

see Deaton (1988):

$$y_i = a_i + \ln x_i + \theta \ln p_i + u_i, \quad \ln v_i = a_z + \beta_2 \ln x_i + \ln P_c + u_{i,c}, \quad (33)$$

where i is a household, and c is the cluster or village in which it lives. The first equation explains the demand for the good — for example the logarithm of quantity or the budget share of the good — in terms of household total expenditure x , the unobservable price p , and village fixed effect f , and a random error term. The price is assumed to be the same for all households in the village and is therefore not indexed on i . The fixed (or random) effect is uncorrelated with the price, but can be correlated with x or with any other included variables that are not constant within the cluster. The unobservable price also manifests itself through the unit value v which is the dependent variable in the second equation. The parameter β_2 is the elasticity of unit value to total expenditure, or quality elasticity — Prais and Houthakker (1955) — while θ allows for possible quality shading in response to price changes. If price and unit value were identical, θ would be unity and β_2 would be zero, but quality effects will make $\theta > 0$ and $\beta_2 > 0$.

Once again, identification is a problem, and as is intuitively obvious from using the second equation to substitute out for the unobservable log price, only the ratio θ/β_2 can be estimated. The β_2 -parameters can be estimated by a within-estimator in which village effects are swept out, a procedure that also provides estimates of the variances and covariances of u_i and $u_{i,c}$. Given the β_2 's from the within-village estimates, construct the corrected village averages

$$z_{i,c} = y_i - \beta_2 \ln x_{i,c}, \quad z_c = \ln v_c - \beta_2 \ln x_c. \quad (34)$$

At the second stage, we calculate the estimate

$$\frac{\text{cov}(z, z_2) - n^{-1} \sum_c r_c \text{cov}(z_c, z_{2c})}{\text{var}(z_2) - n^{-1} \sum_c r_c \text{var}(z_{2c})} \quad (35)$$

where the covariances are taken over villages, where r_c is the number of households per cluster, and the 0's are estimated from the first stage variance covariance matrix of the residuals. Using (33), it is straightforward to show that (35) is a consistent estimate of the ratio θ/β_2 .

Note that (35) is a standard errors-in-variables estimator in which the OLS estimator, which is the ratio of the covariance to the variance, is corrected for the component that is attributable to measurement error. The standard error for θ/β_2 can

be obtained from first principles by application of the "delta method", or by adapting the formulas in Fuller (1987) for the effects of the