may be worthwhile to compare these with the guesses in the existing conversion tables and rough information on prices[1]. Our proposal tries to tackle the problem of measurement errors, which surely are present in the data, as a matter of choice of functional form.

We propose to evaluate purchased goods at consumer prices. Sold goods on the contrary are producer goods and therefore we need to estimate producer prices too. Different rationales for distinguishing between producer and consumer prices can be found. The first one is peculiar to the ERHS-data: consumer and producer goods are not subdivided in a similar way, so that the price of commodities is dependent on the quantities of the different subgroups contained in it. Differences of consumer and producer prices might be due to the different composition of the good-basket of a particular group. Anyway, in a more general fashion this problem occurs in every data set containing information on quantities and monetary amounts, since goods always perform some quality differences. A second reason might be the presence of some possibly informal indirect tax system, though this surely is not the case in Ethiopia. If we would have no information at all about the latter, it might be convenient to let the data detect whether there is a gap. Thirdly, price differences could reflect the difference between retail trade and wholesale prices. Remark that it is always possible to test whether the estimated consumer and producer prices differ significantly. If not, a reestimation without distinction might improve the results. Finally one has to take into account that there is a time delay in the information concerning sold harvest (last harvest) and consumption (typically, last week). In a country with high inflation this naturally causes a price difference.

With regard to the argument concerning quality differences A. Deaton argues in a series of papers (1987, 1988, 1990) that one could only retrieve information on unit values (the value of a unit meat say) rather than prices of goods from data containing quantities and expenditures (or sales) only. He proposes then a method to estimate price elasticities without information on prices, taking into account measurement errors as well as the price-unit value difference. A careful look at his empirical results learns however that the measurement error problem is far greater than the bias caused by quality choice, especially in rural areas. "The quality elasticities are not large, and in the rural sector they are not significantly different from zero. Presumably there is a good deal less choice than there is in the cities." (Deaton, 1987, 23). In a country like Ethiopia, where repressed consumption surely is more of a problem than quality choice, we do not expect this to be otherwise. Moreover the stronger the disaggregation of goods the more the problem of quality choice tend to disappear of course. Therefore we will concentrate on the

---

[1] During the recollection of the data of the ERHS 1994 a price survey on some local markets was done. However the aggregation of quantities and qualities and the conversion of local measures there also was a problem. It may be worthwhile to compare present results with the independently collected price information. You could see it as a first empirical investigation into the validity of the frequently claimed necessity to recollect price information independently. Alternatively, one could retrieve this from the data, a practice which is usually dissuaded. A firmly grounded statistical test however does not seem possible here.

problem of measurement errors rather than bias due to quality.[2] Contrary to Deaton we won't propose a simple measurement in errors estimator but we try to tackle the natural tendency to heteroskedasticity by means of the choice of functional form of the error term.

2 Notation

We introduce the following basic notation:

$q(i)_k =$ quantity of good i purchased or sold, measured in unit k

$p(i)_c =$ consumer price of good i per kilogram

$p(i)_p =$ producer price of good i per kilogram

$w(i)_k =$ amount of money spended on or recieved for quantity $q(i)_k$ of good i measured in unit k

a_k = conversion rate of measurement unit k in to kilograms (or pieces)

Clearly $a_k \equiv 1$ whenever the measurement unit j equals kilograms (or pieces, when this unit is the more "natural" one for that particular good, as in the case of bananas for example). Another basic unit might be necessary whenever we don't have any observation on the currently used numéraire.

Per Peasant Association (PA) the data allow us to recollect all observations on quantities of a certain good i purchased or sold, as well as the associated amounts of money received from or expended to it. This gives us the following analytical equation, which defines exactly the relation between unknown prices and conversion rates on the one hand and "ideally measured" quantities and amounts of money on the other hand:

$$\overline{w}(i)_{tk} = (1-\delta_c)\cdot a_k \cdot \overline{q}(i)_{tk} \cdot p(i)_p + \delta_c \cdot a_k \cdot \overline{q}(i)_{tk} \cdot p(i)_c \qquad (1)$$

with $\delta_c = 1$ if the good is consumed (purchased) and 0 otherwise

and index t runs over the available observations

Bars denote the ideally measured quantities and monetary amounts. They have an exact relation with prices and conversion rates. Retrieving the latter would be a matter of pure calculation. Since we are working with only one particular good i, this index will be dropped from now on for simplicity[3]. Before going on, we rewrite equation (1) by means of a couple of definitions, which will allow us to write down things in a convenient matter. Define therefore:

---

[2] Moreover, we definitely will need prices for the joint estimation of demand and production in a later phase. The phenomenon of consuming self produced goods is mentioned in all of Deatons papers but is nowhere appropriately treated.

[3] Note that there may be a case for jointly estimating the conversion rate of different goods if there is a clear indication that these don't differ across different goods, which is not always the case. However nothing substantially will change w.r.t. the proposals in the present note so that we easily could perform the joint estimation if this was esteemed to be better. Also, estimating jointly conversion rates and prices for the same good over different PA's is only a matter of adding more dummies. Statistical tests whether prices as well as conversion rates differ over the PA's (or goods) are readily available.

$\mathcal{J} = \{$ consumer good, producer good $\} = \{c, p\}$ indexed by j

$\mathcal{K} = \{$ measurement units $\} = \{1, 2, \ldots, k, \ldots, K\}$

$\mathcal{T} = \{$ indices of observation points $\} = \{1, 2, \ldots, t, \ldots, T\}$

Then define the following mappings:

$\tau_1 : \mathcal{T} \to \mathcal{J} \times \mathcal{K} \quad : \quad t \to jk,$ which maps every observation unit onto the unit in which
it is measured and its adherence to the group of
consumer or producer goods

$\tau_2 : \mathcal{T} \to \mathcal{K} \quad : \quad t \to k,$ which maps every observation unit onto the unit
in which it is measured

$\tau_3 : \mathcal{T} \to \mathcal{J} \quad : \quad t \to j,$ which maps every observation unit onto its adherence to the group
of consumer or producer goods

$\phi : q_t \in \mathcal{R}_+ \to q_{jk(t)} \in \mathcal{R}_+^{|\mathcal{J}| \times |\mathcal{K}|} : \quad q_{jk(t)} = q_t \quad$ if $\tau_1(t) = jk$
$q_{jk(t)} = 0 \quad$ otherwise

These definitions make it possible to write down (1) as follows:

$$\overline{w}_t = \Sigma_{j \in \mathcal{J}} \Sigma_{k \in \mathcal{K}} \overline{q}_{jk(t)} a_k \cdot p_j \tag{1′}$$

3 Let the right relation please stand up

The data unfortunately contain a lot of measurement errors. The most straightforward procedure would then be to add the familiar disturbance structure to equation (1′) and estimating it by means of non-linear least squares.

$$w_t = \Sigma_{j \in \mathcal{J}} \Sigma_{k \in \mathcal{K}} q_{jk(t)} a_k \cdot p_j + u_t \tag{2}$$

with: u_t a distrubance term with the following properties: $E[u_t] = 0$ and $E[u_t^2] = \sigma^2$

where we use now the actual observations, to retrieve estimates from the data. In the sequel these estimates will be denoted by: $\hat{p}_j^{\mathrm{NLS}}$ and $\hat{a}_k^{\mathrm{NLS}}$. However this might not be an adequate estimation procedure, as the following example illustrates. Suppose that the measurement errors are additive for both monetary values and quantities:

$$w_t = \overline{w}_t + \varepsilon_t \qquad E[\varepsilon_t] = 0 \text{ and } E[\varepsilon_t^2] = \sigma_\varepsilon^2 \tag{3}$$

$$q_t = \overline{q}_t + \delta_t \qquad E[\delta_t] = 0 \text{ and } E[\delta_t^2] = \sigma_\delta^2 \tag{4}$$

$$E[\delta_t \varepsilon_t] = \sigma_{\varepsilon\delta}^2$$

In that case we could rewrite equation (2) as follows;

$$w_t = \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} q_{jk(t)} a_k \cdot p_j - \delta_t p_{j(t)} a_{k(t)} - \varepsilon_t \tag{5}$$

with: $-\delta_t p_{j(t)} a_{k(t)} - \varepsilon_t \equiv u_t$

Consequently:

$$E[u_t] = E[-\delta_t p_{j(t)} a_{k(t)} - \varepsilon_t] = 0$$

$$\begin{aligned} E[u_t^2] &= E[(-\delta_t p_{j(t)} a_{k(t)} - \varepsilon_t) \cdot (-\delta_t p_{j(t)} a_{k(t)} - \varepsilon_t)] \\ &= E[\delta_t^2 p_{j(t)}^2 a_{k(t)}^2 + 2\delta_t \varepsilon_t p_{j(t)} a_{k(t)} + \varepsilon_t^2] \\ &= \sigma_\delta^2 p_{j(t)}^2 a_{k(t)}^2 + 2\sigma_{\delta\varepsilon} p_{j(t)} a_{k(t)} + \sigma_\varepsilon^2 \end{aligned}$$

The estimated standard errors of the NLS-coefficients would therefore suffer from heteroskedasticity. Beside, since NLS minimizes the quadratic sum of the error terms, it does not take into account that we would like to filter out the measurement errors rather than minimizing them. This could cause biased estimates. If the problem were confined to the presence of heteroskedasticity, this could be quickly resolved for a special case.

Presume that there is no measurement error in the observed money amounts. Then our basic equation (2) could be written as:

$$w_t = \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} q_{jk(t)} a_k \cdot p_j - \delta_t p_{j(t)} a_{k(t)} \tag{6}$$

In that case we could easily try a non-linear version of weighted least squares, by minimizing $\sum_t \delta_t^2$. With regard to interpretation this error-structure assumes that measurement errors are higher in absolute value for goods measured in larger units or with higher prices (i.e. usually the consumer prices). A rationale might be that one is less inclined to use complicated decimal figures. Since w_t is assumed to be measured without error here, the same result could be obtained by estimating the following equation with NLS:

$$q_t = \frac{w_t}{a_{k(t)} \cdot p_{j(t)}} + \delta_t \tag{6′}$$

From the moment however that both variables are measured erroneously, a lack of further information on the disturbance structure, even when we assume that the covariance $\sigma_{\delta\varepsilon}$ between both errors is zero, would bar any further results. One very rough procedure would be to use the estimated prices $\hat{p}_j^{\text{NLS}}$ and reestimate our first equation for the conversion rates only now. This relation is now linear and could be estimated with the aid of instrumental variables. The only straightforward set of instruments here are the per measurement unit means of observed quantities (assuming overestimations and underestimations will cancel out). Thus we would then estimate the a_k's with the following estimator:

$$a = (Z'Q)Z'w \tag{IV}$$

where the Q-matrix contains the observed quantities multiplied by the NLS-estimates of their respective prices, i.e. the following block-diagonal matrix:

$$Q = \begin{pmatrix} \hat{p}^{NLS}_{j(t_1)} q_{jk(t_1)} & 0 & \dots & \dots \\ \cdot & 0 & 0 & \dots \\ 0 & \hat{p}^{NLS}_{j'(t_s)} q_{jk'(t_s)} & 0 & \dots \\ 0 & \cdot & 0 & \dots \\ 0 & 0 & \cdot & \dots \end{pmatrix}$$

and the instrument matrix Z replaces the observed quantities by its means:

$$Z = \begin{pmatrix} \hat{p}^{NLS}_{j(t_1)} \mu(q_{jk(t_1)}) & 0 & \dots & \dots \\ \cdot & 0 & 0 & \dots \\ 0 & \hat{p}^{NLS}_{j'(t_s)} \mu(q_{jk'(t_s)}) & 0 & \dots \\ 0 & \cdot & 0 & \dots \\ 0 & 0 & \cdot & \dots \end{pmatrix}$$

where: :

$$\mu(q_{jk(t_s)}) = \frac{\sum_{t_s \in \mathcal{T}} q_{jk(t_s)}}{T_{jk}}$$

$$T_{jk} = | \{ t \in \mathcal{T} | \tau_1(t) = jk \} |$$

Though the instruments are ideally not correlated any more with the disturbance term, the possible bias in the estimated price may still cause a bias (in the reverse direction) in the IV-estimate. It might also be the case of course that the measurement error is related to the type of measurement unit (that for example measurement in quintal (about hundred kilograms) is systematically overstated compared to say kilograms). In that cases the instruments may exaggerate the bias.

The other way out is assuming more information on the measurement error. One familiar assumption about the disturbance structure, especially made for estimating Cobb-Douglas and CES-production functions conveniently is the multiplicative disturbance. Assume for example that:

$$w_t = \overline{w}_t$$

$$q_t = \overline{q}_t \cdot e^{u_t}$$

with u_t normally distributed and hence $e^{u_t} \equiv v_t$ lognormally distributed. In that case OLS-estimates on the following equation are possible:

$$\ln w_t - \ln q_{jk(t)} = \ln a_{k(t)} + \ln p_{j(t)} + u_t \tag{7}$$

This relation suggests that errors tend to be greater with large quantities. Moreover it has the advantage that within our model the explaining variables are exogeneous here, so that we would get rid of the possible bias in the estimates.

To this multiplicative error an additive one can be added, for example to capture additional (additive) errors in the measurement of monetary amounts ($w_t = \overline{w}_t + \varepsilon$). Our final equation is then:

$$w_t = \Sigma_{j \in \mathcal{J}} \Sigma_{k \in \mathcal{K}} q_{jk(t)} a_k \cdot p_j \cdot e^{u_t} - \varepsilon_t \tag{8}$$

If we assume ε_t to be distributed normally, the density of w_t can be retrieved from the densities of ε_t and u_t by convolution (see Goldfeld and Quandt, 1972)[4]. It is:

$$l(w_t) = \frac{1}{2 \cdot \pi \cdot \sigma_u \sigma_\varepsilon} \int_0^\infty \frac{1}{x} \cdot e^{-\left(\frac{\left(\log x - \log\left(\Sigma_{j \in \mathcal{J}} \Sigma_{k \in \mathcal{K}} q_{jk(t)} a_k \cdot p_j\right)\right)^2}{2\sigma_u^2} + \frac{(w_t - x)^2}{2 \cdot \sigma_\varepsilon^2} \right)} dx \tag{9}$$

Taking logarithms, summing over the observations and maximizing w.r.t a_k, p_j, σ_u^2 and σ_ε^2 yield maximum likelihood estimates of prices and conversion rates. The last procedure is however not easy and it maybe worthwhile first to apply a Box-Cox transformation in order to see whether one of both disturbance structures (the normal or the lognormal) is sufficient to capture the measurement errors (see Judge et. al. 1985). This transformation amounts to:

$$\frac{w_t^\lambda - 1}{\lambda} = \frac{(\Sigma_{j \in \mathcal{J}} \Sigma_{k \in \mathcal{K}} q_{jk(t)} a_k \cdot p_j)^\lambda - 1}{\lambda} + \zeta_t \tag{10}$$

Clearly, taking limits for λ going to 1, (10) becomes identical to equation (2) and for λ going to zero the loglinear specification (7) is obtained. The parameter λ can be estimated by means of maximizing the following loglikelihood function:

$$L(\lambda, \alpha_k, p_j; w, q) = -\frac{T \cdot \ln(\pi)}{2} - \frac{T \cdot \ln \sigma_\zeta^2}{2} + (\lambda - 1) \cdot \ln w_t - \frac{\sum_{t \in \mathcal{T}} \left(\frac{w_t^\lambda - 1}{\lambda} - \frac{(\Sigma_{j \in \mathcal{J}} \Sigma_{k \in \mathcal{K}} q_{jk(t)} a_k \cdot p_{j(t)})^\lambda - 1}{\lambda} \right)^2}{2 \cdot \sigma_\zeta^2} \tag{11}$$

Maximizing (11) w.r.t. λ, σ_ζ^2, α_k and p_j gives ML-estimates which allow for testing whether λ significantly differs from 1 or 0. If so, estimating (8) by means of ML may be appropriate. If not we could stick to the simpler form (2) or (7).

4 Example: conversion rate teff for PA=5 (Yetmen)

For PA=5 and crop=1-2 (white and black,mixed teff), in case of production, and = 100 (teff), in case of consumer goods, we have 35 observations and 3 measurement units: 2=quintal (100 kg according to conversion table) 36=Madaberia (bag) (50 kg according to conversion table) and finally 39=Sahin, plate (1.389 kg according to conversion table). The price survey gives a mean price of 1.397 (1.327 for red teff).

Table 1 gives the results of nonlinear estimation of equation (2)[5]

---

[4] Independency between both errors is assumed here, though in principle there is no problem at all to allow for non-zero covariance between both errors (Goldfeld and Quandt, 1972, p. 146).

[5] Deze verschillen van de schattingen gerapporteerd in een eerdere versie omdat een ander, meer exacte optimeringsprocedure in SAS werd gebruikt. Controle d.m.v. GAMS geeft dezelfde resultaten! Mathematica is hier niet zo exact.

Table 1

parameter	estimate	Asymptotic standard error
$\hat{a}_2^{\text{NLS}}$	102.4436626	9.241526271
$\hat{a}_{36}^{\text{NLS}}$	54.9797906	31.254535225
$\hat{a}_{39}^{\text{NLS}}$	1.4660206	0.630194502
$\hat{p}_c^{\text{NLS}}$	1.3662291	2.618141671
$\hat{p}_p^{\text{NLS}}$	1.3641376	0.115249261
$\hat{\sigma}_u^2$	1777.6350	
-2Loglik	355.838	

In an appendix we report the absolute values of the error terms which suggest indeed some heteroskedasticity. Therefore we estimated equation 6. Table 2 reports this correction for heteroskedasticity.

Table 2[6]

parameter	estimate	Asymptotic standard error
$\hat{a}_2^{\text{NLS}}$	104.627	64.166659061
$\hat{a}_{36}^{\text{NLS}}$	54.571	314.63564666
$\hat{a}_{39}^{\text{NLS}}$	1.455	0.16916835
$\hat{p}_c^{\text{NLS}}$	1.379	.71683682
$\hat{p}_p^{\text{NLS}}$	1.374	.02174093
$\hat{\sigma}_\delta^{2*}$	1827.778	

*The variance was estimated as follows:

$\hat{\sigma}_\delta^2 = \frac{\sum_{t \in \mathcal{T}} \left(w_t - \Sigma_{j \in \mathcal{J}} \Sigma_{k \in \mathcal{K}} q_{jk(t)} \hat{a}_k^{\text{NLS}} \cdot \hat{p}_j^{\text{NLS}} \right)^2}{T - l}$, with *l* the number of parameters to estimate and the estimated parameters of the table 2 are used.

---

[6] De schatting werd met GAMS gemaakt. Voor standaardfouten werd versie (6′) met SAS geschat. SAS vertoont voor sommige startwaarden convergentie (en andere) problemen.

The fact that errors are presumed to be greater for greater measurement units pushes the estimated quintal conversion rate upwards, both other measures slightly downwards. This might indicate that quintal-observations are rather overreported as compared to other measures (remind that we do not correct for this correlation in the present regression). The fact that observations in quintal and madaberia are presumed to contain less exact information on the rates to be estimated, considerably increases the estimated standard errors of these parameters. Notice however the improvement in accuracy of the estimated consumer price and the conversion rate for Sahin. However these estimated standard errors should be compared with the corrected standard errors for the non-heteroskedastic model, taking into account their possible bias when estimated by the usual procedure.[7]

In order to correct for the positive correlation between error terms and variables, the IV-estimates (see equation (IV)) are given in table 3:

Table 3[8]

parameter	estimate	estimated quasi-standard error
$\hat{a}_2^{IV}$	105.04374	3.8776
$\hat{a}_{36}^{IV}$	54.979791	30.2455
$\hat{a}_{39}^{IV}$	1.4657307	0.71182
$\hat{\sigma}_u^2$	1702.3079	

Contrary to what might have been expected the estimated conversion rate for quintal is not corrected downwards. This could be because the instrument is still (positively) correlated with the error term, hence that quintal observations tend to be higher compared to the ideal fit than other measures.

Another way of treating heteroskedasticity and endogeneity of explanatory variables jointly is the loglinear specification (eq. 7). Estimates are reported in table 4

---

[7] Ik heb dat niet uitgerekend omdat ik dan eerst eens moet kijken hoe die assymptotische covarianties berekend worden bij niet-lineaire schattingen. Is het nodig dit te doen???

[8] De IML-module van SAS werd gebruikt

Table 4

parameter	estimate	standard error of logs	Variance of ML estimates[9]
$\hat{a}_2^{\text{OLS}}$	100.292207497	.06857873	40.5485
$\hat{a}_{36}^{\text{OLS}}$	53.2257339363	.18663975	84.5866
$\hat{a}_{39}^{\text{OLS}}$	1.3973410769	.11675123	00.0228126
$\hat{p}_c^{\text{OLS}}$	1.45383660995	.12907334	00.0301824
$\hat{p}_p^{\text{OLS}}$	1.40909266887	.05503706	00.00515521
		LS	ML
$\hat{\sigma}_v^2$		.03181	.0272616
R^2 (adjusted for absence of intercept)		.998	--
-2Loglik		--	309.640

Note the downward movement of the estimate for quintal conversion and the upward movement of both prices. This is in line with the correction we expected already earlier to take place. This regression performs remarkably well. All estimates are significant at the .05 level or higher and so it may be doubtful whether this can be still improved. To see the possible additional explanation of adding an additive disturbance term, table 5 reports the results of ML-estimation of the Box-Cox transformation (equation (11))

---

[9] This is the (estimated) Cramer-Rao lower bound, a χ^2- distributed value, asymptotically reached by the ML-estimates. The estimated ML-parameters of course coincide with the loglinear least squares estimates.

Table 5[10]

parameter	estimate	as. var.[11]	parameter	estimate	q.st.err.
$\hat{a}_2^{ML}$	96.4429	47.0861	$\hat{a}_2^{2SNLLS}$	89.9752	6.5720
$\hat{a}_{36}^{ML}$	51.5438	65.8823	$\hat{a}_{36}^{2SNLLS}$	48.8292	7.5751
$\hat{a}_{39}^{ML}$	1.35553	00.0125961	$\hat{a}_{39}^{2SNLLS}$	1.31771	0.0709
$\hat{p}_c^{ML}$	1.48697	00.0114277	$\hat{p}_c^{2SNLLS}$	1.50870	0.0469
$\hat{p}_p^{ML}$	1.45507	00.0065111	$\hat{p}_p^{2SNLLS}$	1.53597	0.0732
$\hat{\lambda}^{ML}$	-.179403	00.018661	$\hat{\lambda}^{2SNLLS}$	-.5246357	0.04443
$\hat{\sigma}_\zeta^2$	.00461458			.0329022	.00023491
-2Loglik	307.978			--	
H0 $\lambda = 0$	1.6832				
H0 $\lambda = 1$	47.881				

The likelihood-ratio test indicates that multiplicative disturbances are superior in explaining the error structure. It suggests furthermore that adding an additive disturbance might be useless. The 2SNLLS-approach however indicates that the most appropriate functional form is represented by a λ which is significantly lower than zero. Hence, due to measurement errors the likelihood-ratio would be biased in favour of both hypotheses. This is in line with the asymptotic properties of both estimators as studied by Amemiya and Powell (1981), at least if we accept the "true" λ to be negative. The lower variances of 2SNLLS are not surprisingly either, once again given that we are convinced of the "truth" of a negative λ.

---

[10] Berekend met behulp van Gams en Mathematica.

Amemyia criticised the usual ML-method for estimating Box-Cox transformations, since these explicitly rely on normality assumptions while the proposed transformation explicitly intends to test whether this assumption is acceptible. Therefore he proposed instrumental variable estimation (non-linear two stages least squares). The last three columns present the results of this procedure using as instruments the (average) derivatives of the function to be estimated w.r.t. the six parameters to estimate, evaluated in the ML-point-estimates (for further details see Amemyia, 1985, p. 249-252). In fact a slight modification *vis-à-vis* his proposal was introduced: we minimized the weighted sum of squared residuals *divided by the the geometric mean of the dependent variables raised to the power* λ to avoid the risk of λ going to plus or minus infinity as all variables are smaller or greater than 1. The use of the present instruments causes furthermore an adaption of the point-estimates which we tried to obtain by ordinary IV-estimation, so that the current instruments seem more appriopriate than the ones for example used in table 3. However this might be due to the use of a flexible functional form also. An estimation of 2SNLLS for the model with fixed parameter $\lambda = 1$ could throw more clarity on this.

[11] Sufficiency of the ML-estimates was assumed so that it sufficed to calculate the inverse of the negative Hessian of the loglikelihood function (see Judge, Hill et. al., 1988, p. 558 and Fomby, Hill and Johnson, 1984, pp.426-430).

Adding an additive disturbance to explain the measurment errors can in this case cause numerical problems as the following table indicates (it reports the ML-estimates of equation 8), since the best estimate would be $\hat{\sigma}_\varepsilon = 0$, but in this point the objective function is singular.

Table 6[12]

parameter	estimate	approximate asymptotic variance[13]
$\hat{a}_2^{\mathrm{ML}}$	100.2943	40.5514
$\hat{a}_{36}^{\mathrm{ML}}$	53.2305	84.6208
$\hat{a}_{39}^{\mathrm{ML}}$	1.3969	00.0227952
$\hat{p}_c^{\mathrm{ML}}$	1.453324	0.0302136
$\hat{p}_p^{\mathrm{ML}}$	1.40727	0.0051549
$\hat{\sigma}_u$	.165125	
$\hat{\sigma}_\varepsilon$	3.49898E-7	
-2Loglik	309.714	

As expected the lognormal error structure absorbs almost all of the errors. It is however not easy to compare the performance of this regression with previous ones. No obvious tests are available. Goldfeld and Quandt (1972) suggest the standard likelihood ratio tests (p.146) which would result in a $\chi^2(1)$ test statistic of -.074 compared to the lognormal estimates, indicating no significant contribution of the normal part[14].

---

[12] **12** Berekend met behulp van Mathematica, na heel veel problemen omdat de "normale" variantie naar nul neigt. Oplossingen nogal afhankelijk van startwaarden (tot op 1 cijfers na de komma wijzigingen voor quintal, tot op 2 cijfers op de niet nul plaatsen voor de andere parameters). Controle van Mathematica’s toverdoos werd gedaan door het genereren van afhankelijke veranderlijken overeenkomstig de opgelegde verdeling van het storingspatroon. De resultaten laten toe de software te vertrouwen.

[13] **13** Here also suffiency was assumed as Goldfeld and Quandt did (1972, p.145). Moreover we used the less exact Simpson Rule (parabolic approximation) to calculate out the numerical integrals. A test revealed that the loglikelihood function resulted in 154.82 (35 integrals) with this approximation, compared with 154.867 for the Gaussian approximation, used by Mathematica. The simplification was necessary in order to obtain results in a reasonable time with the currently used machine (486SX with co-processor), i.e. about 8 hours. A comparison with variances calculated on the assumption that the error structure was lognormal, which is nearly the case here, shows that the results are close to each other as was expected (the "lognormal variances" are: 40.5562, 84.6165, .0228007, .0302103, .0051557).

[14] **14** The fact that the χ^2 value becomes negative is due to truncation of the numerical integral.

Appendix

Table : Error terms in absolute value of equation (2)

error	quantity	measurement unit
0	1.0	36
0.004	30.0	39
0.006	40.0	39
0.015	5.0	39
0.029	10.0	39
0.253	1.0	2
0.253	1.0	2
0.268	2.0	1 kilograms
0.505	2.0	2
1.264	5.0	2
2.253	1.0	2
3.793	50.0	1
3.857	194.0	1
4.483	125.0	1
5.495	2.0	2
5.586	100.0	1
6.010	70.0	1
7.121	150.0	1
7.126	0.5	2
8.266	13.0	1
8.848	60.0	1
9.495	2.0	2
9.747	1.0	2
9.747	1.0	2
9.747	1.0	2
9.874	0.5	2
19.495	2.0	2
20.379	150.0	1
20.505	2.0	2
29.242	3.0	2
33.697	120.0	1
39.747	1.0	2
48.736	5.0	2
84.379	1.5	2
195.379	1.5	2

A Goldfeld-Quandt test on the basis of increasing order of measurement unit and within measurement unit according to order of magnitudes of quantities yield an F(20,18) value of 39.9967 and when the sample is split up according to order of magnitude of estimated quantities in kilograms we get F(20,14)=146.074. Heteroskedasticity cannot be rejected therefore.

REFERENCES

AMEMIYA, T., *Advanced Econometrics*, Basil Blackwell, Oxford, 1986

AMEMIYA, T., and POWELL, J.L., 1981, *A comparison of the Box-Cox maximum-likelihood estimator and the non-linear two-stage least squares estimator*, Journal of Econometrics, 17, 351-381

DEATON, A., 1987, *Estimation of own- and cross-price elasticities from household survey data*, Journal of Econometrics, 36, 7-30

DEATON, A., 1988, *Quality, quantity, and spatial variation of prices,* American Economic Review, 78, 418-430

DEATON, A., 1990, *Price elasticities from survey data. Extensions and Indonesian results,* Journal of Econometrics, 44, 281-309

FOMBY, T.B., HILL, C.R., JOHNSON, S.R., *Advanced Econometric Methods,* Springer-Verlag, Berlin, 1984

GOLDFELD, , S.M., QUANDT, R.E., *Nonlinear Methods in Econometrics,* North-Holland, Amsterdam, 1972

JUDGE, G.C., HILL, R.C., GRIFFITHS, W.E., LÜTKEPOHL, H., LEE, T., *The Theory and Practice of Econometrics,* 2nd ed., John Wiley & Sons, New York, 1985

JUDGE, G.C., HILL, R.C., GRIFFITHS, W.E., LÜTKEPOHL, H., LEE, T., *Introduction to the Theory and Practice of Econometrics,* 2nd ed., John Wiley & Sons, New York, 1988