Public Disclosure Authorized

Public Disclosure Authorized

WPS6372

Policy Research Working Paper 6372

What Does Variation in Survey Design Reveal about the Nature of Measurement Errors in Household Consumption?

John Gibson
Kathleen Beegle
Joachim De Weerdt
Jed Friedman

The World Bank
Development Research Group
Poverty and Inequality Team
February 2013

Policy Research Working Paper 6372

Abstract

This paper uses data from eight different consumption questionnaires randomly assigned to 4,000 households in Tanzania to obtain evidence on the nature of measurement errors in estimates of household consumption. While there are no validation data, the design of one questionnaire and the resources put into its implementation make it likely to be substantially more accurate than the others. Comparing regressions using data from this benchmark design with results from the other questionnaires shows that errors have a negative correlation with the true value of consumption, creating a non-classical measurement error problem for which conventional statistical corrections may be ineffective.

This paper is a product of the Poverty and Inequality Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at kbeegle@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

What Does Variation in Survey Design Reveal about the Nature of Measurement Errors in Household Consumption?

John Gibson, University of Waikato
Kathleen Beegle, World Bank
Joachim De Weerdt, EDI Tanzania
Jed Friedman, World Bank

JEL: C21, C81, D12

Keywords: Consumption, Engel curves, Household surveys, Measurement error

Sector board: Poverty Reduction (POV)

Acknowledgements: This work was supported by the World Bank Research Committee. We would like to thank Bonggeun Kim and seminar participants at Hitotsubashi and Monash for very useful comments.

'Measurement error is an ever-present, generally significant, but usually neglected, feature of survey based income and expenditure data.' Chesher and Schluter (2002: 377)

I. Introduction

Household consumption surveys are at the heart of attempts to measure living standards, especially in developing countries where labor and income surveys cover neither the full population nor all economic activity. All surveys likely suffer from measurement error but the complexity of consumption surveys makes them especially prone to error. Yet surprisingly, validation studies to describe the nature and consequences of measurement error are mainly for labor market settings, using either administrative data (Bound and Krueger, 1991) or firm records (Pischke, 1995) as their measure of truth. Such validation studies also are limited to rich countries, including for the one type of consumption – out of pocket health expenditure – for which validation has been attempted (Cohen and Carlson, 1994).

This note reports on a consumption survey experiment in Tanzania that is potentially informative about the nature of measurement error in developing country consumption data. We randomly assigned eight alternative survey designs to 4,000 households (three households per design in each of 168 sampling units), covering urban and rural areas.[1] Validation from store data is neither feasible nor useful in this setting, but one resource intensive design – individually kept diaries with daily supervision for 14 days – represents a "gold standard" for consumption measurement in developing countries. The cost per household for this design was almost ten times the cost to survey consumption by recall interview (Beegle, *et al.* 2012), indicative of the resources needed for careful tracking of all commodity in-flows (harvests, purchases, gifts, stock reductions), out-flows (sales, gifts, stock increases, food fed to animals), daily attendance at meals, and acquisitions and disposals by children and other dependents the diary-keeper reported on. Survey diagnostics, such as the time profile for diary entries and mean daily consumption, show no diary fatigue for this "gold standard" module and just three households that started a diary did not complete their final interview.[2]

We deem the data from the individually-kept diary a suitable benchmark for assessing measurement error in data from the other survey designs. Beyond random assignment, several steps were taken to ensure that differences in measured consumption are solely due to survey design. The experiment was conducted by small, intensively supervised teams of experienced

[1] Full details on the design and implementation of the experiment are reported in Beegle *et al.* (2012).

[2] Across the full experiment, of the original 4032 households assigned a module, only 13 were replaced due to refusal. The final achieved sample was 4025.

interviewers, with fieldwork spread over 12 months to balance each module over both time and space. Each interviewer implemented all eight modules in equal proportion in order not to confound module effects with interviewer effects. To prevent potentially uncontrollable cross-module spillovers within households, each household faced just a single survey design.[3]

Since the 'gold standard' measure of consumption is not available for each household in the experiment, we use two indirect methods to study the nature of measurement errors in these data.[4] First, we average the logarithm of total household consumption, for each sampling unit (hereafter 'village') and each survey design. To the extent that measurement errors in consumption are random, averaging should reduce bias, with estimates converging to the true village average. We find that errors are not random. If the error is calculated by subtracting the benchmark average from the average calculated with each of the other seven survey designs, the error is found to be negatively correlated with the benchmark (true) village average. Exactly the same pattern, found with earnings data, is referred to by Bound and Krueger (1991) as *mean-reverting measurement error*.

Since economists rarely work with village averages, our second exercise for assessing the nature and consequences of the measurement error is to estimate a widely used regression on household consumption data – a food Engel curve. We treat the estimates from the intensively supervised sample given the individually-kept diaries as being closest to what true consumption would reveal. We then compare with coefficients estimated using the data from the other designs, and with patterns of bias expected from simulated random and non-random measurement errors. The food Engel curve is useful because of extant results on expected biases from different types of measurement error that are motivated by an Engel curve puzzle raised by Deaton and Paxson (1998).[5] This exercise also shows that the most plausible model of

---

[3] Most previous attempts to compare different consumption survey designs apply them sequentially to the same household (e.g., Ahmed et al, 2010). Sequential designs suffer not only from lack of balance in timing, but also from *conditioning bias*, whereby respondents who previously recorded in diaries might be atypically accurate in a recall interview, while those initially surveyed by recall might shirk at the prospect of daily recording in diaries, making these respondents uninformative about how these modules would perform in practice.

[4] In the typology of validation studies provided by Bound *et al.* (2001: 3743) our methods would be considered 'macro-level comparisons' of survey estimates versus estimates generated under preferred survey conditions. None of the problems listed by Bound *et al*. for such comparisons apply to our experiment given the balance we achieved over time, space, and samples.

[5] The puzzle is that food shares fall as household size rises at constant per capita consumption. The effective income increase from sharing public goods in larger households should outweigh the substitution effect (public goods are cheaper in large households) so food shares should rise for the

measurement error in these consumption data is that errors are negatively correlated with the true value of consumption.

Negatively correlated errors matter because they bias regression coefficients even if present in just the dependent variable. Surveyed consumption is often an outcome measure in impact evaluations, so negatively correlated errors may lead to understated impacts. Household consumption is also used as a key explanatory variable in many studies as a proxy for permanent income. Negatively correlated errors in an explanatory variable may bias regression coefficients either toward or away from zero. In contrast, random (classical) errors cause no bias when just in the dependent variable and always attenuate the coefficient on a single error-ridden explanatory variable. Moreover, economists' main correction for measurement error bias – instrumental variables (IV) – is inconsistent when errors are correlated with true values (Black, Berger and Scott, 2000), while bounding estimates based on reverse regression are unlikely to be effective in practice (Gibson and Kim, 2010).

II. Motivation: Effects of Random and Non-random Measurement Errors

Consider some true (bivariate) model:

$$y = \alpha + \beta x + u \tag{1}$$

for an outcome of interest *y*, an independent variable *x*, which may be a binary treatment variable, a response coefficient β, and a pure random error, *u*. Let measurement error cause the observed value of the dependent variable, y^* to be related to the true value by:

$$y^* = \theta + \lambda y + v. \tag{2}$$

The textbook case of classical measurement error places stringent restrictions on equation (2), specifically that $\theta = 0$, $\lambda = 1$ and $E(v) = \text{cov}(y,v) = \text{cov}(x,v) = \text{cov}(u,v) = 0$, so that just white noise is added to the true value. In contrast to this widely used assumption, validation

poor. But Deaton and Paxson find the most negative effects of household size in their Engel curves for poor countries. Nothing in the current note resolves the puzzle, since there is a significant negative effect of household size with all survey designs. But the prior literature on the effect of measurement error on the food Engel curve (Gibson, 2002; Gibson and Kim, 2007; Ahmed e*t al.* 2010) provides diagnostics for how different types of errors affect the estimated coefficients.

studies of labor survey data find that $0 < \lambda < 1,$ which Bound and Krueger (1991) call mean-reverting measurement error.

The estimator of the response coefficient with the error-ridden dependent variable is:

$$\beta_{y^* x} = \frac{\text{cov}(y^*, x)}{\text{var}(x)} = \frac{\text{cov}(\lambda\alpha + \lambda\beta x + \lambda u - v, x)}{\text{var}(x)} = \lambda\beta \tag{3}$$

showing that in the special case of classical errors there is no bias in the response coefficient. But if measurement errors negatively co-vary with true values $(0 < \lambda < 1)$ the estimated response coefficient will be attenuated. Hence, knowing what type of measurement error afflicts consumption data is likely to be important for empirical research in poor countries. For example, many studies of program impacts use household consumption as an outcome measure (Khandker, 2005). Treatment effects may be greatly understated by these studies if measurement errors in household consumption are negatively correlated with true values.[6]

Consider next the case of no error in the dependent variable (or just white noise error) while the observed value for the independent variable, x^*, is related to the true value by:

$$x^* = \theta + \lambda x + v. \tag{4}$$

The estimator of the response coefficient in equation (1) is then:

$$\beta_{yx^*} = \frac{\text{cov}(y, x^*)}{\text{var}(x^*)} = \frac{\text{cov}(\alpha + \frac{\beta}{\lambda} x^* - \frac{\beta\theta}{\lambda} - \frac{\beta}{\lambda} v + u, x^*)}{\text{var}(x^*)} = \beta \frac{\lambda\sigma_x^2}{\lambda^2\sigma_x^2 + \sigma_v^2} \tag{5}$$

With classical measurement error, where $\lambda = 1,$ the rescaling of the response coefficient is the familiar attenuation in proportion to the explanatory variable's 'reliability ratio' (the variance in the true data relative to that in the mis-measured data, $\sigma_x^2 / [\sigma_x^2 + \sigma_v^2]$). But with mean-

[6] Understated treatment effects are likely even in panel studies. Validation of panel labor surveys shows that just as errors in reported earnings levels negatively co-vary with true values, so too do errors in reported changes in earnings negatively co-vary with true changes, causing understated estimates of growth-related processes (Gibson and Kim, 2010). Validation studies for panel consumption surveys appear to be a long way off.

reverting error, attenuation is not guaranteed. If the 'shrinkage' of the variance in the first term in the denominator due to multiplying by λ^2 (for $0<\lambda<1$) exceeds the effect of adding the variance of the random noise term (σ_v^2), the response coefficient is overstated rather than understated. Also, the variance of the error-ridden variable is less than that of the true variable, contrary to what is possible with the reliability ratio interpretation.

Knowing if the coefficient on the proxy for permanent income is either attenuated or exaggerated seems important for many of the policy conclusions that economists may draw from regressions where surveyed household consumption is a key explanatory variable. Moreover, when economists attempt to deal with possible attenuation by using instrumental variables for household consumption (e.g., Alderman et al, 2006) they are using an estimator that is inconsistent when measurement errors are correlated with true values (Black *et al.* 2000). Once again, knowing more about the nature of measurement error in household consumption data may help improve modeling practice and interpretation of empirical results.

III. The Survey Experiment and Evidence from Village Averages

The designs of the eight alternative consumption modules are described in Table 1. They differ by method of data capture (diary versus recall), by respondent (individual versus household reporting), by recall period (7 day, 14 day and usual month), and by number of items in the recall list. The designs were strategically selected to reflect the most common consumption modules used in multi-topic living standards surveys in developing countries (which typically seek less commodity detail than do household budget surveys). Design variation was restricted to foods and frequently purchased non-foods, but assignment to different modules also affects reports on infrequently purchased items, which were covered by a set of annual recall questions common to all households. Beegle et al (2012) ascribe this impact on the infrequent items as due to either respondent conditioning or fatigue, since questions on infrequent items came after the lengthy food and frequent non-food recall sections in modules 1-5 and after the two-week diary for modules 6-8.

Table 1. Survey Experiment Consumption Modules

Module	Description	Details	Number of Households
1	Long list (58 food items) 14 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	503
2	Long list (58 food items) 7 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
3	Subset list (17 food items; subset of 58 foods), scaled by 1/0.77[a] 7 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
4	Collapsed list (11 food items covering universe of food categories) 7 day	Tshilling value of consumption	504
5	Long list (58 food items) Usual 12 month	Consumption from purchases: number of months consumed, quantity per month, Tshilling value per month Consumption from own-production: number of months consumed, quantity per month, Tshilling value per month Consumption from gifts/other sources: total estimated value for last 12 months	504
6	Household diary, frequent visits 14 day diary		502
7	Household diary, infrequent visits 14 day diary		501
8	Individually-kept diary, frequent visits 14 day diary		503
			4,025

Notes: Frequent visits entailed daily visits by the local assistant and visits every other day by the survey enumerator for the duration of the 2-week diary. Infrequent visits entail 3 visits: to deliver the diary (day 1), to pick up week 1 diary and drop off week 2 diary (day 8), and to pick up week 2 diary (day 15). Households assigned to the infrequent diary but who had no literate members (about 18 percent of the sub-sample) were visited every other day by the local assistant and the enumerator.

Non-food items are divided into two groups based on frequency of purchase. Frequently purchased items (charcoal, firewood, kerosene/paraffin, matches, candles, lighters, laundry soap, toilet soap, cigarettes, tobacco, cell phone and internet, transport) were collected by 14-day recall for modules 1-5 and in the 14-day diary for modules 6-8. Non-frequent non-food items (utilities, durables, clothing, health, education, contributions, and other; housing is excluded) are collected by recall identically across all modules at the end of the interview (and at the end of the 2-week period for the diaries) and over the identical one or 12-month reference period, depending on the item in question.

[a] The 17 foods account for 77 percent of the food budget, so the measured value of food consumption is scaled up by 1/0.77.

Random assignment successfully balanced over consumption-related characteristics.[7] All modules were fielded within villages at the same time, so no controls for timing or other covariates are used when examining the village averages. These averages reveal reported consumption to be highest for the 'gold standard' design of an intensively supervised, individually-kept diary (Table 2).[8] The variances and ratios reported in the middle columns of Table 2 are inconsistent with the assumptions of classical measurement error. For three of the designs (subset 7-day recall and both household diaries) the variance of the error-ridden variable is less than that of the benchmark variable. This understatement of the variance could not happen if the measurement error was just in the form of white noise.

Table 2. Tests for Non-classical (correlated) Measurement Error in Log Consumption

	Mean	Ratio to Benchmark $\frac{E(\ln \bar{x}_k)}{E(\ln \bar{x}_8)}$	Variance	Ratio to Benchmark $\frac{\text{var}(\ln \bar{x}_k)}{\text{var}(\ln \bar{x}_8)}$	$\hat{\lambda}$ (S.E.)	*t*-test for Correlated errors $H_0: \lambda = 1$ $H_1: \lambda < 1$ $p = \Pr(t < \hat{t})$
1. Long 14 day	14.104	0.987	0.350	1.081	0.569 (0.068) [a]	p =0.000
2. Long 7 day	14.225	0.996	0.337	1.040	0.596 (0.064)	p =0.000
3. Subset 7 day	14.195	0.994	0.320	0.988	0.535 (0.065)	p =0.000
4. Collapsed 7 day	14.039	0.983	0.343	1.060	0.583 (0.066)	p =0.000
5. Long usual month	14.084	0.986	0.423	1.307	0.662 (0.072)	p =0.000
6. HH diary frequent	14.128	0.989	0.289	0.891	0.494 (0.062)	p =0.000
7. HH diary infrequent	14.155	0.991	0.269	0.832	0.422 (0.063)	p =0.000
8. Benchmark (indiv, freq)	14.283	1.000	0.324	1.000		

Notes: The $\hat{\lambda}$ are from separate regressions for each module, where the independent variable is village-averaged log annualized total household consumption from the individually-kept, frequent visit diary (the benchmark).

N=168.

[a] Standard errors in parenthesis.

[7] Beegle *et al*. (2012) find just 13 of 420 pairwise comparisons to have statistically significant differences, at the five percent level, for 15 baseline household characteristics.

[8] Beegle *et al.* (2012) show understatement by the other designs occurs through both food and non-food; average food consumption is statistically significantly lower than the benchmark for six of the modules, average frequent non-food consumption is significantly lower for four of the modules and average non-frequent non-food consumption is significantly lower for three of the modules.

The nature of the measurement is revealed by estimating equation (2) seven times, with village-averaged log total consumption from modules 1 to 7 as the measured dependent variable, y^* and the village average of log consumption from the benchmark individual diary taken as the approximation to the true *y*. We find that $\hat{\theta} > 0$, $\hat{\lambda} < 1$ in all seven regressions, with $\hat{\lambda}$ ranging between 0.42 and 0.66 and always statistically significantly less than one. The three survey designs with lower variance than in the benchmark design have the strongest degree of mean reverting error, with $0.42 > \hat{\lambda} > 0.54$. For these designs, the shrinkage due to the λ^2 term in the denominator of equation (5) outweighs the effect of adding the variance of the random noise term. Furthermore, finding $\hat{\lambda} < 1$ in all regressions implies that there is a negative correlation between the errors and the true values (as proxied by the village average of log consumption from the 'gold standard' design).

IV. Evidence from Food Engel Curve Regressions

Consider a simplified version of the food Engel curve of Deaton and Paxson (1998), where the demographic composition and control variables are ignored to reduce clutter:

$$w_{f,i} = \alpha + \beta \ln\left(\frac{x}{n}\right)_i + \gamma \ln n_i + u_i. \tag{6}$$

The food share for household *i*, $w_{f,i}$ depends on household total consumption, x_i and household size, n_i along with a random disturbance, u_i. The data on log per capita consumption, $\ln(x/n)_i$ are affected by measurement error in x_i and since these errors occur through both food and non-food, the food share is also affected unless there are equi-proportional errors in both consumption components.[9] Without more structure on the nature of the measurement error, it is impossible to analytically sign the direction of bias in $\hat{\beta}$ and $\hat{\gamma}$ because it depends on the relative degree of measurement error in food and non-food.[10]

---

[9] Beegle *et al.* (2012) show equi-proportional errors in food and non-food to be unlikely. Errors in household size affecting per capita consumption are also unlikely; just one module (usual month recall) had slightly significantly different household size than the benchmark module.

[10] Equations (14) and (15) of Gibson and Kim (2007) show the direction of bias in each coefficient depends on the relative magnitude of several terms if the measurement error is allowed to have a general (potentially non-random) nature along the lines of what is described in equation (4) above. If

Our strategy is to estimate equation (6) with the benchmark diary data, and then compare with the coefficients estimated using the data from the other designs. Treating those other designs as more error-ridden, a cross-module comparison shows how measurement error affects $\hat{\beta}$ and $\hat{\gamma}$. We then estimate Engel curves on simulated data with three types of measurement error: (i) random, (ii) negatively correlated with true values, and (iii) negatively correlated with household size, so as to see which type of measurement error best matches the empirically observed cross-module pattern. In these simulations, the parameter values for the Engel curve with error-free data are based on the empirical results from the benchmark diary, in keeping with our maintained assumption that this is closest to truth.

The motive for the first simulation is that random (classical) error is the typical view of measurement error in consumption data. For example, authors using instrumental variables to treat attenuation bias (e.g., Alderman *et al.* 2006) implicitly assume random errors in measured consumption. For the second simulation, the small literature on errors in consumption surveys supports the hypothesis that these errors negatively co-vary with true values since survey reporting tasks become harder as the household gets richer and the consumption pattern more varied (Pradhan, 2001; Ahmed *et al.* 2010). For the third simulation, a negative correlation with household size is plausible since one person often reports on behalf of the household, and larger households consume more within a given period, increasing the reporting burden. For example, Gibson (2002) finds the understatement by recall surveys especially apparent for the food consumption of larger households.

---

the more restrictive assumptions of classical measurement error are used, $\hat{\gamma}$ is biased upwards but no result is reported for $\hat{\beta}$.

Table 3: OLS Coefficient Estimates and Hypothesis Test Results for the Food Engel Curves

	Consumption Survey Module Number:							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: No other covariates[a]								
ln per capita cons	-0.055***	-0.084***	-0.100***	-0.076***	-0.074***	-0.091***	-0.077***	-0.056***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
ln household size	-0.063***	-0.051***	-0.073***	-0.053***	-0.067***	-0.069***	-0.059***	-0.061***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)
Observations	503	504	504	504	504	502	501	503
Adjusted-R^2	0.299	0.286	0.380	0.351	0.286	0.354	0.324	0.308
p-value $H_0: \hat{\beta}_k = \hat{\beta}_8$	0.909	0.027	0.001	0.134	0.170	0.001	0.092	0.003[b]
p-value $H_0: \hat{\gamma}_k = \hat{\gamma}_8$	0.892	0.454	0.392	0.553	0.678	0.553	0.871	0.742[b]
Panel B: Including other covariates								
ln per capita cons	-0.039***	-0.068***	-0.083***	-0.060***	-0.056***	-0.059***	-0.059***	-0.026***
	(0.008)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
ln household size	-0.072***	-0.043***	-0.046***	-0.058***	-0.048***	-0.050***	-0.044***	-0.037**
	(0.013)	(0.013)	(0.015)	(0.014)	(0.016)	(0.014)	(0.013)	(0.015)
share of kids <6	0.102**	0.019	0.029	0.066	-0.036	0.073	0.007	0.010
	(0.040)	(0.042)	(0.041)	(0.043)	(0.046)	(0.046)	(0.041)	(0.043)
share of kids 6-15	0.117***	0.064*	0.049	0.049	0.044	0.085**	0.007	0.058
	(0.034)	(0.033)	(0.036)	(0.037)	(0.038)	(0.036)	(0.033)	(0.037)
share of elderly	0.011	0.038	0.079*	0.014	0.048	0.098**	0.066	0.075**
	(0.034)	(0.036)	(0.040)	(0.042)	(0.042)	(0.039)	(0.041)	(0.036)
Head is female	-0.029*	-0.039**	-0.060***	-0.017	-0.032*	-0.040**	-0.006	0.011
	(0.017)	(0.017)	(0.017)	(0.018)	(0.019)	(0.019)	(0.019)	(0.020)
Head's age	0.001**	0.000	-0.001	0.001	-0.000	0.000	0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Marital status	0.011**	0.006	0.020***	-0.001	0.007	0.011**	-0.000	-0.002
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Head school years	-0.006***	-0.008***	-0.008***	-0.006***	-0.008***	-0.011***	-0.007***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	503	504	504	504	504	502	501	503
Adjusted-R^2	0.360	0.332	0.438	0.369	0.318	0.430	0.351	0.386
p-value $H_0: \hat{\beta}_k = \hat{\beta}_8$	0.310	0.002	0.000	0.016	0.032	0.019	0.016	0.002[b]
p-value $H_0: \hat{\gamma}_k = \hat{\gamma}_8$	0.078	0.763	0.667	0.324	0.632	0.553	0.731	0.773[b]

[a] Except for month and district fixed effects which are included in all models in this table.

[b] Joint test that the coefficients are equal across all eight columns.

Sample sizes for each module are reported in Table 1. Standard errors in parentheses; *, ** and *** denote statistical significance at 10%, 5% and 1% level. The consumption modules that match the numbers in the column headings are:

1 Long list (58 items), 14 day
2 Long list (58 items), 7 day
3 Scaled subset list (17 items), 7 day
4 Collapsed list (11 items), 7 day
5 Usual month (58 items)
6 Household diary, Frequent
7 Household diary, Infrequent
8 Individually-kept diary, Frequent

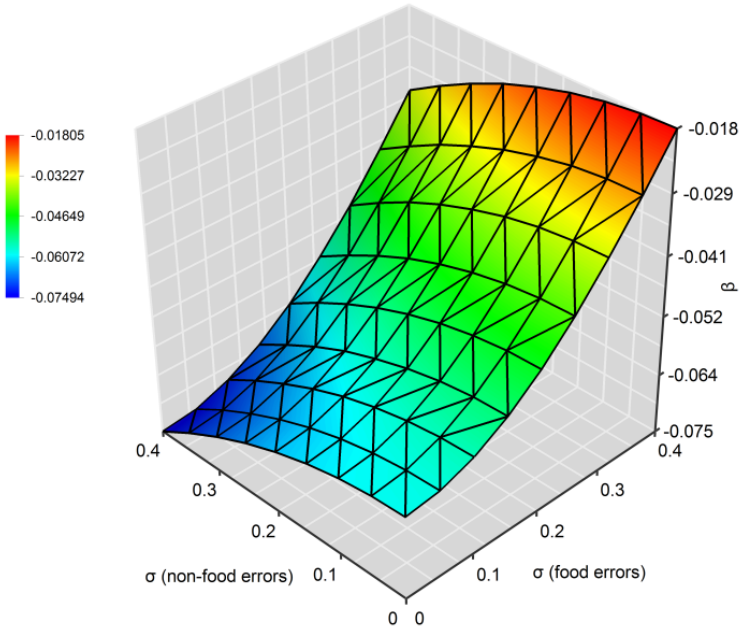
There is significant variation in $\hat{\beta}$ over the different consumption modules (Table 3). The coefficient on log per capita consumption in the benchmark module is -0.056 (in panel A, with demographics and household head characteristics excluded). In three of the other modules, $\hat{\beta}$ is statistically significantly more negative, ranging from -0.084 to -0.100. The hypothesis of equal $\hat{\beta}$ across all eight modules is strongly rejected (p=0.003). In contrast, there is no module effect on $\hat{\gamma}$, which ranges between -0.051 and -0.073 with no statistically significant differences. The significant impact on $\hat{\beta}$ and lack of impact on $\hat{\gamma}$ strengthens when covariates are included (Table 3, panel B).[11] Except for module 1, all other modules give a significantly more negative $\hat{\beta}$ than the benchmark, with p-values that range from 0.000 to 0.032. In contrast, there are no statistically significant effects on $\hat{\gamma}$.

What type of measurement errors are consistent with this pattern of a significantly more negative $\hat{\beta}$ and little impact on $\hat{\gamma}$? The simulation results in Figure 1 show that it requires errors in food consumption to be more strongly (negatively) correlated with true values than are the measurement errors in nonfood consumption.[12] This combination of errors causes the $\hat{\beta}$ estimated from the error-ridden module to be more negative than the $\hat{\beta}$ estimated with the benchmark survey, while leaving $\hat{\gamma}$ largely unchanged. For example, a simulated error process where errors in food consumption are generated from a regression on true food consumption with a coefficient of $\varphi = -0.3$ while errors in non-food are generated from a regression on true non-food consumption whose coefficient is just $\psi = -0.05$ gives results like those observed in the Table 3 regressions, with $\hat{\beta} = -0.082$ and $\hat{\gamma} = -0.068$.

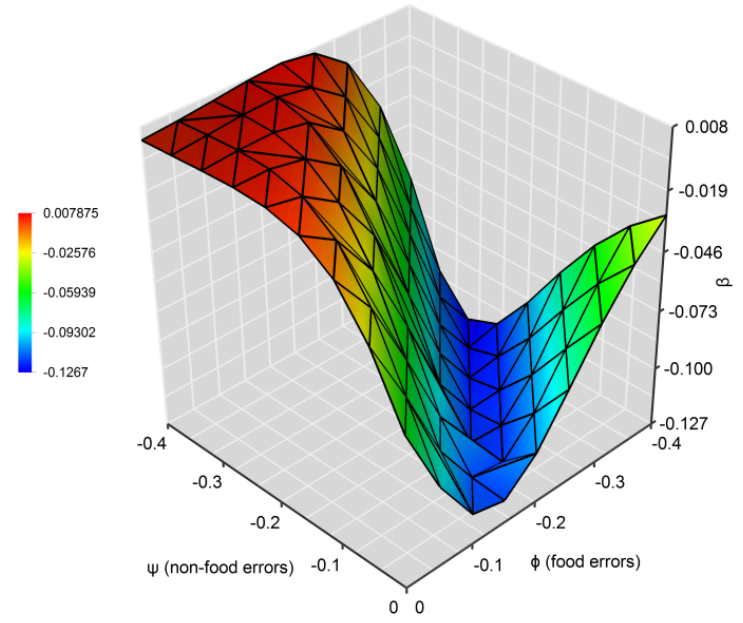
[11] We include three demographic ratios (the share of children aged less than six, children aged six to fifteen, and elders aged over 65), and the age, education, gender and marital status of the household head.

[12] Appendix A describes the Monte Carlo experiments yielding the simulations illustrated in Figure 1.

Panel A
Random Measurement Errors in Food and Non-Food

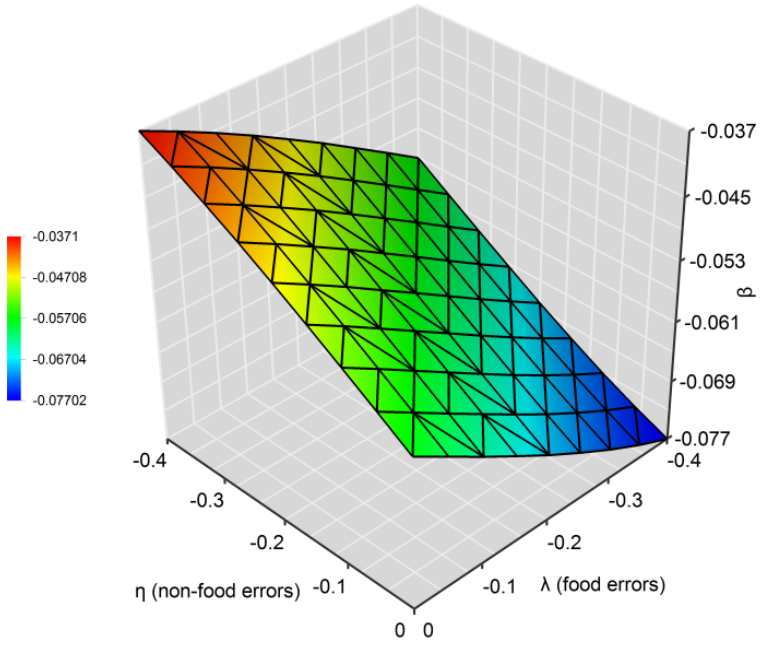


Panel B
Errors in Food and Non-Food Negatively Correlated with True Values



Panel C

Errors in Food and Non-food Negatively Correlated with Household Size



The simulations also rule out the other hypotheses of either random errors or errors negatively correlated with household size. In panel A of Figure 1, larger random errors in measuring food consumption make $\hat{\beta}$ less negative rather than more negative. It takes error-free food consumption and large, random, non-food errors to make $\hat{\beta}$ more negative, but this effect is never strong enough in the simulations to cause the $-0.10 < \hat{\beta} < -0.08$ observed in the Table 3 regressions. Similarly, no simulated correlation with household size is strong enough to yield $\hat{\beta}$ values in Panel C that are negative enough to match those from the Table 3 regressions on data from the more error-ridden survey modules, even with simulated error-free non-food consumption. Moreover, there is a high correlation (r=0.99) between $\hat{\gamma}$ and $\hat{\beta}$ in the simulation results (so just patterns for $\hat{\beta}$ are shown since they are so similar for $\hat{\gamma}$); setting the food errors at their largest values and simulating error-free non-food data so as to make $\hat{\beta}$ as negative as possible causes $\hat{\gamma}$ to be much more negative than is observed in any of the Table 3 regressions ($\hat{\gamma} < -0.15$).

V. Conclusions

In this paper we examined measurement errors in household consumption data. This is inherently difficult due to lack of a gold standard for comparing with survey estimates so as to reveal the nature of the errors. Nevertheless we provide indirect evidence on the nature of measurement errors using two different regression approaches and an experiment where eight different consumption questionnaires were randomly assigned to households. The results are most consistent with errors in measured consumption that are negatively correlated with true values, especially for food. Such a correlation with true value is likely because for most surveys reporting tasks become harder as the household gets richer and the consumption pattern more varied. A negative correlation with true values implies mean reversion, so even when mis-measured consumption is a dependent variable there may be bias in regression coefficients, and when consumption is an explanatory variable the usual attenuation bias may not apply. Both cases should be of serious concern to economists who rely on accurate household consumption data for their measuring and modeling of living standards.

Appendix A. The Monte Carlo Experiments

The Monte Carlo experiments use 10,000 replications of the model: $w_f = \alpha + \beta \ln(x/n) + \gamma \ln n + u$, where α = 1.4, β = -0.06, γ = -0.06 and each series is 1000 observations. Parameter values match the results using data from the benchmark diary, in column (8) of Table 3. To implement the experiments, total consumption, x was partitioned into food consumption, $x_F = x \cdot w_F$ and non-food consumption, $x_{NF} = x - x_F$ and three different types of errors were added to food (v_F) and non-food (v_{NF}) and the error-ridden total expenditure and food share variables were reconstructed as $\tilde{x} = \tilde{x}_F + \tilde{x}_{NF}$ and $\tilde{w} = \tilde{x}_F / \tilde{x}$, before the food Engel curve regressions were re-estimated.

In case (1) the measurement errors were independent of any of the variables in the model and of each other, with $v_F \sim N(0, \sigma^2_{v_F})$, and $v_{NF} \sim N(0, \sigma^2_{v_{NF}})$. The errors, σ_{v_F} and $\sigma_{v_{NF}}$ took each of nine values, ranging from 0, 0.05, 0.1, 0.15, …, 0.35, 0.4.

In case (2) the errors were correlated with true values, $v_F = \varphi \ln x_F + \varepsilon$ and $v_{NF} = \psi \ln x_{NF} + \varepsilon$ where $\varepsilon \sim N(0,0.1)$ and the values used for φ and ψ were 0, -0.05, -0.1, -0.15, …, -0.35, -0.4.

In case (3) errors were correlated with household size, n: $v_F = \lambda \ln n + \varepsilon$ and $v_{NF} = \eta \ln n + \varepsilon$ where $\varepsilon \sim N(0,0.1)$ and the values used for λ and η were 0, -0.05, -0.1, -0.15, …, -0.35, -0.4.

References

Ahmed, A., Brzozowski, M. and Crossley, T. (2010). 'Measurement errors in recall food consumption data.' *Mimeo* University of Cambridge.

Alderman, H., Hoogeveen, H. and Rossi, M. (2006) 'Reducing child malnutrition in Tanzania: Combined effects of income growth and program interventions.' *Economics and Human Biology* 4(1): 1-23.

Beegle, K., de Weerdt, J., Friedman, J., and Gibson, J. (2012). 'Methods of household consumption measurement through surveys: Experimental results from Tanzania.' *Journal of Development Economics* 98(1): 3-18.

Black, D., Berger, M., and Scott, F. (2000). 'Bounding parameter estimates with nonclassical measurement error.' *Journal of the American Statistical Association* 95(451): 739-748.

Bound, J., and Krueger, A. (1991). 'The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?' *Journal of Labor Economics* 9(1): 1-24.

Bound, J., Brown, C., and Mathiowetz, N. (2001). 'Measurement error in survey data.' In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics: Volume 5*, Elsevier, pp. 3705-3843.

Chesher, A., and Schluter, C. (2002). 'Welfare measurement and measurement error.' *Review of Economic Studies* 69(2): 357-378.

Cohen, S., and Carlson, B. (1994). 'A comparison of household and medical provider reported expenditures in the 1987 NMES.' *Journal of Official Statistics* 10(1):3-29.

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

Gibson, J. (2002). 'Why does the Engel method work? Food demand, economies of size and household survey methods.' *Oxford Bulletin of Economics and Statistics* 64(4): 341-360.

Gibson, J. and Kim, B. (2007). 'Measurement error in recall surveys and the relationship between household size and food demand.' *American Journal of Agricultural Economics* 89(2): 473-489.

Gibson, J. and Kim, B. (2010). 'Non-classical measurement error in long-term retrospective surveys.' *Oxford Bulletin of Economics and Statistic*s 72(5): 687-695.

Khandker, S. (2005). 'Microfinance and poverty: evidence using panel data from Bangladesh.' *World Bank Economic Review* 19(2): 263-286.

Pischke, J-S. (1995). 'Measurement error and earnings dynamics: Some estimates from the PSID Validation Study.' *Journal of Business and Economics Statistics* 13(3): 305-314.

Pradhan, M. (2001). 'Welfare analysis with a proxy consumption measure: Evidence from a repeated experiment in Indonesia.' *mimeo* Cornell University.