Why Does the Engel Method Work?

Food Demand, Economies of Size and Household Survey Methods

John Gibson[†]

Abstract

Estimates of household size economies are needed for the analysis of poverty and inequality. This paper shows that Engel estimates of size economies are large when household expenditures are obtained by respondent recall but small when expenditures are obtained by daily recording in diaries. Expenditure estimates from recall surveys appear to have measurement errors correlated with household size. As well as demonstrating the fragility of Engel estimates of size economies, these results help resolve a puzzle raised by Deaton and Paxson (1998) about differences between rich and poor countries in the effect of household size on food demand.

JEL: C81, I32

Keywords: Food demand, Economies of size, Household surveys, Measurement error

Acknowledgements:

I am grateful to Ron Bewley, Angus Deaton, Les Oxley, Peter Phillips, Scott Rozelle, Christopher Scott, Jeffrey Williams, Fred Zimmerman and an anonymous referee for helpful advice. Data for this research are part of a World Bank poverty assessment for Papua New Guinea, for which financial support from the governments of Australia (TF-032753), Japan (TF-029460), and New Zealand (TF-033936) is gratefully acknowledged. All views in this paper are those of the author and should not be attributed to the World Bank.

Word count: 7679

[†]Department of Economics, University of Waikato, Private Bag 3105, Hamilton, New Zealand.
Fax: (64-7) 838-4331. E-mail: jkgibson@waikato.ac.nz

I. Introduction

Economists are increasingly aware of the effect that assumptions about economies of household size have on poverty and inequality comparisons (Coulter, Cowell and Jenkins, 1992). For example, whether people in Indian households headed by women are more likely to be poor than are those in male-headed households depends on the adjustment made for size economies (Dreze and Srinivasan, 1997). In the transition economies, the rising relative cost of housing has made size economies more important, shifting the incidence of poverty toward small households and affecting conclusions about whether public interventions should be aimed at children or at the elderly (Lanjouw, Milanovic and Paternostro, 1998).

But despite their importance, there is no generally accepted method of measuring household size economies. The Engel method is popular because it is simple, using food budget shares to indicate the welfare of different sized households (Lancaster, Ray and Valuezuela, 1999).[1] But this method is silent on how larger households actually achieve economies and it also lacks theoretical justification. For example, Deaton (1997) constructs two separate cost functions that a household of size n faces to reach utility level u at prices p, $c(u, p, n)$. The size elasticities, $\partial \ln c / \partial \ln n$, of these cost functions differ but the same food Engel curve is derived from both of them, indicating a lack of identification. The Engel method also appears contradictory; if there are size economies, a larger household with the same per capita expenditures as a smaller household is better-off, and so should have a lower food share. But a decline in the food share with constant per capita expenditures can occur only if there is a decline in food spending per person, which is not what is expected when welfare increases, especially in poor countries (Deaton and Paxson, 1998).

In addition to these theoretical problems, this paper reports evidence on the empirical fragility of Engel estimates of household size economies. The results use a survey where a sample of households had each adult report daily purchases in diaries, while a matched sample had a single respondent give a verbal recall of the household's expenditure for the past fortnight.[2] Both of these methods are widely used by household expenditure surveys in developed and developing countries. Regression results from the matched samples, when compared with the results from Monte Carlo experiments, suggest that recalled food expenditures have measurement errors correlated with household size. These correlated errors cause a negative bias in the coefficient on household size in regression models of food budget shares and, consequently, cause Engel estimates of size economies to be overstated. This error is shown to affect the cross-sectional pattern of poverty in the sampled population.

Given these problems with the Engel method, it would be helpful to have another simple method of estimating household size economies. A promising approach is based on the effect of public goods within the household (Deaton, 1997). The intuition is that increases in household size (holding outlay per head constant) allow the resources released by the wider sharing of public goods (e.g., heat and light) to be spent on both public and private goods. The effective price of public goods is lower in larger households, so substitution effects are away from private goods. Hence, there is a positive income effect and a negative substitution effect on the demand for private goods. But for private goods like food, the income effect should prevail because the absolute value of the own-price elasticity is likely to be lower than the income elasticity, and food demand should rise, especially in poor countries. A measure of size economies could be calculated from the notional reduction in outlay needed to prevent the rise in per head spending on food.

However, the development of this public goods method is currently blocked by an empirical puzzle about the relationship between household size and measured food demand. The puzzle is that food expenditure per head *falls* as household size rises, holding outlay per head constant. Deaton and Paxson (1998) find that a unit increase in the logarithm of household size decreases the budget share of food by up to 10 percentage points in a group of poor countries (Thailand, Pakistan, and African households in South Africa), holding outlay per person constant.[3] The food share falls by 1-2 percentage points in Taiwan and the U.S. and by less in France and Britain.[4] Not only do these empirical results contradict the prediction of the public goods model, they do so in a pattern across countries that is just the opposite to what would be expected. The positive relationship between per capita food expenditure and household size should be strongest in poor countries, where the income elasticity of food demand is most likely to exceed the (absolute) own-price elasticity.

Deaton and Paxson (1998) reject a measurement error explanation for this puzzle because they cannot see why the respondent in a recall survey should be worse informed about others' food consumption than about non-food consumption. But there are grounds for this asymmetry in reporting errors, and for expecting the errors to rise with household size. For example, in the survey used here, a household with two people makes an average of 50 food purchases per fortnight, while a household with 10 people makes 140 food purchases per fortnight. Thus, the respondent from the larger household is the one more likely to forget food purchases when giving her verbal report on expenditures in the previous fortnight. But whether the household has two people or 10 people, it still needs only one gas stove, so the reporting task for non-foods is easier and less proportional to household size. While purchase frequency of some non-food items rises with household size (e.g., bus fares), the rate of increase is less than it is for food purchases,[5] and these items are only a small part of total non-food spending. Furthermore, household surveys often impute the value of consumption of certain non-food items (e.g., durables and rent), using observations and measurements made by the interviewers. The measurement error in these imputations is likely to be orthogonal to household size, in contrast to the errors in food expenditures.

Perhaps because of these differential reporting requirements, a food Engel curve estimated on the sample who received the recall method looks most similar to Deaton and

Paxson's results for African households – the food budget share falls sharply with increasing household size (at constant outlay per person). But when the sample who were given diaries are used, the results look most similar to their results for Britain – there is a statistically insignificant relationship between the food share and household size.

The next section of the paper reviews evidence from previous studies of measurement error in household expenditure surveys. Food budget share models and the Engel and public goods methods of estimating size elasticities are introduced in Section III, and the effects of measurement error are examined using Monte Carlo methods. Section IV describes the field experiment where data on household expenditures were gathered using the diary method and the recall method on matched samples. Section V presents the estimation strategy and the empirical results. In this section, food Engel curves are estimated and size elasticities are derived, which are then used to examine the pattern of poverty across household size groups. Section VI concludes.

II. Measurement Errors in Diary and Recall Surveys of Household Expenditures

Although some reports suggest that recording in diaries gives more accurate data, especially for frequently purchased items like food (for example, see Branch, 1994; Republic of China, 1990), studies are hampered by lack of data on actual expenditures. This makes it difficult to interpret the comparisons of diary and recall surveys (Kemsley and Nicholson, 1960; McWhinney and Champion, 1974). These comparisons are also clouded because they vary not only in terms of data collection methods – diary versus recall – but also in terms of whether the time frame for the reporting is bounded by visits from the interviewer. In these studies, interviewers ask respondents to recall expenditures from the previous week and then give the family diaries to be completed in the week following. While an interview marks the start of the diary-keeping period, no such event marks the beginning of the recall period, so the recall estimates may be biased upwards by *telescoping errors* – the incorrect placement by the respondent of earlier expenditures in the recall period (Sudman and Bradburn, 1973). Furthermore, these previous studies mainly compare total expenditure estimates and do not show whether discrepancies between the two survey methods vary with household size.

Other relevant evidence comes from experiments with different recall periods. Scott and Amenuvegbe (1991) find that average daily expenditures reported by respondents fall by almost three percent for every day added to the recall period, with the greatest decline for the more frequently purchased items.[6] This tendency to forget earlier transactions – what Deaton (1997) calls "progressive amnesia" – is relevant to the comparison of diaries with recall. Diary recorders usually make written reports on the day an expenditure occurs, which should minimize recall errors, while respondents in an interview are asked to recall expenditures that may have been made many days earlier. Scott and Amenuvegbe (1991) also suggest that after a threshold number of purchases during the recall period, respondents switch from reporting their actual expenditures to what they think are their usual expenditures. This change in reporting style means that exceptional expenditures tend to be overlooked; typically, exceptional means exceptionally high, so this leads to underestimation. Because larger households generate more transactions per time period, they are more likely to reach the threshold where respondents switch away from reporting actual expenditures. Moreover, a larger household may have a higher proportion of people who are non-nuclear family members and who make purchases that the respondent does not know about.[7]

By themselves, errors in recalling expenditures do not explain why the measured food share may vary with data collection methods. If all types of expenditures were collected by either recording in diaries or respondent recall in an interview, recall errors would reduce the sub-total of expenditures on each category of consumption by the same proportion. But recording in diaries is feasible for only two out of the four components of consumption that surveys typically cover – food expenditures and day-to-day non-food expenses. Diaries are not feasible for infrequent expenses (e.g., school fees) and services from durable goods and dwellings, and these two components may account for two-thirds of non-food consumption. Any difference between diaries and recall affects all of reported food expenditure but perhaps only one-third of non-food expenditure, thus affecting the measured food share.

III. Food Demand Models and Measurement Error

The relationship between measured food demand and household size is central to both the Engel method and the public goods method of measuring household size economies. Deaton and Paxson (1998) use a model with household consumption of food, q_f and a non-food good, such as housing, q_h, to show the necessary conditions for increases in household size to raise per capita food demand. Both goods are subject to some commodity-specific scale economies, so that effective household size for the consumption of each good is not n, but rather $f_i(n)$, where $i=f$, h and the commodity-specific economy of scale measure is:

$$s_i = 1 - \frac{\partial \ln f_i(n)}{\partial \ln n} \tag{1}$$

The per capita food demand function is:

$$\frac{q_f}{n} = \frac{f_f(n)}{n} g_f\left(\frac{x}{n}, \frac{p_f f_f(n)}{n}, \frac{p_h f_h(n)}{n}\right) \tag{2}$$

where p_f and p_h are the price of food and non-food, x is household total expenditure, and $g_f(x, p_f, p_h)$ is the food demand function for a single person household. When the logarithm of equation (2) is differentiated with respect to ln n, the condition for per capita food consumption to increase with household size, holding x/n constant, is:

$$s_h(e_{fx} + e_{ff}) - s_f(1 + e_{ff}) > 0 \tag{3}$$

where e_{ff} and e_{fx} are the own-price and income elasticities of demand for food. Thus, as long as non-food contains some public goods, so that $s_h \neq 0$, while food is a pure private good, equation (3) requires the absolute value of the own-price elasticity of food demand to be less than the income elasticity of food demand. This condition is most likely to hold for poor consumers, so the positive effect of household size on per capita food consumption and food budget shares should be strongest in poor countries.

To test whether the empirical evidence is consistent with this pattern, Deaton and Paxson (1998) using the following food share model:

$$\frac{p_f q_f}{x} = w_f = a + b \ln\left(\frac{x}{n}\right) + g \ln n + \sum_{j=1}^{J-1} h_j r_j + d \cdot z + u \qquad (4)$$

where $r_j = n_j / n$ is the proportion of persons in the household in demographic group j, z is a vector of other household characteristics, u is a disturbance term, and a, b, g, h, and d are parameters to be estimated. If the condition in equation (3) holds, $\hat{g}$ should be positive, with the largest values estimated from household data in poor countries. In fact, the empirical results of Deaton and Paxson show exactly the opposite pattern.

The same food share model, when reparameterised, can provide Engel estimates of size economies, although the assumptions underlying the model are substantially different to those used by Deaton and Paxson (1998). In the case of the Engel method, no distinction is made between private and public goods and the food budget share is a sufficient welfare measure. Under the Engel method interpretation, if some economies of scale are present, $\hat{g}$ would be expected to be negative, rather than positive as argued by Deaton and Paxson. A recent example of this Engel method approach is provided by Lanjouw and Ravallion (1995) who use data from Pakistan to estimate:

$$w_f = a + b \ln\left(\frac{x}{n^{1-s}}\right) + \sum_{j=1}^{J-1} h_j r_j + d \cdot z + u, \qquad (5)$$

which is identical to equation (4) because $g=bs$.[8] According to equation (5), if x^0 is the outlay of a 1-person household, an n-person household of the same composition needs total outlay of $x^0 n^{1-s}$ to have the same food share (and the same welfare level, by assumption). Lanjouw and Ravallion estimate s to be 0.4, suggesting that ten individuals, each spending, say, \$1 per day in separate single-dweller households, can achieve the same welfare level living as a 10-person single household with total expenditures of just \$4 per day ($10^{0.6}$=3.98). These size economies imply surprisingly large falls in food spending per head for consumers in a poor country. With an average food share of 0.5 in Pakistan, the per-head spending on food in the 10-person group declines by 60 percent, from, say, 50 cents per day to 20 cents per day.

The Monte Carlo experiments may show whether measurement errors could cause such large estimates of size economies, since it is possible to retrieve an estimate of s from the ratio of g to b. Even errors in expenditures that are uncorrelated with any explanatory variables may matter because $\ln(x/n)$ and w_f are constructed from the same information ($x = p_f q_f + p_{nf} q_{nf}$ and $w_f = p_f q_f / x$). Because $\ln(x/n)$ and $\ln n$ are negatively correlated by construction, errors in $\ln(x/n)$ are likely to bias $\hat{g}$, but in an unpredictable direction (Deaton and Paxson, 1998). Bias in $\hat{g}$ is even more likely if the errors are correlated with household size or with the true value of expenditures.[9]

The Monte Carlo experiments use a simplified form of equation (4):

$$w_f = a + b \ln\left(\frac{x}{n}\right) + g \ln n + u, \qquad (6)$$

where a=1.6, b=-0.14, and g=-0.007. The value chosen for g is similar to the estimates of g made by Deaton and Paxson using a household survey from France, with expenditures reported in diaries. The experiments aim to answer the following question: what sort of measurement errors could cause equation (6) to give estimates of g like those found when using household surveys from poor countries with expenditure data collected by recall, specifically, $-0.09 \le \hat{g} \le -0.05$. This experimental design is not meant to imply that reports of expenditures made in diaries have no errors and that estimates of g coming from such data are the true values. Rather, this design is used to reveal *differential bias* – the extent to which estimates of g coming from recall surveys could diverge from the estimates coming from diary surveys due to measurement error in recall surveys that is over and above the measurement error in diary surveys.

The experiments allow uncorrelated errors in expenditures, errors that are correlated with the true value of expenditures, and errors that are correlated with household size. Each experiment moves from a situation of no error to cases of increasingly severe measurement error. Initially, just food expenditures are measured with error, with the experiments carried out as follows: Samples of 1000 observations on (log) total expenditure, x and household size, n were generated from a bivariate normal distribution: $N_2[m_x, m_n, s_x^2, s_n^2, r]$, with m_x=9.5, m_n=6.7, s_x=0.7, s_n=3.5, and r=0.2.[10] The regression errors, u were normal, with mean zero and standard deviation 0.1. Any draws with food budget shares outside the range 0.05-0.95 were dropped.[11] Total expenditure, x was partitioned into food expenditures, $x_f = x \cdot w_f$ and non-food expenditures, $x_{nf} = x - x_f$. A proportionate error was added to true food expenditures, so that the observed variable was $\ln \tilde{x}_f = \ln x_f + v$. In the first experiment the measurement error is independent: $v \sim N(0, s_v^2)$, with three values of s_v used; 0.1, 0.2, and 0.3. In the second experiment $v = j \ln x_f + e$, where $e \sim N(0, s_e^2)$ and $E(e,x_f)$=0. In the third experiment $v = l \ln n + e$, where $e \sim N(0, s_e^2)$ and $E(e,n)$=0. In the second and third experiments, the values used for j and l were -0.3, -0.2, -0.1, 0.1, 0.2, and 0.3. The error-ridden total expenditure and food share variables were reconstructed as $\tilde{x} = \tilde{x}_f + x_{nf}$ and $\tilde{w}_f = \tilde{x}_f / \tilde{x}$, and equation (6) was estimated.

The results of the Monte Carlo experiments can be summarized by the following three observations: First, errors in measuring food expenditures that are negatively correlated with either household size or with the true value of food expenditures are the only type of errors that could cause estimates of g to be like those found when using surveys from poor countries with expenditure data collected by recall, specifically, $-0.09 \le \hat{g} \le -0.05$ (see Table 1, row 2b. and 3b.). Second, if measurement errors are correlated with the true value of expenditures, the coefficient on ln (*x*/*n*), $\hat{b}$ will suffer attenuation bias (i.e., towards zero) but if errors are correlated with household size, there will be no effect on $\hat{b}$ (see row 2a. and 3a.). Third, if the true level of size economies (according to the Engel method) is s=0.05, errors in measuring food expenditures that are negatively correlated with either true values (row 2c.) or with household size (row 3c.) will cause $\hat{s}$ to be biased upwards, with a range of values that includes the estimate of s=0.4 found by Lanjouw and Ravallion (1995).

(Table 1 about here)

When measurement errors in non-food expenditures that are uncorrelated with the errors in food expenditures are introduced, the pattern of results is largely unchanged. If the errors in non-food expenditures are independent, i.e., $\ln \tilde{x}_{nf} = \ln x_{nf} + g$ where $g \sim N(0{,}0.4)$, the effect of food expenditure errors is amplified slightly. If the errors in non-food expenditures vary negatively with household size, $g = -0.2 \ln n + z$ where $z \sim N(0{,}0.4)$ and there is only a weak correlation between food expenditure errors and household size (l=-0.1), the estimate of g tends to be positive. But as food expenditure errors become more strongly correlated with household size (l≤-0.2) the expected value of $\hat{g}$ moves into the range $-0.09 \le \hat{g} \le -0.05$. Errors that are correlated with household size have no effect on $\hat{b}$, which is the same as was found when there were no errors in non-food expenditures.

IV. The Household Survey

The household survey was carried out between April and December 1996 in Port Moresby, the capital of Papua New Guinea (PNG). Interviewers were graduates of the national university while the field supervisors were senior academics. The interview teams were monitored by the manager of the university's research company and by the current author. Interview teams did not specialize; the same team would use diaries to gather expenditure data on some households and recall interviews for other households.

Data Collection

The questionnaire that used respondent recall was typical of expenditure surveys currently used in developing countries. To avoid telescoping errors, there were two interviews, two weeks apart, so that the start of the recall period was signaled by the first interview. This first interview collected information on demographics, education, income sources and housing conditions, while the second interview collected data on expenditures, durable assets and transfers. The expenditure data were collected for 36 categories of food and for 20 categories of other frequent expenses during the recall period. The expenditure estimates include the imputed value of own-production, net gifts received, and food stock changes (measured by the interviewer) for the period since the first interview. In addition, information was collected on 31 categories of infrequent expenses for the 12 months prior to the interview. This longer-term recall is 'unbounded' in the sense that there was no definite event, like the initial visit by the interviewer, to mark the beginning of the period over which the respondent was meant to recall expenditures. The final component of the survey was an inventory of 16 durable assets, which was used with data about the dwelling to estimate the value of the flow of annual services from durable goods and dwellings.

For the questionnaire where each adult recorded daily expenditures, the diaries were organized by means of acquisition (purchases, gifts, own-production) and lists of easily forgotten items were included. Interviewers visited every three to four days to check that respondents had made written daily reports. Interviewers also measured food stocks at the beginning and end of the 14-day period. Data on the infrequent expenditures of these households were collected using the same 12-month recall as used by the other questionnaire. The data on durable goods and dwellings also were obtained in the same manner as the other questionnaire.

Sample Design

Households within finely defined area units were randomly allocated into two groups: one receiving the diary method and the other receiving the recall method. This is a break from previous studies that apply the two methods *sequentially* to the same household. This sequential design was not used because it may be subject to *conditioning bias* – persons who learn to report their expenditures in diaries may, subsequently, be atypically accurate recall respondents, while persons initially given recall questionnaires may find daily recording to be onerous and be atypically bad diary-keepers.

Two-stage sampling was used, with 38 clusters initially selected with probability proportional to estimated size from a frame made up of the 1990 Census plus areas of recent settlement (500 clusters in total). The frame was divided into nine strata, corresponding to major districts of the city, with equal sampling rates used for each strata. At the second stage, circular systematic sampling was used to select six households in each cluster, and a further four households who were "reserves" that were surveyed only if some of the original six selections were absent or refused at the time of the initial interviews. Households were divided into the diary and recall samples at this stage, using a coin toss to decide whether the 1st, 3rd, and 5th households received diaries and the 2nd, 4th and 6th received the recall questionnaire, or *vice versa*. The alternating pattern of diary and recall samples continued into the “reserve” households if they were needed.

There was a potential sample of 228 households (38×6) but eight households had missing data due to absence at the time of the second visit for the recall interview or the departure of a household during the diary-keeping period. The reserve households were not used as replacements when this problem arose because a bounded recall would not be possible for these households. To keep the matched nature of the two samples for this analysis, a non-response by a recall household causes the closest diary household within the same cluster to be dropped from the sample (and *vice versa*). This leaves a sample of 212 households. The sample is weighted because selection based on the 1990 Census under-represents clusters found to be larger during the household listing.

Summary statistics for the two samples are reported in Table 2. The average food budget share was 51 percent for the sample where the diary method was used and 45 percent when the recall method was used. This statistically significant difference in the food budget share is driven by the lower estimates of per capita food expenditures when the recall method is used, while there is no significant difference in non-food expenditures between the two samples. The difference in food expenditures and food shares is unlikely to be due to differing characteristics of the households across the two samples. All of the demographic variables, except the share of seven to 14 year old males, had no significant differences in means between the two samples. This is to be expected because the demographic and employment variables were collected at the initial interview and are thus independent of the choice of questionnaire. The other variable affected by the choice of questionnaire - per capita expenditure - had an average that was 13 percent lower when the recall method was used, but this difference was not statistically significant.

(Table 2 about here)

V. Estimation Methods and Results

The model is based on the specification used by Deaton and Paxson, excluding variables not relevant in an urban areas (see equation 1). In addition to per capita expenditure and household size, there are eight demographic ratios, and the adults employment rate. Employment may affect the food share because of higher caloric requirements for workers or because of the higher cost of meals eaten out of the home. The model also includes dummy variables for the calendar quarter in which the household was surveyed. The model is separately estimated on the sample of households whose expenditures were reported in diaries and on the sample whose expenditures were recalled in an interview. Slope and intercept dummy variables from a model estimated on the pooled sample are used to test which, if any, of the coefficients differ between the two samples. The model is estimated by both Ordinary Least Squares (OLS) and Instrumental Variables (IV), and the estimation methods account for the clustered, stratified, and weighted nature of the sample.

OLS Results

The method used to collect expenditure data affects the relationship between household size and food share but does not affect any of the other coefficient estimates of the food Engel curve (Table 3). When the Engel curve is estimated on the sample of households whose expenditures were recalled in an interview, household size appears to exert a negative and statistically significant effect on the food budget share, holding per capita outlay constant. The effect is even larger than that found for the poorest households in Deaton and Paxson's sample; a unit increase in the logarithm of household size decreases the budget share of food by 12 percentage points (10 percentage points if using the unweighted data). But when the Engel curve is estimated on the sample of households whose expenditures were reported in diaries, household size has no statistically significant effect on the food budget share, and the point estimate is almost zero – a result that is similar to what Deaton and Paxson find for Britain. Hence, the variation in data collection methods across the group of countries studied by Deaton and Paxson may account for some of their results.

(Table 3 about here)

The Monte Carlo experiments suggest that measurement errors in expenditures that are negatively correlated with either household size or with the true value of food expenditures could cause negative bias in the coefficient on household size. According to the Monte Carlo results, we can distinguish which of these two types of measurement error are present by what happens to the coefficient on ln (x/n). If errors are correlated with true values, $\hat{b}$ will be biased toward zero but if errors are correlated with household size, there will be no effect on $\hat{b}$. The coefficient on ln (x/n) when the diary sample is used is identical to the coefficient when the recall sample is used. This suggests that the measurement errors in the recall sample are negatively correlated with household size.

Instrumental Variables Results

The IV specification of Deaton and Paxson, where per capita cash income instrumented for ln (x/n), was not exactly replicated because income data were not gathered by the PNG survey.[12] Measures of wealth were available (dwelling quality and value of durables) but were closely related to imputed non-food expenditures and hence directly influenced the food budget share. Therefore, the instruments chosen were the average number of school years of each adult in the household and the age of the household head, which are both good predictors of per capita expenditures. These two regressors raise the R^2 in the first stage regression of ln (x/n) on the exogenous variables from 0.41 to 0.71 and the F-test for excluding them is highly significant, $F_{(2,28)}$=34.64. Over-identification tests suggested that these two variables did not play a direct role in the explanation of food budget shares, so they should be valid instruments, even though they have sometimes appeared in food demand studies elsewhere. One concern with using schooling as an instrument is that illiteracy (a close correlate) could cause measurement error in expenditure data that are collected by the diary method. However, this should not be a problem in the current study because adult literacy rates are high by developing country standards (approximately 90 percent).[13]

The IV estimates of the Engel curve, reported in Table 4, show the same pattern as the OLS estimates. Just as before, the only variable with a statistically significant difference in coefficients between the diary and recall samples is household size. The use of an instrument for ln (x/n) causes only small changes in the value of the coefficients on household size, compared with the OLS results.[14] There also are small changes in the coefficients on per capita expenditures and a widening of the standard errors surrounding all coefficients. However, the Durbin-Wu-Hausman test suggests no significant difference between the OLS and the IV estimates.

(Table 4 about here)

Engel Estimates of Size Economies

The results reported in Tables 3 and 4 suggest that an appropriate model for the pooled sample needs only one slope dummy variable, for household size, and that it is safe to use OLS to estimate such a model. The results of this model are:

$$w_f = 1.480 - \underset{(9.85)}{0.140} \ln (x/n) - \underset{(3.41)}{0.083} \ln n + \underset{(4.36)}{0.049} [\ln n * \text{Diary Dummy}]$$

$$+ \text{demographic ratios} + \text{adult employment rate} + \text{quarterly dummies}$$

$$R^2=0.44; \quad F_{(13,17)} = 15.84$$

t-statistics in ( ) corrected for clustering, sampling weights and stratification

These results suggest that when an Engel curve is estimated with expenditure data collected by recall, a unit increase in the logarithm of household size will cause the observed food share to fall by five percentage points more than it would if the data were collected by respondents reporting expenditures in diaries. The elasticity of per capita food expenditure with respect to household size (which equals g/w_f given per capita expenditure) is estimated as –0.184 when using expenditure data collected by recall but only –0.067 when using expenditure data reported in diaries (at the average food shares in Table 2).

According to the Engel method, the size economies parameter s can be estimated from the ratio of the coefficients on ln *n* and ln (*x*/*n*). The coefficients from the pooled model reported above give estimates of s for the diary and recall samples of:

Diary Sample	Recall Sample
s = 0.24	s = 0.59
s.e.(s) = 0.15	s.e.(s) = 0.18
$H_{0:}$ $s = 0$ $\chi^2_{(1)} = 2.54$ ($p<0.12$)	$H_{0:}$ $s = 0$ $\chi^2_{(1)} = 10.77$ ($p<0.01$)

Engel estimates of size economies appear to be sensitive to the method used to collect household expenditure data. When expenditure data are collected from recall interviews, size economies appear large. But when expenditure data are collected by having a similar sample of households record expenditures in diaries, estimated size economies are much smaller and a per capita normalization of total expenditures (i.e., s=0) would not be rejected. Thus it is possible that the estimate of s=0.4 made by Lanjouw and Ravallion (1995) for households in Pakistan is biased upwards by the use of recall data, which may have measurement errors in expenditures that are correlated with household size.

Figure 1 shows the pattern of headcount poverty by household size class for the various estimates of the economies of household size parameter, σ.[15] These poverty rates are based on normalised estimates of effective household size, n^{1-s} so that the poverty rate is always the same (37 percent) for the average sized household. It is clear that when no allowance is made for economies of size, the headcount poverty rate increases rapidly with household size, approaching almost 50 percent for households with more than ten members. In contrast, when the estimate of size economies calculated from the recall sample (σ=0.59) is used, the largest households have a poverty rate of around 20 percent but the smallest have a poverty rate of 50 percent. The robustness of this pattern must be questioned, however, because with the estimate of σ from the diary sample, a slight rise in poverty with increasing household size is implied.

VI. Conclusions

Economists who want to measure economies of household scale face unpalatable choices. They can use the atheoretical Engel method, which works but makes no sense (Deaton, 1997) or they can try a method – based on public goods within the household – that makes sense but does not work.[16] This public goods method does not work because its prediction of higher food expenditure per head in larger households (of equal per capita outlay) is rejected by the data (Deaton and Paxson, 1998). The results reported here suggest that it is no coincidence that the prediction of higher food expenditures fails most in countries where household budget surveys have a single respondent give a verbal recall of the household's expenditures. When the diary and recall methods were used to collect data on the expenditures of two random samples of households in the same location, the elasticity of per capita food expenditure with respect to household size, given per capita total expenditure, was estimated to be -0.18 when using data collected by recall, but just -0.07 when using data reported in diaries. Similarly, if the expenditure data collected by recall are used, Engel estimates of scale economies appear much larger than if the expenditure data reported in diaries are used.

The most plausible interpretation of these results is that food expenditure data collected with the recall method have measurement errors that are correlated with household size. As household size increases, it becomes increasingly harder for a survey respondent to accurately recall expenditures on food because of the rise in the number of transactions. It is easier to recall expenditures on non-food items because these may be purchased only sporadically. These measurement errors in expenditure data cause a negative bias in the coefficient on household size in food Engel curves. The Engel method of measuring household size economies may mistake these errors in the food expenditures of large households for genuine size economies. However, evidence from other settings is needed before firmer conclusions can be reached, because the difficulties for a respondent recalling the food expenditures or food consumption of a large household may be less in a rural area.

References

Anand, S., and Harris, C. (1994). ‘Choosing a welfare indicator,’ *American Economic Review* 84 (2): 226-231.

Branch, E. R. (1994). ‘The Consumer Expenditure Survey: a comparative analysis.’ *Monthly Labor Review,* 117(Dec): 47-55.

Coulter, F.A.E., Cowell, F.A. and Jenkins, S.P. (1992). ‘Equivalence scale relativities and the extent of inequality and poverty.’ *Economic Journal*, 102(Sept): 1067-1082.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: Johns Hopkins University Press.

Deaton, A., and Paxson, C. (1998). ‘Economies of scale, household size, and the demand for food.’ *Journal of Political Economy,* 106(5): 897-930.

Dreze, J., and Srinivasan, P. (1997) ‘Widowhood and poverty in rural India: some inferences from household survey data.’ *Journal of Development Economics* 54 (2): 217-234.

Kemsley, W., and Nicholson, J. (1960). ‘Some experiments in methods of conducting Family Expenditure Surveys.’ *Journal of the Royal Statistical Society* 123 (3): 307-328.

Lancaster, G., Ray, R., and Valenzuela, M. (1999). ‘A cross-country study of equivalence scales and expenditure inequality on unit record household budget data.’ *Review of Income and Wealth* 45(4): 455-482.

Lanjouw, P. and Ravallion, M. (1995). 'Poverty and household size.' *The Economic Journal,* 105(Nov): 1415-1434.

Lanjouw, P., Milanovic, B., and Paternostro, S. (1998). 'Poverty in the transition countries: a case of children pitted against the elderly?' *mimeo* Development Economics Research Group, The World Bank.

McWhinney, I., and Champion, H. (1974). 'The Canadian experience with recall and diary methods in consumer expenditure surveys.' *Annals of Economic and Social Measurement,* 3(2): 411-435.

Republic of China (1990). *Report on the Survey of Personal Income Distribution in Taiwan Area of the Republic of China,* Taipei: Directorate-General of Budget, Accounting and Statistics.

Rodgers, W., Brown, C. and Duncan G. (1993). 'Errors in survey reports of earnings, hours worked, and hourly wages.' *Journal of the American Statistical Association,* 88(424): 1208-1218.

Scott, C. and Amenuvegbe, B. (1991). 'Recall loss and recall duration: an experimental study in Ghana.' *Inter-Stat,* 4(1): 31-55.

Sudman, S. and Bradburn, N. (1973). 'Effects of time and memory factors on response in surveys.' *Journal of the American Statistical Association,* 68(344): 805-815.

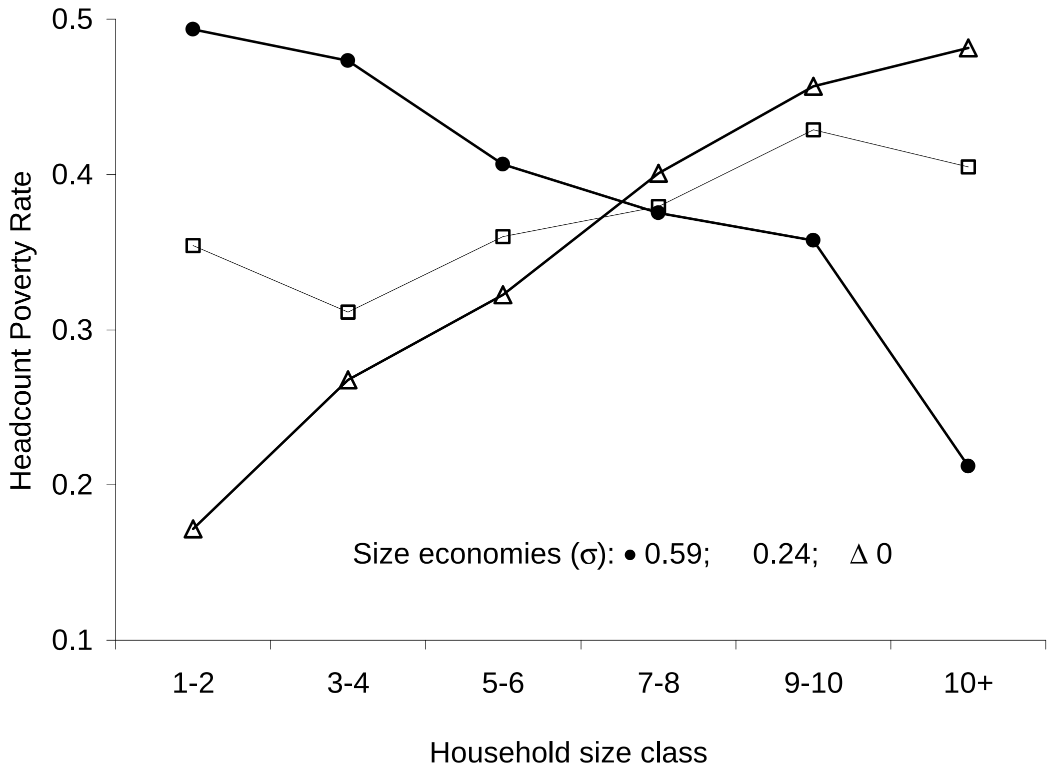

Figure 1: Poverty and household size (headcount index)

Table 1: Monte Carlo Results for Food Share Model

		Experiment 1: Independent measurement errors $v \sim N(0, s_v^2)$			
		No error	$s_v = 0.1$	$s_v = 0.2$	$s_v = 0.3$
1a.	$E(\hat{b})$	-0.1379	-0.1344	-0.1241	-0.1082
1b.	$E(\hat{g})$	-0.0073	-0.0047	0.0030	0.0146
1c.	$E(\hat{s})$	0.0518	0.0339	-0.0254	-0.1377
		Experiment 2: Errors correlated with true values $v = j \ln x_f + e$, $e \sim N(0,0.4)$			
		no error	j = -0.1	j = -0.2	j = -0.3
2a.	$E(\hat{b})$	-0.1379	-0.1282	-0.0940	-0.0560
2b.	$E(\hat{g})$	-0.0073	-0.0383	-0.0448	-0.0331
2c.	$E(\hat{s})$	0.0518	0.2986	0.4763	0.5904
			j = 0.1	j = 0.2	j = 0.3
2d.	$E(\hat{b})$		-0.0728	-0.0221	0.0001
2e.	$E(\hat{g})$		0.0547	0.0706	0.0543
2f.	$E(\hat{s})$		-0.7594	-3.3308	1636.7
		Experiment 3: Errors correlated with household size $v = l \ln n + e$, $e \sim N(0,0.4)$			
		no error	l = -0.1	l = -0.2	l = -0.3
3a.	$E(\hat{b})$	-0.1379	-0.1263	-0.1262	-0.1242
3b.	$E(\hat{g})$	-0.0073	-0.0289	-0.0582	-0.0844
3c.	$E(\hat{s})$	0.0518	0.2282	0.4603	0.6792
			$l = 0.1$	$l = 0.2$	$l = 0.3$
3d.	$E(\hat{b})$		-0.1200	-0.1142	-0.1072
3e.	$E(\hat{g})$		0.0357	0.0686	0.1003
3f.	$E(\hat{s})$		-0.2999	-0.6039	-0.9393

Note:

Results based on 10,000 replications of the model: $w_f = a + b \ln(x/n) + g \ln n + u$.

The true values are a=1.6, b=-0.14, g=-0.007, and $s = g / b$, so the implied true value is s=0.05

Each series is 1000 observations.

Table 2: Description of the Data (N = 212)

	Diary Sample		Recall Sample		
Variable	Mean	Std. Deviation	Mean	Std. Deviation	t-test for equal means
Food budget share	0.5085	0.1677	0.4502	0.1610	2.83**
ln (food expenditure per capita)	6.9603	0.7055	6.7307	0.7687	3.00**
ln (non-food expenditure per capita)	6.9210	1.1831	6.9441	0.9921	0.21
ln (per capita expenditure)	7.6998	0.8969	7.5958	0.8389	1.25
ln (household size)	1.6843	0.6612	1.7394	0.6954	0.83
rm06	0.0884	0.1226	0.0776	0.1088	0.60
rf06	0.0918	0.1211	0.0985	0.1237	0.39
rm714	0.1122	0.1365	0.0787	0.1029	2.55**
rf714	0.0721	0.0997	0.0933	0.1165	0.93
rm1550	0.3417	0.2573	0.3735	0.2351	0.75
rf1550	0.2529	0.1543	0.2450	0.1567	0.29
rm51+	0.0352	0.0799	0.0122	0.0364	1.68
rf51+	0.0058	0.0262	0.0211	0.0845	1.16
Adult employment rate	0.6159	0.2550	0.6037	0.2702	0.43
School years of adults	8.3614	3.8887	8.3071	3.7467	0.21
Age of household head	37.6023	10.3757	36.7177	8.8388	0.64

Note: Variables beginning with *r* are demographic ratios, so that e.g., rf714 is the ratio of the number of females aged 7-14 to total household numbers.

Means and standard deviations are calculated using household sampling weights.

The *t*-test uses standard errors corrected for clustering, sampling weights and stratification.

** =significant at p<0.05 (2 sided).

Table 3: OLS Estimates of the Food Engel Curve

	Diary Sample		Recall Sample		*t*-test for equal
Explanatory variable	coefficient	|*t*|	coefficient	|*t*|	coefficients
ln (per capita expenditure)	-0.1329	9.26	-0.1328	4.66	0.01
ln (household size)	-0.0026	0.17	-0.1262	2.73	2.64
rm06	0.3998	1.08	0.1991	1.16	0.50
rf06	0.2304	0.63	0.1657	1.14	0.17
rm714	0.0766	0.22	0.1301	0.64	0.14
rf714	0.4232	1.00	0.2813	1.82	0.36
rm1550	0.3228	0.96	0.3034	3.68	0.06
rf1550	0.3532	0.99	0.1952	1.25	0.45
rm51+	0.3406	0.88	0.8271	2.15	0.98
Adult employment rate	0.1207	2.41	0.0123	0.19	1.24
Constant	1.1788	3.32	1.4919	5.34	0.67
R^2	0.5457		0.4008		
$F_{(12,18)}$	30.46		7.86		

Note: Variables beginning with *r* are demographic ratios, so that e.g., rf714 is the ratio of the number of females aged 7-14 to total household numbers. The omitted group is elderly females. Models also contain two quarterly dummies.

Reported absolute *t*-values are corrected for clustering, sampling weights and sample stratification.

$F_{(12,18)}$ is an adjusted Wald (*W*) test for zero slopes: $\frac{d-k+1}{kd}W \sim F(k, d-k+1)$, where *d* is the number of clusters minus the number of strata (29), and *k* is the number of slope variables.

Table 4: IV Estimates of the Food Engel Curve

Explanatory variable	Diary Sample coefficient	Diary Sample |*t*|	Recall Sample coefficient	Recall Sample |*t*|	*t*-test for equal coefficients
ln (per capita expenditure)	-0.1536	6.70	-0.1039	2.73	1.12
ln (household size)	-0.0147	0.81	-0.1049	2.42	1.77
rm06	0.3513	0.97	0.2335	1.23	0.29
rf06	0.1685	0.48	0.2058	1.28	0.10
rm714	0.0411	0.12	0.1853	0.83	0.39
rf714	0.3559	0.81	0.2928	1.84	0.16
rm1550	0.2895	0.88	0.3354	3.61	0.15
rf1550	0.3067	0.89	0.2048	1.27	0.30
rm51+	0.2876	0.77	0.8862	2.29	1.28
Adult employment rate	0.1265	2.43	0.0154	0.22	1.23
Constant	1.3958	3.47	1.2046	3.34	0.34
R^2	0.5385		0.3894		
$F_{(12,18)}$	22.42		7.07		
Durbin-Wu-Hausman test	t=1.06		t=1.35		
Over-identification test	$\chi^2_{(2)}$=0.25		$\chi^2_{(2)}$=0.02		

Notes: See Table 3.

Instruments for ln (per capita expenditure) are the average number of school years of each adult in the household and the age of the household head.

Notes

[1] Anand and Harris (1994) review the empirical performance of various welfare indicators and conclude that the food share has little merit as an overall welfare indicator.

[2] The respondent was usually the senior female, especially for food expenditures. However, for certain items like alcohol and tobacco, the senior male usually provided the information. The principle of asking the respondent best informed to answer questions regarding each type of consumption is widely used by recall surveys.

[3] With an average food share of 0.5, a 20 percent decrease in per capita food expenditures is implied.

[4] It is perhaps no coincidence that the two countries in Deaton and Paxson's sample with the least puzzling results (Britain and France) collect household expenditures using the diary method, while the other countries gather data by recall.

[5] A household with two people makes an average of 25 non-food purchases per fortnight while a household with 10 people averages 50 non-food purchases per fortnight.

[6] This result is also found by the field survey described below. The foods whose expenditures recalled by respondents were lowest, compared with expenditures reported in diaries, were the foods that were most frequently purchased, and this correlation is significant at the $p<0.01$ level. An alternative hypothesis, that recall errors reflect respondent fatigue, is not supported because there was no significant association between the placement of a food in the questionnaire and the size of the recall error.

[7]For example, the "intergenerational family" (the head and their spouse, the parents and parents-in-law of the head, children of the head and their spouses, and grandchildren of the head) comprises 78 percent of the residents of the average household in urban areas of Papua New Guinea. A unit increase in the logarithm of household size raises the share of residents who are not part of the intergenerational family by 14 percentage points.

[8] By rewriting $b\ \ln(x/n^{1-s})$ as $b \ln x-(1-s)b \ln n$ it is clear that $b\ \ln(x/n)+bs\ \ln n = b\ \ln(x/n^{1-s})$.

[9] Rodgers, Brown, and Duncan (1993) report evidence that errors in survey responses are not independent of either the true values of the variable being measured or of other explanatory variables in the causal model.

[10]These values were drawn afresh at each experiment. Values for household size were rounded to the nearest integer, with the minimum constrained to n=1.

[11] The mean food budget share generated by these assumptions was 0.50 (standard deviation=0.16). The means and standard deviations of the generated x, n, and w_f variables match the data collected in the field.

[12] Deaton and Parxon use IV to deal with random measurement errors in ln (x/n) that might bias the g coefficient because of the correlation between ln (x/n) and ln n.

[13] Illiterates were typically the elderly, so sometimes other family members recorded in the diaries on their behalf or interviewers visited more frequently than usual (roughly every second day) and did the recording.

[14] This is also true when the unweighted data are used. For the recall sample, the OLS coefficient on household size is –0.1000 and the IV coefficient is –0.0928. For the diary sample, the OLS coefficient is -0.0061 and the IV coefficient is –0.0237.

[15] The size classes are used to give smoother estimates because of the small sample size. For the same reason, the poverty rates are calculated from all areas of Papua New Guinea rather than just from the capital city.

[16] Other methods based on estimation of a system of demand equations (for example, see Lancaster *et. al.*, 1999) may be too involved for many applied economists who just want to estimate an equivalence scale as an input into some other modeling task.