Bounding Expected Per Capita Household Consumption in the Presence of Demographic Change

by Timothy J. Halliday,

Department of Economics and John A. Burns School of Medicine,

University of Hawaii at Manoa

Working Paper No. 06-10

June 13, 2006

Abstract

This paper deals with the measurement of per capita household consumption expenditures when the household's underlying demographic structure changes during the survey period. To do this, we provide a formal definition of precisely what it means to mis-measure the household's demographic structure. We then use assumptions on demographic processes within the household during the survey period to construct bounds on expected per capita consumption expenditures. We estimate these bounds using two surveys from El Salvador, a country in which household demographic structures are very fluid. Our results reveal that these bounds can be wide suggesting that the measurement error in the household's demographic structure is non-trivial. We conclude by showing that mis-measured household size can have important implications when identifying economies of scale within the household.

JEL Classification: J12, C14
Key Words: Migration, Measurement Error, Semi-Parametric Bounds, Economies of Scale

Address: Timothy Halliday, Department of Economics, University of Hawaii at Manoa, 2424 Maile Way, Honolulu, HI 96822. Phone: (808) 956-8615. Fax: (808) 956-4347 <halliday@hawaii.edu>

1 Introduction

It has long been understood by economists that the household's demographic structure can have serious implications for how we define and measure individual living standards using household survey data. For example, in order to make welfare comparisons across households with different demographic compositions, it is essential to account for the fact that household members of different ages and genders will have different needs. Accordingly, it is reasonable to expect that a household with two adults and one child will require less compensation to achieve the same welfare level as a family with three adults. However, as has been argued by Pollak and Wales (1979), this task of identifying "equivalence scales" is impossible without imposing stringent assumptions on the household's preferences over goods and family members. Another way in which the household's demographic structure can influence welfare is through its impact on the price of intra-family public goods such as housing. The basic idea behind this is that public goods effectively become cheaper as the household becomes larger which, in effect, makes the household better off.[1] Unfortunately, however, despite the importance of this interplay between demographics and household consumer behavior, the literature on the topic has often raised more questions than it has resolved.

Further complicating matters is that most empirical investigations which attempt to identify equivalence scales or economies of scale require calculating consumption expenditures or income in *per capita* terms which can be quite difficult when the underlying structure of the household is dynamic. Indeed, most household surveys solicit retrospective information which ostensibly measures consumption expenditures over the duration of the survey period, but only solicit demographic information which measures the household's demographic structure at a point-in-time. This is potentially problematic because the household's demographic structure can be fluid since household compositions may change over the survey period due to fertility, marriage, divorce, migration and/or mortality. Consequently, the household's structure at the end of the survey period may not adequately reflect the household's structure during the survey period as a result of any one of these demographic processes. In this paper, we attempt to better understand what these demographic changes imply for the measurement of *per capita* household consumption expenditures.

To accomplish this, we make assumptions on the dynamics of the household's structure over

[1] Empircal evidence from a variety of studies suggests that the share of food in the household's budget decreases as the household becomes larger holding expenditures constant (see Lanjouw and Ravallion (1995) and Deaton and Paxson (1998) for two examples). If one believes Engel's assertation that the share of food in the household's budget is a proxy for the household's welfare then the empirical evidence suggests that larger households are better off than smaller households. However, recently, Deaton and Paxson (1998) have argued that Engel's assertion is unsound on the grounds that there are economies of scale within the household. Essentially, their argument is that, holding expenditures constant, larger households face lower prices for public goods which, in turn, makes them richer. Provided that the income elasticity of food is sufficiently high and provided that there are relatively small economies of scale in food consumption, we should expect to see the share of food in the household's budget increases as the household's size increases. Accordingly, in their example, an increase in welfare is accompanied by an increase in the food share which contradicts Engel's assertion.

the survey period and use these assumptions to derive bounds *a la* Manski (2003) on expectations of the household's size and *per capita* consumption over the survey period. We estimate these bounds using two surveys from El Salvador which provides us with a great venue to carry out this exercise due to the high volume of migration in the country. We find that these bounds are often wide, particularly, for households who report having members residing in the United States. This suggests that the extent of measurement error in the household's demographic composition is non-trivial.

The balance of this paper is organized as follows. In Section 2, we formally define what it means to mis-measure the household's demographic structure. In Section 3, we discuss the data that we employ. In Section 4, we derive the bounds. In Section 5, we discuss our results. In Section 7, we conclude by discussing the implications of measurement error in the household's demographic composition for the identification of economies of scale within the household.

2 The Problem

We assume that the household's decision process unfolds in continuous time. We let $C(s)$ denote the household's total consumption expenditures at time s and we let $N(s)$ denote the household's size at time s. For the bulk of this paper, we remain agnostic about the underlying decision process which determines $C(s)$ and $N(s)$. We assume that $N(s) \geq 1$ for all s. At any given point in time, (log) *per capita* consumption expenditures are given by

$$x(s) \equiv \log\left(\frac{C(s)}{N(s)}\right) \equiv c(s) - n(s). \tag{1}$$

Throughout this paper, we adopt the convention that upper-case variables correspond to levels and lower-case variables correspond to logs.

Unfortunately, the survey instrument does not collect data at every point in time and, thus, researchers do not observe the quantities $C(s)$ and $N(s)$ for *all* s. Instead, data is collected at discrete intervals such as every year or every two years which, in turn, means that pin-pointing $x(s)$ at a particular point-in-time can be quite difficult if not impossible. As a result, researchers are forced to summarize $x(s)$ over discrete time intervals such as $[t-1, t]$ or $[t, t+1]$.

To help fix ideas about how one would do this, we define the following objects:

$$C_t^* \equiv E[C(s)|s \in [t-1, t]] \tag{2}$$

and

$$N_t^* \equiv E[N(s)|s \in [t-1, t]]. \tag{3}$$

These quantities denote the average of the household's consumption expenditures and size over the interval $[t-1, t]$. In an ideal world, consumption surveys would enable precise measurement of these quantities and we would then measure *per capita* consumption expenditures *via* $x_t^* = c_t^* - n_t^*$.

However, in reality, C_t^* and N_t^* can be quite difficult to measure precisely. For example, while it is true that consumption surveys do solicit retrospective information which ostensibly should enable precise measurement of consumption expenditures over the survey period, this task is fraught with difficulties such recall bias and problems associated with survey design to name just a few. However, despite the importance of the measurement issues concerning C_t^*, we abstract from them in this paper. Instead, for the balance of the paper, we focus on the difficulties in measuring N_t^* and what it implies for the measurement of *per capita* consumption expenditures.

The fundamental problem with the measurement of N_t^* is that the household's structure often changes during the survey period. This may happen as a consequence of birth, death, migration, marriage or divorce. The fact that the household's demographic composition is fluid during the survey period is problematic when calculating consumption in *per capita* terms since it forces the researcher to re-evaluate precisely what it means to measure *anything* in *per capita* terms. Moreover, these problems are exacerbated by the fact that household surveys usually do not have adequate information about demographic transitions that occur during the survey period. Consequently, if the survey was administered at time t, researchers typically proxy for the household's size over the survey period with $N_t = N(t)$ which is the household's size at the exact point-in-time when the survey was administered and measure *per capita* consumption expenditures with $x_t = c_t^* - n_t$.

Often, n_t^* and x_t^* will deviate from n_t and x_t. In such a scenario, *per capita* consumption expenditures are measured with error which can be written as

$$e_t = x_t - x_t^* = n_t^* - n_t. \tag{4}$$

This expression for e_t summarizes the fundamental problem with n_t and x_t which is that the consumption component measures consumption over the whole time interval, whereas the demographic component only measures household size at an instant in time. Accordingly, if the household's demographic structure is constant over the time interval so that $N(s) = N$ for all

$s \in [t-1, t]$, then there will be no measurement error and, thus, we will have $x_t^* = x_t$. Otherwise, *per capita* consumption will be measured with error and we will have a distorted picture of the household's living standards over the time period. This could be particularly problematic for developing countries such as El Salvador where there is a tremendous amount of migration.

3 The Data

In this study, we utilize data from El Salvador. Our data come from two sources. The first is the *Encuesta de Hogares Propositos Multiples* (EHPM) which is a consumption survey that is administered annually by the Salvadoran Economic Ministry. We use the 2001 survey. There are a total of 11953 households in the survey. The second source is the BASIS panel which was administered by the Ohio State University. The advantage of the BASIS data is that their dynamic nature allows us to make inferences about how household demographic structures are changing across time. However, unlike the EHPM, they do not have comprehensive consumption data. We use data from the 2001 and 1999 waves of the panel. Since BASIS only surveyed households every other year, there is no wave from 2000. These data contain a total of 672 households.

3.1 *Encuesta de Hogares Propositos Multiples*

Table 1 summarizes the consumption expenditure data that we use from the EHPM. All values are in 2001 dollars. Our consumption data are divided into six main categories which are listed and defined in the table. The first three categories include all items that were bought, produced *via* home production and given to the household as aid.[2] Total consumption expenditures are defined to be the sum of these six categories. Average consumption expenditures in our data were roughly $3000.00 per household. To give the reader a better idea of the distribution of each component of consumption expenditures, we provide non-parametric density estimates of each expenditure component in Figure 1.

Column 1 of Table 2 summarizes the other variables that we use from the EHPM. We employ data on the number of migrants in the household, the household's size and the number of babies in the household. In this column, we also report the average amount remitted by all migrants in the household as well as total household income, although these data will not be used in the coming analysis. In Figure 2, we provide non-parametric density estimates of total consumption, family income and remittances.

The EHPM has a complex survey design. In the first stage, the country is divided into geographical strata. The Salvadoran Economic Ministry used a census which took place after the civil war concluded in 1992 to determine sample sizes within strata which ostensibly resulted in a nationally representative sample.[3] Accordingly, no weighting procedures should be required with these data. Within strata, primary sampling units or clusters were sampled. Because it is likely that observations within clusters will be correlated, it is necessary to adjust all standard errors throughout this analysis. We use the bootstrap to address these complex aspects of the survey's design (Efron and Tibshirani 1993). Additional detail concerning the bootstrapping

[2] We did not include rent expenditures in our calculation of total household expenditures. The reason for this was that the rent expenditure data in the EHPM appeared to be a bit suspicious. Discussions with a researcher at FUSADES, a Salvadoran think tank, confirmed our suspicion that there were problems with the rent data.

[3] Whether or not the weights which came from the 1992 census are still correct is an open question. Nevertheless, if these weights are incorrect, aside from running a new census, there is little that we can do to determine the correct weights.

procedure that we employ can be found in Section 4.

3.2 BASIS

Column 3 of Table 2 summarizes the BASIS data. From this sample, we also employ data on the number of migrants in the household, the household's size and the number of babies in the household. In addition, we employ a variable which we call *migration* which is the difference in the household's migrant stock across the 1999 and 2001 waves of the survey. Its mean is close to zero, but its standard deviation is very large, reflecting the large amount of migration in El Salvador. In this column, we also report the average amount remitted by all migrants in the household.

According to people at The Ohio State University, the survey has a stratified design with two strata: households with land and households without land. As with the EHPM, the sample sizes within strata were determined according to the 1992 census so as to (hopefully) ensure a representative sample. Consequently, no weighting scheme should be necessary. To the best of our knowledge, the survey contains no cluster design. However, we find it to be implausible that all of the observations in the sample are independent of one another, particularly, within small geographic units. Accordingly, as with the EHPM, we use the bootstrap to address any possible issues with the survey design. Once again, additional detail about this procedure can be found in Section 4.

3.3 Comparability of the Two Surveys

Comparing columns 1 and 3 of Table 2, one can see that there appears to be a lack of concordance between the EHPM and BASIS data. For example, there are, on average, 0.34 migrants per household in the EHMP, whereas there are 0.65 migrants per household in the BASIS data. What accounts for this difference?

One possible reason for this difference is that the migration modules in the two surveys differ and these differences have resulted in different migrant numbers.[4] However, when we look at the data on household size, which is measured in the same way in both surveys, we see, once again, that household sizes are substantially higher in the BASIS data than the EHPM. This suggests that differences in the migration modules are not responsible for the different migrant numbers in the two surveys.

Another possible reason for the discrepancies between the two surveys is that the BASIS data only sampled rural Salvadoran households, whereas the EHPM sampled all Salvadoran households. To shed light on this issue, in the second column of Table 2, we provide summary statistics from the EHPM data for only rural households.[5] We see that once we do this, the discrepancies between the two surveys persist. The average number of migrants among the rural households in the EHPM is 0.35 and the average household size is 4.92. Both of these

[4] In the EHPM, the number of migrants in a household was solicited simply by asking the household how many household members are residing abroad. The module does not ask where the migrant is resding, although a reasonable assumption is that the vast majority of these migrants are residing in the US. On the other hand, in the BASIS data, the migration module is substantially more complicated. First, the household is asked if there are any members who have left the household to work abroad including anybody who may have subsequently returned. From here, we had to use the remaining questions in the module to count the number of people who have returned from abroad and have not, subsequently, returned to the US. Using this battery of questions, we defined a migrant to be any household member who is *currently* residing in the US. While, ostensibly, the migrant numbers in the two surveys should be measuring the same quantities, the complex nature of the BASIS migration module and the simple nature of the EHPM migration model may be resulting in a higher estimated number of migrants in the BASIS data than the EHPM.

[5] The definitions of “rural” are the same in both surveys.

numbers are substantially lower than what we saw in the BASIS data. Moreover, even when we look within each of El Salvador's fourteen departments, most of which are rural, we still observe differences across the two surveys. In the interest of saving space, we do not report summary statistics by department, but these results are available upon request. Consequently, this suggests that the discrepancies between the two surveys are not the result of the BASIS data sampling only rural households.

We conclude that the lack of concordance between the two surveys cannot be explained by differences in survey design nor can it be explained by the BASIS data sampling rural households. Overall, the reasons for the discrepancies across the two surveys elude us. Unfortunately, there is not that much that we can do to rectify this. Accordingly, we proceed with the analysis using the two surveys as they stand while providing the caveat that the differences across the two surveys preclude us from making as precise of a statement about population parameters as we would like.

4 Bounding Expected *Per Capita* Consumption

We now turn ourselves to the task of using information from our two surveys to construct bounds on the household's expected *per capita* consumption expenditures. Because we do not consider the possibility that consumption is measured with error, this task only requires the construction of bounds on the household's size over the survey period. In order to address the ways in which N_t^* can be mis-measured, we introduce some notation. We let M_t denote the number of migrants in the household at the time t which is the time that the survey was administered. We define a migrant to be a household member residing outside of the household's dwelling. It is important to note that N_t only includes home dwellers and, thus, does not include any migrants. We let B_t denote the number of births that took place in the household during the survey year. Finally, we let D_t denote the number of deaths which took place during the survey year. Throughout this section, we do not address marriage or divorce.

Abstracting from marriage and divorce, we will have the following identity

$$N_t = N_{t-1} - \Delta M_t + B_t - D_t. \tag{5}$$

Since the household size is always positive, we will also have $N_t \geq 1$. It is also important to emphasize that the quantity ΔM_t is *net* migration. This simple identity suggests some sensible assumptions which will allow us to construct bounds on $N(s)$ for $s \in [t-1, t]$.

To see this how this can be done, suppose that the only demographic change that takes place in the household over the survey period is migration. Then, we will have that $N_{t-1} = N_t + \Delta M_t$. If $\Delta M_t > 0$, then this implies that

$$N_t < N_{t-1} = N_t + \Delta M_t. \tag{6}$$

Accordingly, this suggests that a reasonable assumption is that $N(s)$ was in the interval $[N_t, N_t + \Delta M_t]$ for all $s \in [t-1, t]$.

We use this logic to make three assumptions on the process for $N(s)$. They are:

$$N(s) \in [N_t - B_t, N_t + D_t] \text{ for } \Delta M_t = 0 \text{ and } s \in [t-1, t], \tag{A1}$$

$$N(s) \in [N_t - B_t, N_t + D_t + j] \text{ for } \Delta M_t = j > 0 \text{ and } s \in [t-1, t] \quad \text{(A2)}$$

and

$$N(s) \in [\max\{N_t - B_t + j, 1\}, N_t + D_t] \text{ for } \Delta M_t = j < 0 \text{ and } s \in [t-1, t]. \quad \text{(A3)}$$

The lower bound in Assumption 3 results from the assumption in Section 2 that the household size is always positive at any point in time.

It is important to emphasize that these conditions are assumptions and are not simply implied by the identity in equation (5). To better understand this, we consider a hypothetical scenario in which the household size was five at the end of the survey period and net migration out of the household was two during the survey period. For the sake of simplicity, we assume that no births or deaths took place during the survey period. In this scenario, Assumption A2 implies that $N(s)$ will lie in the interval $[5, 7]$ for all $s \in [t-1, t]$. However, in the absence of Assumption A2, this need not be the case. The reason for this is that ΔM_t is net migration during the time interval and this may mask some more subtle movements in the household's demographic structure which have occurred during the survey period. Going back to our example, it could have been that, shortly after the start of the survey period, just after time $t-1$, four members migrated out of the household. Now, suppose that just prior to the end of the survey, at time t, two of these same members subsequently returned to the household. In this hypothetical case, net migration out of the household would still be two. However, for the survey period, $N(s)$ would be in the interval $[3, 7]$ not $[5, 7]$. Assumptions A1 through A3 rule these types of scenarios out.

While we concede that these assumptions may be unrealistic in certain circumstances, they are still far weaker than the assumption that the household's demographic structure was constant over the survey period. This latter assumption is employed in the vast majority of studies on household consumption behavior. Thus, it is impossible to take exception to assumptions A1 through A3 without taking exception with the implicit assumptions in much of the previous literature.

These assumptions can easily be used to construct bounds on expectations of x_t^* and n_t^*: the true values of *per capita* consumption and household size. To accomplish this, we proceed in a series of steps. First, we note that these bounds on $N(s)$ imply the following bounds on N_t^*:

$$N_t^* \in [N_t - B_t, N_t + D_t] \text{ for } \Delta M_t = 0 \tag{7}$$

$$N_t^* \in [N_t - B_t, N_t + D_t + j] \text{ for } \Delta M_t = j > 0 \tag{8}$$

and

$$N_t^* \in [\max\{N_t - B_t + j, 1\}, N_t + D_t] \text{ for } \Delta M_t = j < 0. \tag{9}$$

Second, we note that because the logarithm function is monotonic, we will also have that

$$n_t^* \in [\log(N_t - B_t), \log(N_t + D_t)] \text{ for } \Delta M_t = 0, \tag{10}$$

$$n_t^* \in [\log(N_t - B_t), \log(N_t + D_t + j)] \text{ for } \Delta M_t = j > 0 \tag{11}$$

and

$$n_t^* \in [\log(\max\{N_t - B_t + j, 1\}), \log(N_t + D_t)] \text{ for } \Delta M_t = j < 0. \tag{12}$$

Third, we define the vector $W_t \equiv (N_t, M_t, D_t, B_t)$ and note that, by the Law of Iterated Expectations, we can write

$$E[n_t^*|W_t] = \sum_j E[n_t^*|\Delta M_t = j, W_t] P(\Delta M_t = j|W_t). \tag{13}$$

Fourth, equations (10), (11) and (12) imply that

$$\log(N_t - B_t) \leq E[n_t^*|\Delta M_t = j, W_t] \leq \log(N_t + D_t + j) \text{ for } j > 0, \tag{14}$$

$$\log(N_t - B_t) \leq E[n_t^*|\Delta M_t = j, W_t] \leq \log(N_t + D_t) \text{ for } j = 0$$

and

$$\log(\max\{N_t - B_t + j, 1\}) \leq E[n_t^*|\Delta M_t = j, W_t] \leq \log(N_t + D_t) \text{ for } j < 0. \tag{15}$$

Finally, these bounds together with equation (13) imply that

$$L(W_t) \leq E[n_t^*|W_t] \leq U(W_t) \tag{16}$$

where

$$U(W_t) \equiv \log(N_t + D_t) P(\Delta M_t \leq 0|W_t) + \sum_{j>0} \log(N_t + D_t + j) P(\Delta M_t = j|W_t) \tag{17}$$

and

$$L(W_t) \equiv \log(N_t - B_t) P(\Delta M_t \geq 0|W_t) + \sum_{j<0} \log(\max\{N_t - B_t + j, 1\}) P(\Delta M_t = j|W_t). \tag{18}$$

The bounds $L(W_t)$ and $U(W_t)$ can be calculated with the BASIS data since these data are a panel and, thus, contain information on ΔM_t. We can now bound expected *per capita* consumption in the following way:

$$l(W_t) \leq E[x_t^*|W_t] \leq u(W_t). \tag{19}$$

where $u(W_t) \equiv E[c_t^*|W_t] - L(W_t)$ and $l(W_t) \equiv E[c_t^*|W_t] - U(W_t)$.

5 Estimation and Inference

We estimate $u(W_t)$ and $l(W_t)$ in two steps. In the first, we estimate $E[c_t^*|W_t]$ and in the second we estimate $U(W_t)$ and $L(W_t)$. Estimation of $E[c_t^*|W_t]$ is relatively straight-forward and can be accomplished with the following regression:

$$c_{h,t}^* = \alpha + \sum_{j=2}^{10} d_{h,t}^{n,j}\eta_j + \sum_{j=1}^{3} d_{h,t}^{m,j}\mu_j + \sum_{j=1}^{2} d_{h,t}^{b,j}\beta_j + u_{h,t} \tag{20}$$

where $d_{h,t}^{n,j}$, $d_{h,t}^{m,j}$ and $d_{h,t}^{b,j}$ are dummy variables for having j household members, j migrants or j babies, respectively. The use of dummy variables for the household's demographic structure gives us a relatively loose parameterization of the regression function which provides us with a semi-parametric way of estimating the conditional expectation. Because the EHPM does not contain data on mortality, we not address the death of a family member when calculating these bounds.

Estimation of $U(W_t)$ and $L(W_t)$ is a slightly more complicated task. To estimate these objects, we devise two methods. The first is the most straight-forward. It involves using the

BASIS data to estimate the probabilities, $P(\Delta M_t = j|W_t)$, with ordered logit models.[6] We use the same right-hand side variables as equation (20). These fitted probabilities are then used to back out $U(W_t)$ and $L(W_t)$. One of the advantages of the ordered logit model is that it is easy to implement and the use of ancillary parameters for each migration category provides us with a flexible way of treating the regression function. One of the disadvantages of it, however, is that it assumes that the ancillary parameters are the same for households of all sizes. This is undesirable because it can produce positive probabilities of large positive (negative) values of ΔM_t for large (small) households. In practice, however, it turned out that these probabilities were typically small and were of little consequence when estimating the bounds.

Nevertheless, to address this issue, we also employ a simple alternative method where we split the sample into households with five or fewer members and households with more than five members and estimate the ordered logits separately for each sample. Doing this mitigates the problem of predicting large positive (negative) values of ΔM_t for larger (smaller) households since the procedure allows the ancillary parameters to vary with the household's size. After estimating the ordered logits on the split sample, we back out the migration probabilities and calculate the bounds just as before. While this method allows for a more flexible parameterization of the regression function, it has the disadvantage that it is less efficient than the previous method.[7]

We calculate the standard errors for $u(W_t)$ and $l(W_t)$ using the bootstrap. We do so for two reasons. The first is that the analytical standard errors for these bounds are rather complicated.

[6] The BASIS data measure net migration from 1999 to 2001, whereas the EHPM measures consumption over 2001. Ideally, we would have liked to have had a measure of migration from 2000 to 2001 in the BASIS data to have more consistency across the two data sets. Unfortuantely, there is little that can be done about this.

[7] We did not further sub-divide the sample into smaller sub-samples, however. The reason for this is that, doing so, involved estimating the ordered logits on rather small sub-samples of the data. These small samples sometimes resulted in non-convergence of the non-linear maximization routine when we bootstrapped our standard errors and, therefore, created substantial complications.

Calculation of the analytical standard errors of $u(W_t)$ and $l(W_t)$ would entail applying the delta method to the joint asymptotic distribution of the parameter estimates from the regression in equation (20) and the parameter estimates from the ordered logit estimates of $P(\Delta M_t = j|W_t)$. Given that there are a large number of ancillary parameters that need to be estimated to calculate the migration probabilities, this would have been somewhat of a cumbersome task. The second reason for bootstrapping the standard errors is that it allows us to address any issues concerning the complex design of these surveys. As pointed out by Deaton (1997), the bootstrap offers researchers with a convenient, albeit computationally intensive, means of addressing complex survey designs.

The bootstrapping procedure that we employ works as follows. First, from the EHPM and BASIS data, we re-sample from the data with replacement. To address the possibility of spatial correlation across households, we re-sample *municipios* from both the EHPM and the BASIS data. We re-sample as many *municipios* as were present in the data. For example, if the data contained 109 *municipios*, we would re-sample 109 *municipios* with replacement. In the case of the EHPM, the actual clusters or primary sampling units are contained within *municipios* and, thus, our standard errors are actually conservative. In the case of the BASIS data, it is unclear from the survey's documentation and our communication with the Ohio State University whether or not the survey had a cluster design. Nevertheless, to the extent that there is spatial correlation across households in these data, our calculation of the standard errors will address it provided that there is only correlation across observations within *municipios*. Using the re-sampled data, we then calculate $u(W_t)$ and $l(W_t)$. After this, we re-sample from the data again and repeat the process. After 100 replications, we calculate the standard errors of our estimated bounds.

6 Empirical Results

In this section, we estimate the bounds. In Table 3, we regress (log) consumption expenditures and migration on the right-rahnd side variables from equation (20) using OLS and ordered logit esimation, repectively. These regressions are used to calculate the bounds. The table gives the reader some notion of the relationship between consumption expenditures, migration and household demographic characteristics. A perusal of the table reveals few surprises. However, the migration dummies in column 6 are of some interest. Not surprisingly, we see that the migrant dummies are the single biggest predictors of migration. This suggests that we will have the most difficultly making precise inferences on *per capita* variables for households that have migrants residing abroad.

In Table 4, we report estimates of $U(W_t)$ and $L(W_t)$ for households with no babies using the first methodology from the previous section for calculating the bounds. We report the bound estimates for household sizes ranging from three to nine and for households who have between zero and three migrants. In Figure 3, we provide graphs of these bounds for households with zero, one, two and three migrants, respectively. Each panel of the figure reports $U(W_t)$, $L(W_t)$ and the log of the household's size as reported at the time of the survey. Looking at these results, two striking features emerge.

The first is that one of the bounds is always almost identical to the logarithm of the household's size as reported at the time of the survey. In the top left panel, the upper bound is very close to the log of the household's size, whereas, in the remaining three panels, we see the reverse.

However, after inspection of the formulae for the bounds in equations (17) and (18), this is not too surprising. The reason is that the probabilities, $P(\Delta M_t = j|W_t)$ for $j > 0$, will be very high for households with migrants and very low for households without migrants. Consequently, the suggestion is that the log of the household's size at the time of the survey's enumeration will tend to overestimate the household's size during the survey period for households without migrants and underestimate it for households with migrants.

The second interesting feature of these bounds is the relationship between their width and the household's demographic composition. Specifically, we see that the width of these bounds is increasing in the number of migrants in the household and decreasing in the number of members in the household. This suggests that measurement error in household size or any *per capita* variables will not be classical and, most likely, will be systematically correlated with the household's demographic composition.

Table 5 and Figure 4 are similar to Table 4 and Figure 3 in all respects except that these results use the second methodology from the previous section to calculate $U(W_t)$ and $L(W_t)$. The bound estimates are very similar in Tables 4 and 5. The only substantial difference between the two sets of results is that the second set is less efficient as can be seen in the higher standard errors in Table 5 when compared to those in Table 4. This similarity in the point-estimates of $U(W_t)$ and $L(W_t)$ in both tables suggests that the first method of estimation is preferred as it is more efficient and yields the same conclusions.

In Tables 6 and 7, we report the estimates of $u(W_t)$ and $l(W_t)$ which are the bounds on *per capita* consumption expenditures. We use the first method for calculating $U(W_t)$ and $L(W_t)$ in Table 6 and the second method in Table 7. Figures 5 and 6 plot the point-estimates of $u(W_t)$

and $l(W_t)$ as a function of household size. In these figures, we also plot expected consumption expenditures per household member as reported at the time of the survey (*i.e.* $E[c_t^*|W_t] - n_t$). These results on the *per capita* consumption expenditure bounds are very much in the same spirit as the results on the household size bounds.

7 Implications

We conclude this paper by exploring the implications of mis-measured household size for the estimation of Engel curves and, more specifically, for the identification of economies of scale within the household. To do this, we consider an Engel curve of the form:

$$\omega_f = \alpha + \beta x^* + \gamma n^* + \varepsilon \tag{21}$$

where ω_f is the share of food in the household's budget and x^* and n^* are defined as in Section 2. We assume that the residual in this equation is uncorrelated with all of the right-hand side regressors. Throughout this section, we suppress all subscripts. This specification was first estimated by Working (1943) and has been used extensively in the literature on household consumer behavior.[8] As pointed by Deaton (1997) and Deaton and Muellbauer (1980), this Engel curve has the advantage that it fits the data well and is consistent with optimizing household behavior.

Arguments put forth in a seminal piece by Deaton and Paxson (1998) suggest that γ should be positive. The foundation of their argument for this is that public goods within the household

[8] See Deaton and Paxson (1998) or Deaton and Muellbauer (1986) for two examples.

become cheaper as the household's size increases and, if we hold the household's living standards constant, this effectively makes the household richer. To better understand this consider a situation, discussed in Deaton and Paxson's original paper, in which two people decide to move in together. Once these people are living under one roof, they no longer need to pay two separate rents. Provided that their incomes remain constant, each individual has in effect become richer.

Deaton and Paxson go on to argue that provided that the income elasticity of food is sufficiently high, which it is throughout the developing world, the household's consumption of food should increase and, thus, we should expect to see that γ is positive. However, using data from a variety of countries which run the whole gamut of living standards, they show that the share of food in the household's budget actual *decreases* with the household's size. This is the exact opposite of what the theory predicts. They then spend a considerable amount of time trying to rationalize their empirical findings, but, ultimately, they are unable to do so.[9] Consequently, we are left with a puzzle.

To better understand the role that mis-measured household size can play in the identification of economies of scale, we first note that, because the household's size is measured with error, equation (21) cannot be estimated since x^* and n^* are never observed. Instead, the econometrician has to estimate

$$\omega_f = \alpha + \beta x + \gamma n + \upsilon \tag{22}$$

[9] In a comment on Deaton and Paxson (1998), Gan and Vernon (2003) claim to resolve the puzzle. The crux of their argument is that there may be relatively large economies of scale in food consumption and, consequently, it may be reasonable to see that the share of food expenditures in the household's budget decreases with household size. The main reason underlying this assertion is that total household expenditures may include goods that are potentially more private than food such as clothes. Gan and Vernon provide evidence that as the household's size rises, food expenditures as a share of food and housing expenditures also rise. They claim that this resolves the puzzle since housing is known to be more public than food. However, Deaton and Paxson (2003), in a response to the comment, assert that Gan and Vernon's findings are consistent with empirical results in their original piece, but do nothing to resolve the puzzle. Their fundamental contention with Gan and Vernon's comment is that it provides little evidence that there are substantial economies of scale in food consumption.

where $x = x^* + e$, $n = n^* - e$ and $\upsilon = \varepsilon + (\gamma - \beta)e$. Clearly, OLS will not yield consistent estimates of β and γ since υ is correlated with both x and n. Next, we project e onto x and n and obtain

$$e = \kappa + \phi x + \lambda n + u \tag{23}$$

where u is uncorrelated with both x and n. Because $x = x^* + e$ and $n = n^* - e$, it is reasonable to expect that $\phi > 0$ and $\lambda < 0$. Next, we substitute equation (23) into equation (22) and we obtain

$$\omega_f = \widetilde{\alpha} + \widetilde{\beta} x + \widetilde{\gamma} n + \varepsilon + u \tag{24}$$

where $\widetilde{\alpha} \equiv \alpha + (\gamma - \beta)\kappa$, $\widetilde{\beta} \equiv \beta + (\gamma - \beta)\phi$ and $\widetilde{\gamma} \equiv \gamma + (\gamma - \beta)\lambda$.

The probability limit of the OLS estimate of the economies of scale parameter is $\widetilde{\gamma}$. Accordingly, we can write

$$p \lim \widehat{\widetilde{\gamma}} = (1 + \lambda)\gamma - \lambda\beta. \tag{25}$$

This equation illustrates how mis-measured household size can lead to a failure to identify economies of scale even when they are present. To better see this, first note that to the extent that λ is negative, the first term on the right-hand side of the equation will be less than γ and, perhaps, even negative. Second, Engel's Law says that the share of food in the household's budget will fall with the household's living standards and, thus, β will be negative. Indeed, in practically every study of household consumption behavior which involves Working's Engel curve, estimates of β are always negative and very large. Accordingly, to the extent that λ is negative, the second term in the probability limit will be negative and potentially large, depending on the magnitude of λ. What this all means then is that tests for the presence of economies of scale of this type may have low power due to the presence of measurement error in the household's size. Moreover, this calculation also suggests that negative estimates of γ may occur even when economies of scale are present. Finally, it is interesting to point out that Deaton and Paxson find that their puzzle is deepest (i.e. the estimates of γ are the most negative) for the poorest countries which also happen to be the countries where household demographic structures are the most pliable.

We conclude this paper with some *prima facie* evidence which suggests that OLS estimates of γ are positively related to β as is suggested by equation (25). To do this, we estimate

$$\omega_f^j = \alpha^j + \beta^j x + \gamma^j n + \sum_{k=1}^{K-1} \eta_k^j \frac{N_k}{N} + \upsilon^j \text{ for } j = 1, ..., J. \quad (26)$$

The dependent variable in this equation is the budget share of a particular food item. The food items that we use are tortillas, bread, rice, milk, beans, chicken, beef, pork, vegetables and eggs. $\frac{N_k}{N}$ is the share of the total number of household members in a particular age and gender category. We report the estimates of γ^j and β^j in Table 8. What can be seen in the table is that the estimates of γ^j have a lot to do with the estimates of β^j as is suggested by equation (25). Generally, we see that food items with higher income elasticities also have higher estimates of γ^j. To better see this, we plot the pairs $\left(\widehat{\gamma^j}, \widehat{\beta^j}\right)$ in Figure 7 which clearly illustrates a strong positive relationship between the two parameter estimates.[10]

[10] There are two alternative explanations for the positive relationship in Figure 7. The first is that goods that have higher income elasticities also have fewer economies of scale associated with them than the other goods in the household's budget. If this were, in fact, the case, then we would see that, as the household's size increases, the prices of the other goods in the budget would decrease more rapidly than the goods with the higher income elasticities. However, if this were true, then these results suggest that there are fewer economies of scale in beef consumption than in pork consumption. It is unclear to us why this would be the case. The second explanation for the relationship in Figure 7 has to do with the theory in Deaton and Paxson's original work. Specifically,

The results and calculations of this section suggest that mis-measured household size *may* help to explain the paradox that Deaton and Paxson originally posed. However, we are cautious to say anything more than this. The fundamental reason for our caution is that we still do not fully understand the nature of this measurement error. Indeed, the results of the previous section suggest that the measurement error in household size is potentially complicated. Consequently, at this point, we do not understand the magnitude (or even the sign) of the parameter λ that well. While we believe that it is reasonable to suspect that λ is negative, further work is still warranted.

References

[1] Deaton, Angus. 1997. *The Analysis of Household Surveys.* Baltimore: Johns Hopkins University Press.

[2] Deaton, Angus and John Muellbauer. 1980. *Economics and Consumer Behavior.* Cambridge: Cambridge University Press.

[3] Deaton, Angus and John Muellbauer. 1986. "On Measuring Child Costs: with Application to Poor Countries." *Journal of Political Economy.* 94: 720-744.

they show that the consumption of a good should increase with the household's size when the income elasticity of that good is high relative to the absolute value of its price elasticity. The fact that we find positive estimates of the economies of scale parameter for goods that are luxuries (or almost luxuries) like beef or pork suggests that there may be some credence to this. However, working against this explanation is the presumption that the price elasticity of beef or pork is higher than the price elasticity of staples like tortillas. Unfortunately, without data on unit prices, there is no way of verifying this presumption. In addition, this argument suggests that the negative estimates of the economies of scale parameter for tortillas is the result of the absolute value of the price elasticity of a staple being high relative to its income elasticity which we find to be somewhat hard to believe.

[4] Deaton, Angus and Christina Paxson. 1998. "Economies of Scale, Household Size and the Demand for Food." *Journal of Political Economy.* 106: 897-930.

[5] Deaton, Angus and Christina Paxson. 2003. "Engel's What? A Response to Gan and Vernon." *Journal of Political Economy.* 111: 1378-1381.

[6] Gan, Li and Victoria Vernon. 2003. "Testing the Barten Model of Economies of Scale in Household Consumption: Toward Resolving a Paradox of Deaton and Paxson." *Journal of Political Economy.* 111: 1361-1377.

[7] Efron, Bradley and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap.* London: Chapman and Hall.

[8] Lanjouw, Peter and Martin Ravallion. 1995. "Poverty and Household Size." *The Economic Journal.* 105: 1415-1434.

[9] Manski, Charles. 2003. *Partial Identification of Probability Distributions.* New York: Springer-Verlag.

[10] Pollak, Robert A. and Terence J. Wales. 1979. "Welfare Comparisons and Equivalence Scales." *American Economic Review, Papers and Proceedings.* 69: 216-221.

[11] Working, Holbrook. 1943. "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association.* 104: 275-298.

Table 1: Consumption Items from the EHPM

	Mean (Standard Deviation)	Definition
Food Consumption		
- Bought	1221.46	
	(883.08)	Includes tortillas, bread, rice, beans, salt, sugar, cereal, grains, chicken,
- Auto – Production	139.20	beef, pork, fish, eggs, milk, cheese, aceite, vegetables, fruit, restaurant
	(380.39)	meals, prepared meals, coffee, drinks, alcohol, other items
- Aid	82.43	
	(300.92)	
Consumption 1		
- Bought	239.20	
	(264.73)	Includes toiletries, soap, cleaning products, magazines, newspapers,
- Auto – Production	8.33	cosmetics, fuel, transportation, babysitting
	(53.60)	
- Aid	3.69	
	(35.03)	
Consumption 2		
- Bought	149.47	
	(339.84)	Includes travel, jewelry, pots, towels, car repairs, other repairs,
- Auto – Production	0.42	appliances, furniture, clothes, glasses
	(14.35)	
- Aid	22.23	
	(108.20)	
Utilities	357.30	Includes water, electricity, kerosene, propane, candles, carbon, leña,
	(422.39)	telephone, cell phone, cable, garbage
School Expenditures	678.26	Includes tuition, supplies, uniforms, textbooks
	(988.05)	

Medical Expenditures	71.11 (320.30)	Includes doctors visits, lab work, x-rays, hospital days, medicine
Total	2929.19 (2195.43)	

Table 2: Other Variables in the EHPM and the BASIS Data

`	Mean (Standard Deviation)			Definition
	EHMP - All	EHMP - Rural	BASIS	
Migrants	0.34 (0.91)	0.35 (0.96)	0.65 (1.33)	Total number of household members residing the US
Migration	-	-	0.03 (1.30)	Change in the household's migrant stock between 1999 and 2001
Remittances	362.44 (1128.47)	334.40 (929.25)	557.99 (1452.74)	Amount sent back in cash or kind to the HH by all migrants during the year in 2001 Dollars
Family Income	3424.14 (4633.13)	2018.51 (2207.41)	-	Family Income in 2001 Dollars
Household Size	4.43 (2.26)	4.92 (2.48)	5.96 (2.71)	Size of the household in El Salvador
Babies	0.09 (0.29)	0.10 (0.31)	0.13 (0.36)	Number of household members less one year old
Sample Size	11953	4534	672	

Table 3: Consumption, Migration and Demographics

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS Regressions of Log Consumption on HH Demographics			Ordered Logistic Regression of Migration on HH Demographics		
Household Size						
= 2	0.41	0.42	0.41	1.53	1.53	0.57
	(14.08)	(14.12)	(14.44)	(2.15)	(2.15)	(0.81)
= 3	0.73	0.75	0.75	0.79	0.79	0.47
	(23.84)	(24.65)	(24.34)	(1.42)	(1.42)	(0.80)
= 4	0.95	0.97	0.97	1.28	1.28	1.37
	(34.34)	(35.15)	(37.14)	(2.23)	(2.22)	(2.26)
= 5	1.03	1.06	1.07	1.69	1.69	1.69
	(34.13)	(35.52)	(36.53)	(2.88)	(2.88)	(2.81)
= 6	1.07	1.10	1.10	1.51	1.52	1.20
	(33.84)	(35.00)	(35.98)	(2.58)	(2.58)	(1.97)
= 7	1.05	1.09	1.09	0.96	0.98	1.04
	(29.25)	(30.75)	(31.96)	(1.69)	(1.71)	(1.72)
= 8	1.11	1.16	1.16	1.22	1.23	0.93
	(29.51)	(31.21)	(32.91)	(2.04)	(2.04)	(1.35)
= 9	1.07	1.13	1.13	1.15	1.15	0.99
	(27.05)	(30.14)	(31.57)	(1.93)	(1.93)	(1.62)
>= 10	1.21	1.30	1.29	1.05	1.11	1.21
	(28.36)	(31.52)	(32.70)	(1.92)	(1.93)	(2.03)
Babies						
= 1		-0.26	-0.25		-0.05	0.03
		(-11.84)	(-11.27)		(-0.18)	(0.13)
>= 2		-0.26	-0.27		-0.75	-0.94
		(-1.38)	(-1.51)		(-1.15)	(-1.35)
Migrants						
= 1			0.21			2.63
			(7.61)			(5.63)
= 2			0.27			3.79
			(8.24)			(7.82)
>= 3			0.31			4.90
			(8.09)			(9.90)
R Squared	0.190	0.199	0.217	0.010	0.011	0.179

N	11696	11696	11696	672	672	672

All standard errors cluster on municipios. t-statistics are reported in parentheses.

Table 4: Bounds on Household Size – No Babies (Method 1)

HH Size (Log of HH Size)	No Migrants Upper Bound	No Migrants Lower Bound	One Migrant Upper Bound	One Migrant Lower Bound	Two Migrants Upper Bound	Two Migrants Lower Bound	Three or More Migrants Upper Bound	Three or More Migrants Lower Bound
3	1.106	0.885	1.183	1.079	1.287	1.092	1.423	1.097
(1.099)	(0.003)	(0.035)	(0.022)	(0.012)	(0.043)	(0.003)	(0.047)	(0.0010)
4	1.400	1.300	1.513	1.379	1.622	1.384	1.749	1.386
(1.386)	(0.005)	(0.025)	(0.031)	(0.005)	(0.047)	(0.001)	(0.060)	(0.0004)
5	1.625	1.555	1.737	1.605	1.835	1.608	1.949	1.609
(1.609)	(0.005)	(0.013)	(0.021)	(0.003)	(0.034)	(0.001)	(0.035)	(0.0002)
6	1.800	1.719	1.871	1.786	1.946	1.790	2.037	1.791
(1.792)	(0.003)	(0.017)	(0.017)	(0.004)	(0.029)	(0.001)	(0.028)	(0.0003)
7	1.952	1.874	2.008	1.940	2.071	1.944	2.149	1.945
(1.946)	(0.002)	(0.018)	(0.014)	(0.005)	(0.022)	(0.001)	(0.030)	(0.0004)
8	2.084	2.012	2.130	2.074	2.184	2.078	2.253	2.079
(2.079)	(0.002)	(0.018)	(0.014)	(0.004)	(0.021)	(0.001)	(0.033)	(0.0003)
9	2.202	2.141	2.245	2.193	2.294	2.196	2.357	2.197
(2.197)	(0.002)	(0.014)	(0.012)	(0.003)	(0.020)	(0.001)	(0.023)	(0.0003)

This table contains the upper and lower bounds on expected (log) household size conditional on household demographic characteristics. Bootstrapped standard errors are reported in parentheses.

Table 5: Bounds on Household Size – No Babies (Method 2)

HH Size (Log of HH Size)	No Migrants Upper Bound	No Migrants Lower Bound	One Migrant Upper Bound	One Migrant Lower Bound	Two Migrants Upper Bound	Two Migrants Lower Bound	Three or More Migrants Upper Bound	Three or More Migrants Lower Bound
3	1.106	0.891	1.162	1.074	1.328	1.095	1.436	1.097
(1.099)	(0.003)	(0.038)	(0.024)	(0.022)	(0.052)	(0.003)	(0.070)	(0.0015)
4	1.402	1.316	1.500	1.379	1.677	1.385	1.777	1.386
(1.386)	(0.006)	(0.020)	(0.031)	(0.008)	(0.054)	(0.001)	(0.076)	(0.0005)
5	1.626	1.561	1.719	1.605	1.874	1.609	1.962	1.609
(1.609)	(0.006)	(0.014)	(0.025)	(0.006)	(0.036)	(0.001)	(0.052)	(0.0004)
6	1.800	1.717	1.880	1.787	1.918	1.789	2.031	1.791
(1.792)	(0.003)	(0.021)	(0.022)	(0.006)	(0.027)	(0.002)	(0.042)	(0.0008)
7	1.952	1.870	2.013	1.941	2.044	1.943	2.140	1.945
(1.946)	(0.002)	(0.023)	(0.018)	(0.006)	(0.025)	(0.002)	(0.043)	(0.0009)
8	2.084	2.011	2.136	2.075	2.164	2.077	2.250	2.079
(2.079)	(0.002)	(0.029)	(0.020)	(0.007)	(0.027)	(0.002)	(0.050)	(0.0011)
9	2.202	2.140	2.250	2.193	2.275	2.195	2.352	2.197
(2.197)	(0.002)	(0.022)	(0.017)	(0.005)	(0.022)	(0.002)	(0.033)	(0.0007)

This table contains the upper and lower bounds on expected (log) household size conditional on household demographic characteristics. Bootstrapped standard errors are reported in parentheses.

Table 6: Bounds on *Per Capita* Household Consumption – No Babies (Method 1)

HH Size (Log of HH Size)	No Migrants Upper Bound	No Migrants Lower Bound	One Migrant Upper Bound	One Migrant Lower Bound	Two Migrants Upper Bound	Two Migrants Lower Bound	Three or More Migrants Upper Bound	Three or More Migrants Lower Bound
3	6.716	6.494	6.732	6.628	6.781	6.587	6.819	6.492
(1.099)	(0.057)	(0.046)	(0.036)	(0.040)	(0.043)	(0.061)	(0.046)	(0.066)
4	6.528	6.427	6.658	6.525	6.717	6.478	6.757	6.393
(1.386)	(0.047)	(0.040)	(0.032)	(0.044)	(0.039)	(0.061)	(0.040)	(0.072)
5	6.369	6.298	6.529	6.397	6.589	6.362	6.629	6.289
(1.609)	(0.042)	(0.040)	(0.035)	(0.041)	(0.043)	(0.055)	(0.043)	(0.055)
6	6.237	6.156	6.381	6.295	6.440	6.283	6.480	6.233
(1.792)	(0.041)	(0.037)	(0.033)	(0.037)	(0.039)	(0.049)	(0.042)	(0.050)
7	6.067	5.990	6.212	6.144	6.271	6.144	6.311	6.107
(1.946)	(0.037)	(0.032)	(0.031)	(0.034)	(0.037)	(0.043)	(0.039)	(0.049)
8	5.996	5.924	6.145	6.089	6.204	6.098	6.245	6.071
(2.079)	(0.043)	(0.039)	(0.044)	(0.046)	(0.047)	(0.051)	(0.049)	(0.059)
9	5.842	5.780	6.000	5.948	6.060	5.962	6.100	5.940
(2.197)	(0.036)	(0.033)	(0.035)	(0.039)	(0.041)	(0.046)	(0.047)	(0.052)

This table contains the upper and lower bounds on expected (log) *per capita* consumption conditional on household demographic characteristics. Bootstrapped standard errors are reported in parentheses.

Table 7: Bounds on *Per Capita* Household Consumption – No Babies (Method 2)

HH Size (Log of HH Size)	No Migrants Upper Bound	No Migrants Lower Bound	One Migrant Upper Bound	One Migrant Lower Bound	Two Migrants Upper Bound	Two Migrants Lower Bound	Three or More Migrants Upper Bound	Three or More Migrants Lower Bound
3	6.710	6.495	6.737	6.649	6.779	6.546	6.818	6.479
(1.099)	(0.060)	(0.047)	(0.045)	(0.046)	(0.044)	(0.068)	(0.045)	(0.083)
4	6.512	6.425	6.659	6.538	6.716	6.423	6.756	6.336
(1.386)	(0.043)	(0.039)	(0.037)	(0.048)	(0.039)	(0.067)	(0.038)	(0.085)
5	6.363	6.298	6.530	6.415	6.588	6.323	6.629	6.276
(1.609)	(0.038)	(0.036)	(0.036)	(0.043)	(0.040)	(0.054)	(0.041)	(0.066)
6	6.239	6.156	6.380	6.287	6.440	6.311	6.480	6.240
(1.792)	(0.041)	(0.036)	(0.036)	(0.041)	(0.040)	(0.048)	(0.040)	(0.058)
7	6.071	5.990	6.211	6.139	6.272	6.171	6.311	6.117
(1.946)	(0.042)	(0.035)	(0.040)	(0.043)	(0.041)	(0.048)	(0.044)	(0.061)
8	5.997	5.924	6.144	6.083	6.205	6.119	6.245	6.076
(2.079)	(0.047)	(0.037)	(0.043)	(0.047)	(0.048)	(0.055)	(0.046)	(0.067)
9	5.843	5.781	6.000	5.943	6.061	5.981	6.101	5.945
(2.197)	(0.039)	(0.033)	(0.037)	(0.041)	(0.042)	(0.047)	(0.047)	(0.058)

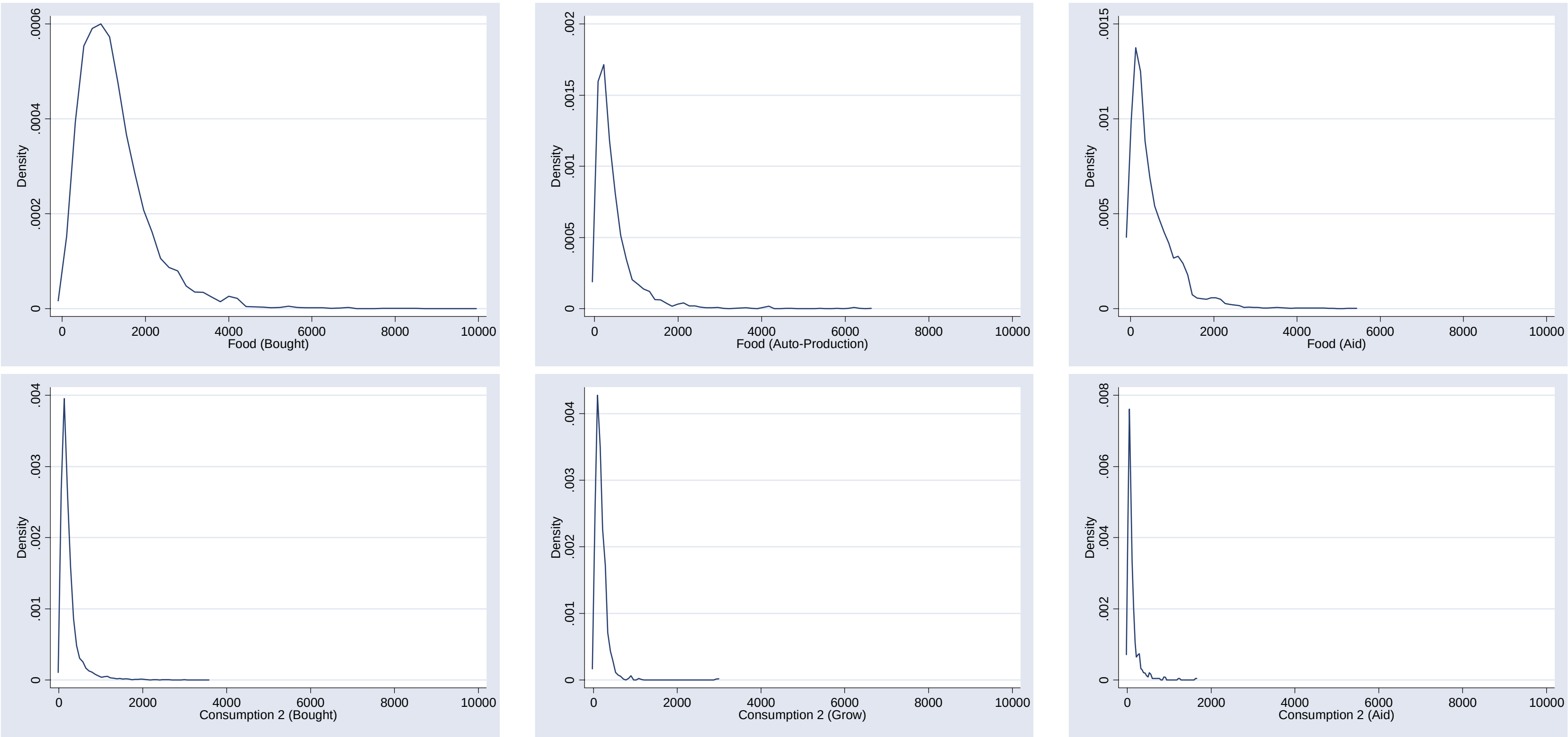
This table contains the upper and lower bounds on expected (log) *per capita* consumption conditional on household demographic characteristics. Bootstrapped standard errors are reported in parentheses.

Table 8: Engel Curve Estimates

	Tortillas	Bread	Rice	Milk	Beans
Log of Per Capita	-0.733	-0.003	-0.014	-0.001	-0.030
Expenditures	(-31.61)	(-2.35)	(-28.38)	(-0.67)	(-27.49)
Log of Household	-0.015	-0.001	-0.003	-0.000	-0.005
Size	(-7.73)	(-0.65)	(-7.65)	(-0.17)	(-7.99)
	Chicken	Beef	Pork	Vegetables	Eggs
Log of Per Capita	-0.003	0.004	0.000	-0.006	-0.023
Expenditures	(-2.80)	(4.39)	(0.95)	(-9.66)	(-33.27)
Log of Household	0.002	0.004	0.001	-0.003	-0.006
Size	(2.27)	(4.94)	(3.63)	(-5.15)	(-7.92)

This table contains OLS estimates of ten separate Engel curves. The dependent variable in each regression is the share of total household expenditures that were allocated to each of the ten food items which are listed above. Each regression contains the log of per capita consumption expenditures, the log of the household size and demographic controls. All regressions adjust the standard errors for clustering on municipios.

Figure 1: Density of the Different Components of Consumption Expenditures in the EHPM



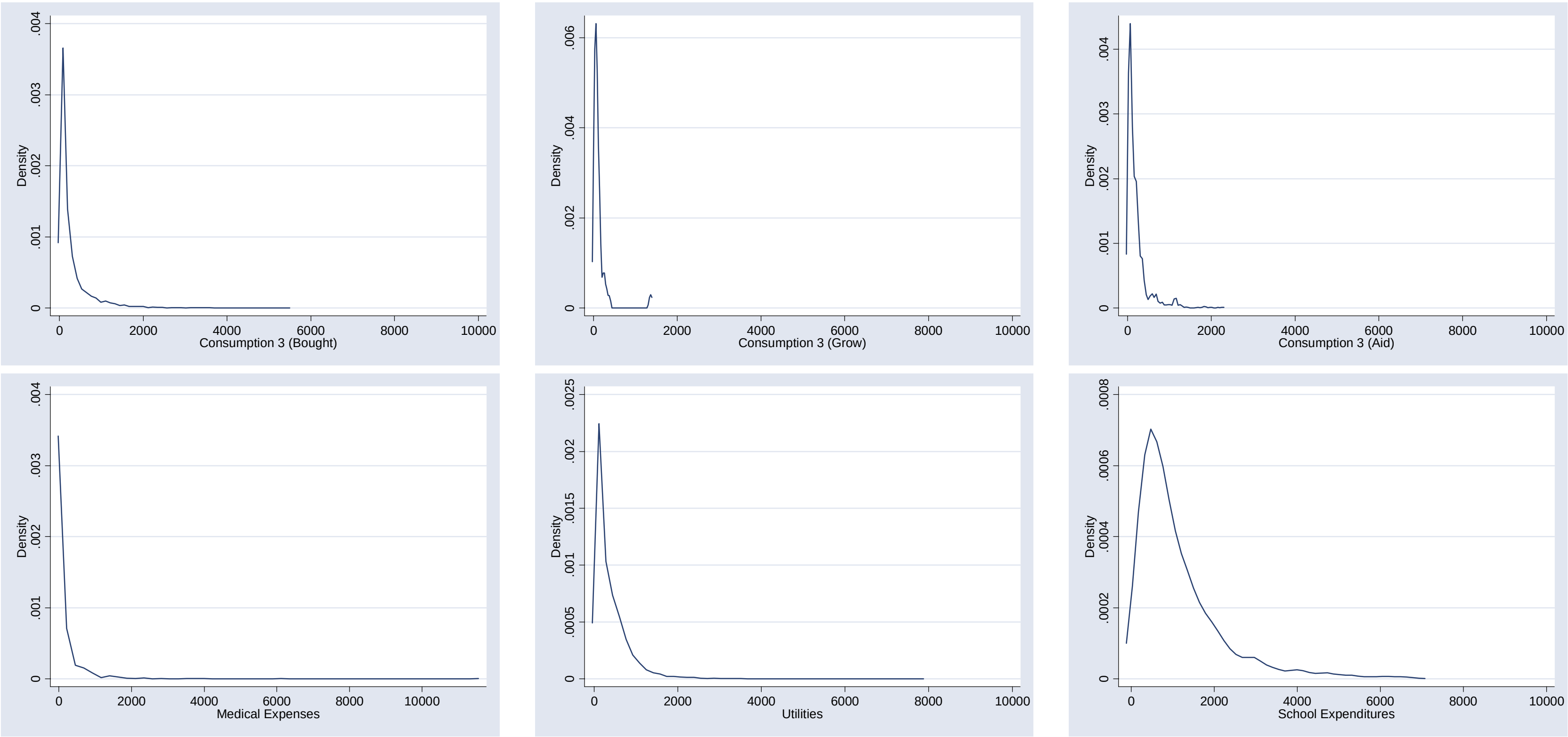

Density
0
.001
.002
.003
.004
0
2000
4000
6000
8000
10000
Consumption 3 (Bought)
Density
0
.002
.004
.006
0
2000
4000
6000
8000
10000
Consumption 3 (Grow)
Density
0
.001
.002
.003
.004
0
2000
4000
6000
8000
10000
Consumption 3 (Aid)
Density
0
.001
.002
.003
.004
0
2000
4000
6000
8000
10000
Medical Expenses
Density
0
.0005
.001
.0015
.002
.0025
0
2000
4000
6000
8000
10000
Utilities
Density
0
.0002
.0004
.0006
.0008
0
2000
4000
6000
8000
10000
School Expenditures

Figure 2: Density of Total Consumption Expenditures, Income and Remittances in the EHPM

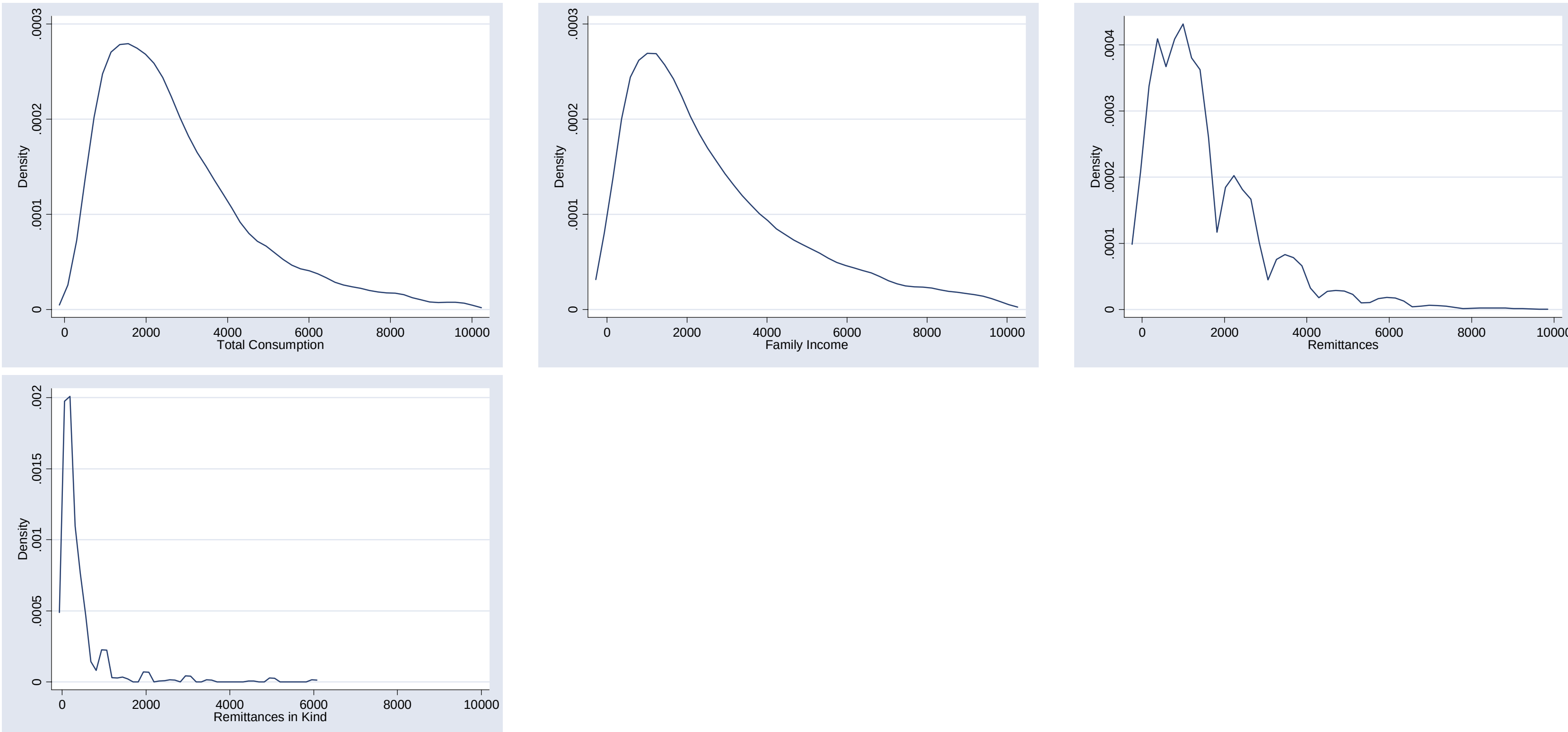


Figure 3: Bounds on Household Size (Method 1)

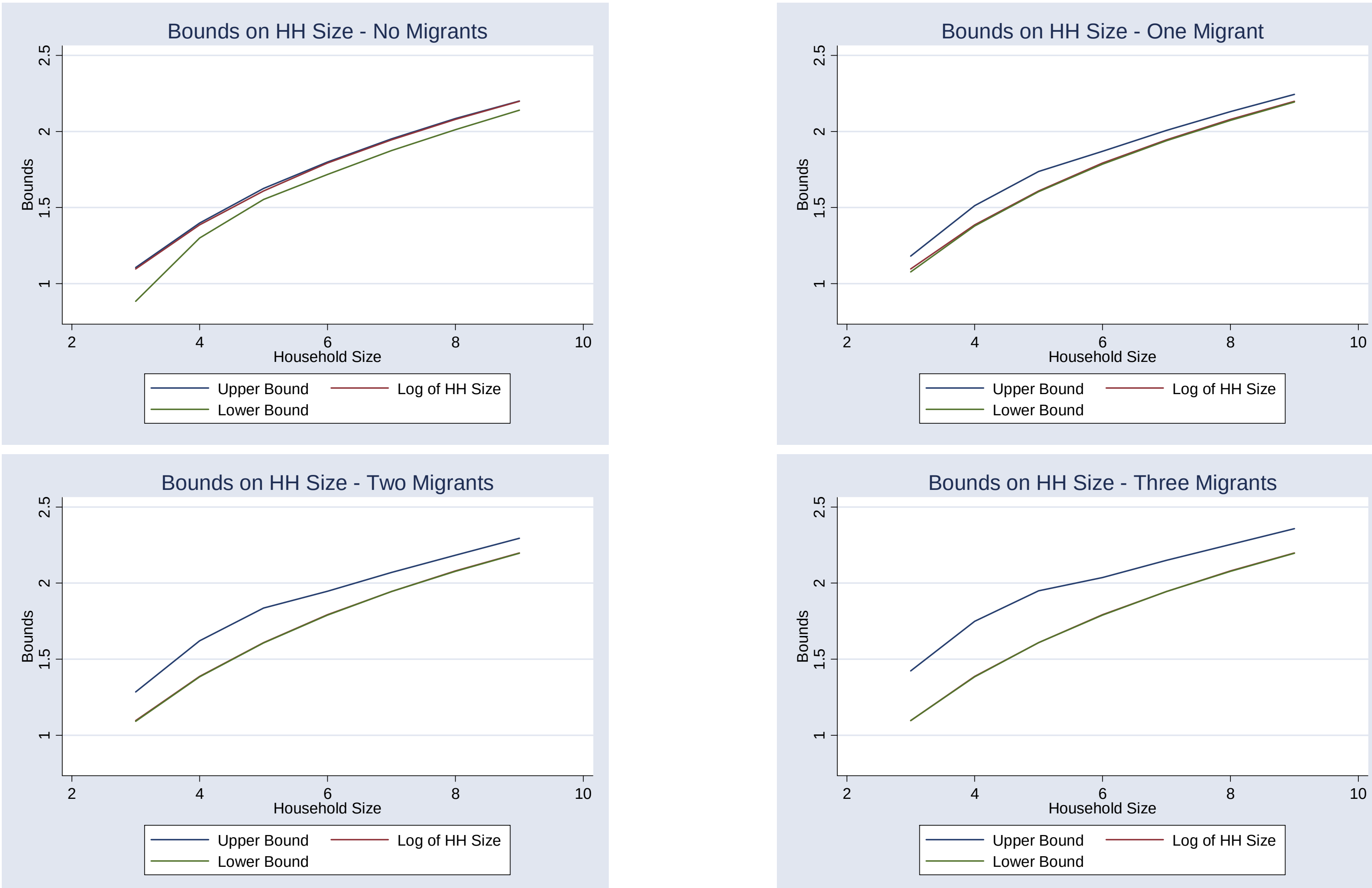


Figure 4: Bounds on Household Size (Method 2)

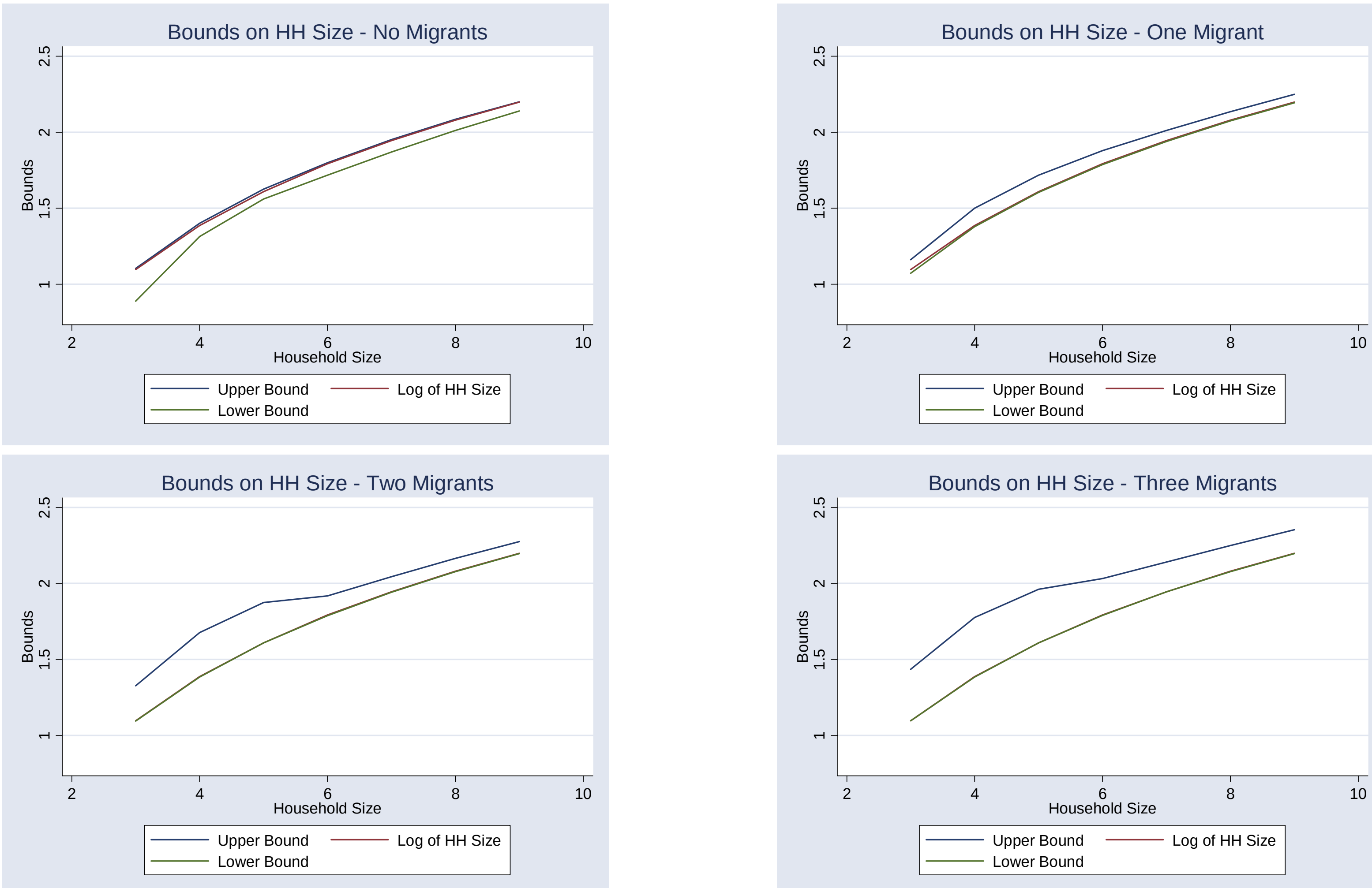


Figure 5: Bounds on Expected *Per Capita* Consumption (Method 1)

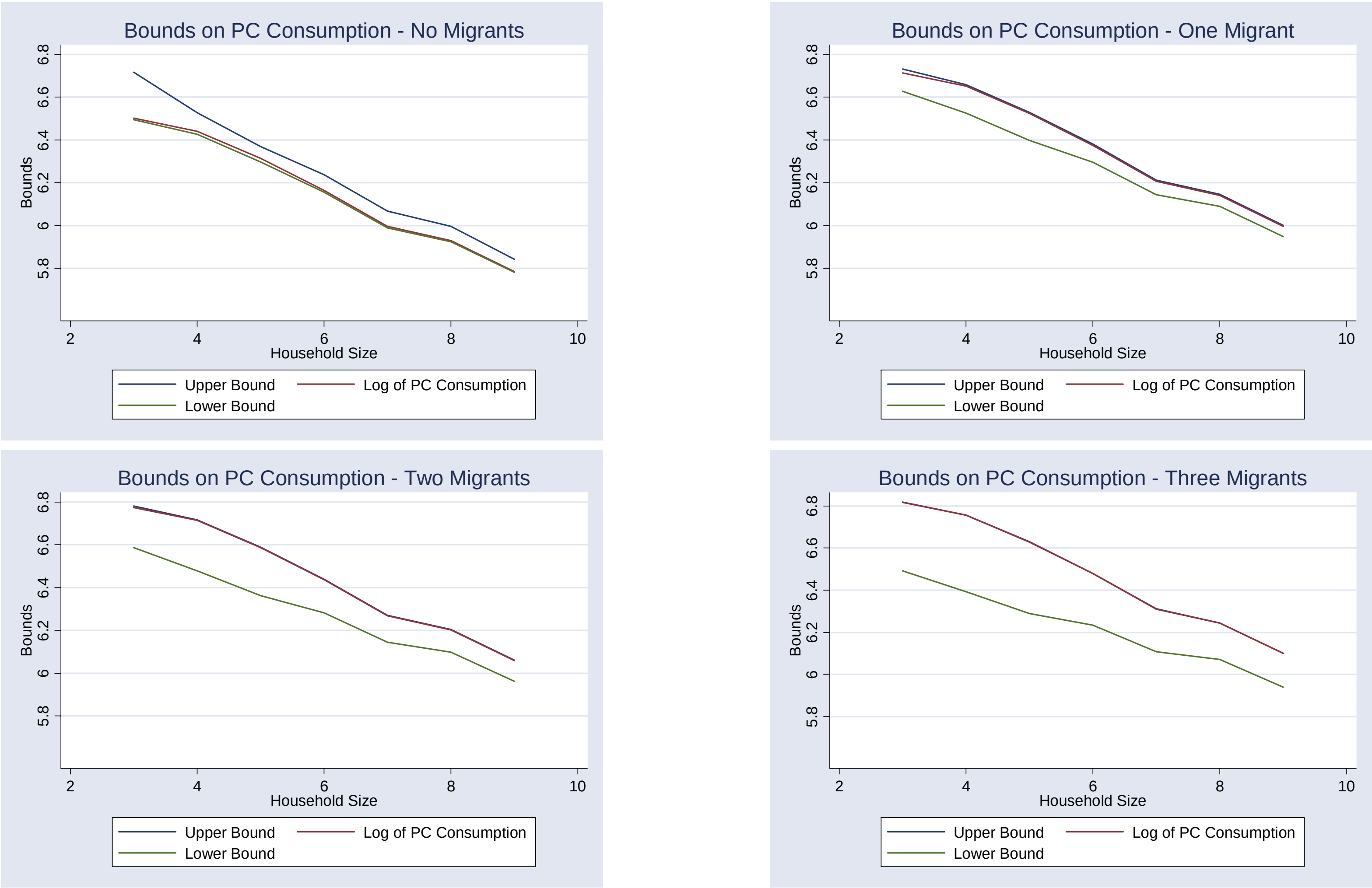


Figure 6: Bounds on Expected *Per Capita* Consumption (Method 2)

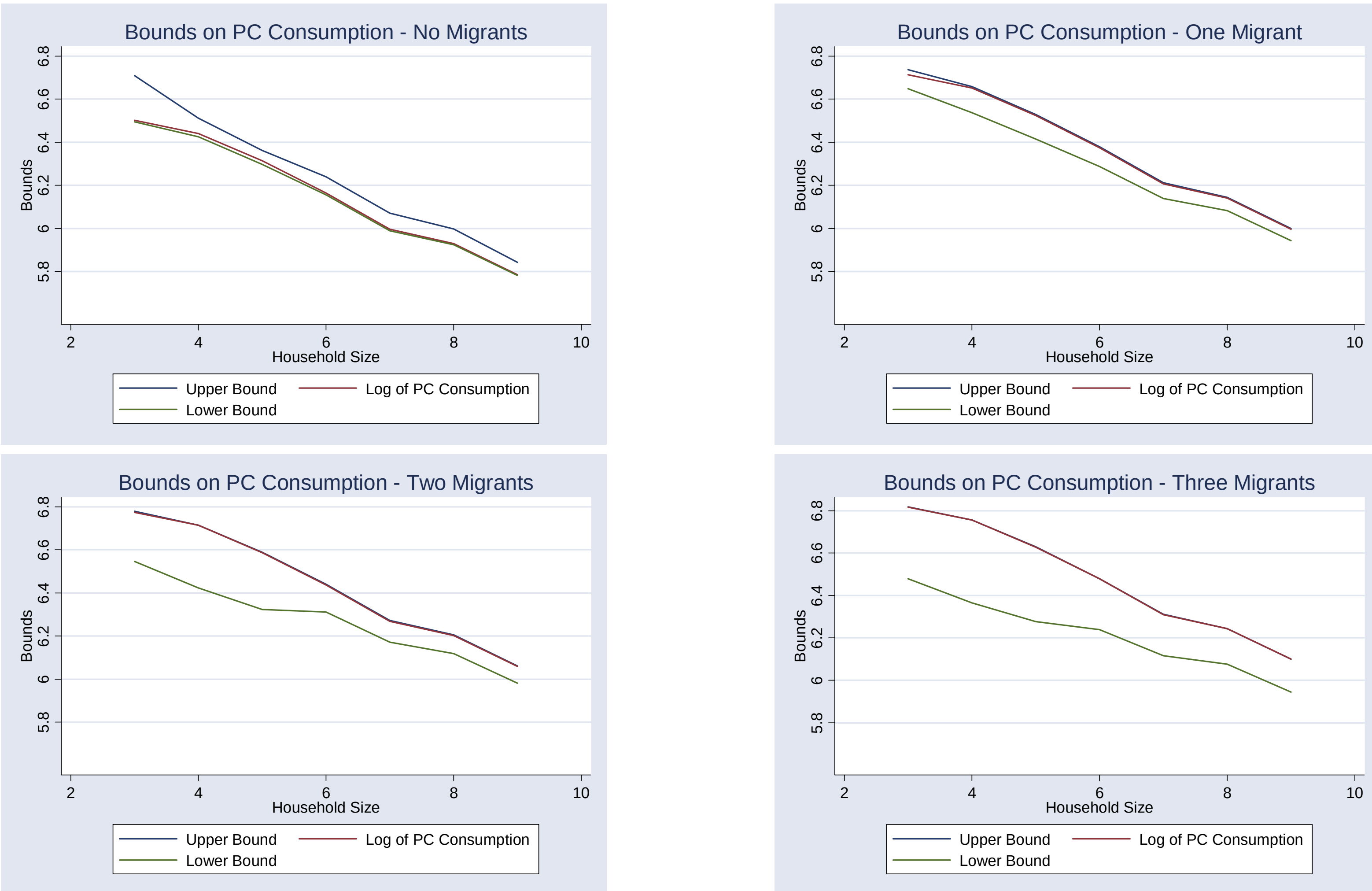


Figure 7: The Relationship between Estimates of Economies of Scale and the Income Elasticity of Food