DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 98-32

The Estimation of Linear Expenditure Demand Equations for Food with an Extended Set of Demographic Variables

Christopher D. Grewe
Department of Economics, University of Colorado at Boulder
Boulder, Colorado

November 1998

Center for Economic Analysis
Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

The Estimation of Linear Expenditure Demand Equations for Food with an Extended Set of Demographic Variables

There are three main themes in the applied demand literature: the formulation and testing of theoretically plausible functional forms, the development of econometric tools, and the consequence household characteristics have on demand. This paper falls squarely into the last camp. The literature has generally examined the effect which children have on a household's demand for goods or the cost of an additional child to a household. The reference household is taken to be either a single adult, or more commonly, two adults. The literature has so far not examined two other questions related to a household's composition. The most obvious oversight is that the age of the adults considered as the reference household has not been taken into account in these studies. Just as a teenager has different demands than an infant, an adult's pattern of consumption will change as he or she ages.

The other question not addressed in the literature is the possibility of extended families. The two adult household used as the reference in the literature is the norm for the developed economy data which underlies most of these studies, but it is not the only household composition available. The extended family, defined for the purposes of this paper as a household composed of more than two married adults and their attendant children, is a common institution in agricultural economies where family labor offers some advantages over hired labor. Families in economies without well developed pension systems commonly include several generations of adults as children care for their elderly parents. Other reasons for the existence of extended families easily come to mind.

The purpose of this paper is to estimate a set of demand equations for food which allow for extended families as one form of household composition. I am interested in the influence which allowing for extended families and multiple adult age groups has on the estimations. Using the exact form of the indirect utility function and the expenditure function, I calculate household equivalence indices which show the cost to the household of an additional member for each of seven age categories including three adult age groups. In turn, I examine these indices for evidence of the rationale for the formation of extended families. The data used to estimate the equations is drawn from the Indonesian Susenas survey of household consumption for 1990.

The functional form I use for the demand equations is derived from the Linear Expenditure System (LES). The LES offers the advantages of a familiar specification, is parsimonious in the number of parameters estimated, and has well defined forms for the indirect utility function and the expenditure function, which are necessary for the calculation of theoretically valid household equivalence scale. The demand equations are estimated individually using a Tobit maximum likelihood equation corrected for the presence of heteroscedasticity in the data.

The data used is from the 1990 Susenas survey of household consumption in Indonesia. Ten food variables are aggregated as the category average of the individual goods surveyed. A household specific price for each food variable is calculated as the share weighted average of a category's prices. The demographic variables which characterize the composition of the household are the number of individuals in seven age groups. In contrast to previous papers in this literature, there are three adult age groups in addition to four children's age groups.

The chapter proceeds as follows: Section 1 presents the Linear Expenditure Model, Barten's demographic scaling, and the transformed demand equations. An example with two goods and one demographic variable is presented to clarify the model and demonstrate the importance of allowing substitution between goods, a characteristic of the demographically scaled model. Section 2 discusses the data, particularly the variation in prices across provinces. The results of the estimations are covered in Sections 3.1 and 3.2. Section 3.1 presents the basic results of the estimations. The significance of the demographic variables are discussed, the average household scaling functions and the average adult equivalent quantities and effective prices are presented. Section 3.2 reports the elasticity of quantity to a change in the demographic variables, as well as the standard price elasticities. I then use the results of Section 3.1 to calculate Household Equivalence in Section 4. In turn, these are used to evaluate the impact of the demographic variables on a household's welfare. The conclusion summarizes the results.

1. The Model

There is a vast array of functional forms available for the specification of demand equations. Incorporating household composition into the demand equations also offers a wide range of choices. The model I present below has the advantage of integrating demographic characteristics into a theoretically plausible system of demand equations. While the restrictiveness of the LES is a drawback, it allows the estimation of the demand equations using available software, while still allowing the problems of zero consumption and heteroscedasticity to be addressed.

The demand equation for the Linear Expenditure System is:

$$(1) \qquad q_i = \beta_i + \frac{\alpha_i}{p_i}(\mu - \sum_k p_k \beta_k)$$

where q_i is the quantity of good *i* consumed by the household, p_i is the market price, μ is the total expenditure on the class of goods ***q***, and α and β are parameters. In a system of equations there are 2n parameters (where *n* is the number of goods) of which only 2n-1 need to be estimated because the restriction that $\Sigma\alpha_i = 1$ allows one equation to be dropped. A single demand equation will have n+1 parameters. The LES is derived from the Stone-Geary direct utility function.[1] The β's are usually interpreted as the necessary quantities of the goods consumed and α is the (constant) marginal budget share. The LES was derived by Stone by algebraically imposing the regularity conditions on a general linear demand function (Deaton and Muellbauer 1989 pg. 65).

Demographic scaling was first proposed by Barten (1964) and the treatment in this paper follows Muellbauer (1977) and Pollak and Wales (1992).[2] Given a theoretically plausible household demand function $q_i = \bar{h}_i(P, \mu)$, demographic scaling replaces this with:

$$(2) \qquad h_i(p, \mu) = m_i \bar{h}_i(p_1 m_1, \ldots, p_j m_j, \mu)$$

where m_i is the scaling function for good *i* and depends on the demographic variables.[3] If the effect of a demographic variable is the same for all goods, then we consider *m* to be the adult equivalent. If the m_i's vary according to the good, then we speak of a good specific adult equivalence scale. An adult equivalence scale is, effectively, a deflator for the quantities consumed by a household. For example, the adult equivalent of an additional child under 14 on a household's consumption of food might be .52 (Deaton and Muellbauer 1989, pg. 193).

Barten's demographic scaling is a generalization of Engel's approach and allows changing demographic variables to not only effect the absolute price of a good to a household but also its relative cost. In Gorman's (1976) oft quoted words, "a penny bun costs threepence when you've a wife and child". Demographic scaling allows for substitution away from relatively expensive

---

[1] $G(q) = \prod_{i=1}^{n} (q_i - \beta_i)^{\alpha_i}, \quad q_i > \beta_i, \quad \sum_{i=1}^{n} \alpha_i = 1$

[2] While this discussion is in terms of Barten's method of demographic scaling, it is worth noting that this is indistinguishable from demographic translating for the LES. The interpretation of the results is not effected.

[3] I am assuming the existence of a well behaved household utility function from which the demand equations are derived.

goods in the face of changes in their effective cost due to changing household composition. Following Pollak and Wales (1992), I use a linear scaling function:

$$m_i(h) = 1 + \sum_{j=1}^{N} g_{ij} h_j \qquad (3)$$

where h_j is the household characteristic variable and g_{ij} its corresponding parameter for the ith price. *N* indexes the demographic variables.

The role equivalence scales play in the estimation of demand requires a little clarification. Engel's idea was that households which have the same expenditure share for food must be enjoying the same level of welfare, regardless of the differences between their income or household composition. The ratio of the two households' money income then gives a measure of the cost of the differences between them. Scaling functions are meant to incorporate the effect household characteristics have on the demand for goods directly into the estimation of the household's behavior.

Consider a reference household *r*, usually thought of as one adult, although two adult households have sometimes been used as the reference.[4] The household maximizes utility by consuming goods, subject to its budget constraint: max $u^r = v(q_1, q_2, \ldots q_n)$ s.t. $\sum p_i q_i = m$. Suppose that household *h* is larger than the reference household and so consumes more of goods $q_1 \ldots q_n$. If the households are getting the same level of utility from their consumption then $u^r = u^h$. The m_i's deflate household *h*'s goods so that they are equivalent to the reference household's. *h*'s direct utility function becomes: $u^h = v(q_1/m_1, q_2/m_2, \ldots, q_n/m_n)$. That is, *h*'s utility is a function of the adult equivalent quantities of the goods. Household *h* also faces a budget constraint. Let

[4] Muellbauer 1977, for example.

$p_i m_i = p_i^*$ and $q_i / m_i = q_i^*$. h then optimizes $u = v(q_q^*, q_2^*, \ldots, q_n^*)$ s.t. $\sum p_i^* q_i^* = \mu$, which is equivalent to the reference households problem. p^*_i is the cost to household h of providing the adult equivalent quantity of good i to the members of h. The cost of a penny bun to h, in Gorman's example, is three cents because that is the cost of the adult equivalent quantity of buns the household consumes. I will refer to p^*_i as the effective price for good i.

When $m_i = 1$, u^h reduces to u^r, the reference household's problem. h's demand for good i is $q_i^* = h_i(p^*, \mu)$ or, multiplying by m_i, $q_i = m_i h_i(p^*, \mu)$ (equation 2 above). The cost function for household h is: $c(u, p^*) = \mu$. Household h's demand is a function of the effective prices $\boldsymbol{p}^*$ and so changes in the household's composition can effect the demand for a good both directly via m_i and indirectly via p^*, the change of the effective prices through m_i and $m_{j \neq i}$.

An example with two goods and one demographic variable helps clarify what is happening in a LES demand equation with demographic scaling.[5] The general demand equation is:

$$(4) \qquad q_i = \beta_i + \frac{\alpha_i}{p_i}(\mu - p_1\beta_1 - p_2\beta_2) \qquad i = 1,2$$

Assume there is one demographic variable, η, and that the scaling function is $m_i = 1 + \gamma_i \eta$. The transformed demand equations are now:[6]

$$(5) \qquad \begin{aligned} q_1 &= \beta_1 + \gamma_1\beta_1\eta + \frac{\alpha_1}{p_1}[\mu - p_1\beta_1 - \gamma_1\beta_1 p_1\eta - p_2\beta_2 - \gamma_2\beta_2 p_2\eta] \\ q_2 &= \beta_2 + \gamma_2\beta_2\eta + \frac{\alpha_2}{p_2}[\mu - p_1\beta_1 - \gamma_1\beta_1 p_1\eta - p_2\beta_2 - \gamma_2\beta_2 p_2\eta] \end{aligned}$$

The general effect of a change in η on the demand for good i is:

[5] The example is drawn from Pollak and Wales (1992) and Deaton and Muellbauer (1986).

[6] $q_1^* = \beta_i + \alpha_1 / p_1^* (\mu - p_1^*\beta_1 - p_2^*\beta_2)$ since $q_1^* = \frac{q_1}{m_1}$ and $p_1^* = p_1 m_1$, multiplying through by m_1 cancels out the m_i in the denominator of the marginal budget share, yielding the form of the equations which follow.

$$(6) \qquad \frac{\partial q_i}{\partial \eta} = \gamma_i \beta_i (1-\alpha_i) - \frac{\alpha_i}{p_i} \sum \gamma_k \beta_k p_k \qquad i \neq k$$

The effect of a increase in η on q_i depends on the signs of the γ's.[7] For case 1, where γ_1 and γ_2 both have the same sign, the effect on q_i is indeterminate. If γ_1=0 and γ_2>0, case 2, then q_1 will decrease and q_2 will rise and vice versa for γ_1>0 and γ_2=0. The last case is γ_1<0 and γ_2>0 which causes q_1 to fall and q_2 to rise. The effect on q_1 and q_2 is reversed when the signs are reversed. The change in q_1 and q_2 do not have to be equal, but if the budget constraint is binding the value of the changes must equal to zero.[8]

To demonstrate this, suppose that an increase in η (the birth of a child, for example) has γ_2>0 but γ_1=0. The effect on the demand for q_1 and q_2, for example wine and milk, can be seen in the partial derivatives: $\frac{\partial q_1}{\partial \eta} = -\alpha_1 \gamma_2 \beta_2 \frac{p_2}{p_1}$, $\frac{\partial q_2}{\partial \eta} = (1-\alpha_2)\gamma_2 \beta_2$. The effect is to increase the demand for milk (q_2) and decrease the demand for wine (q_1), as we would expect.

We can also look at the effect of an increase in η in terms of the equivalent household problem. Figure 1 shows the indifference curve and budget line for the household before and after the change in η for goods q_1^* and q_2^*. The initial equilibrium is at point A. After the change in η the budget line shifts from BB′ to BB″. This shift occurs because the change in η causes a change in the effective price of good 2. It becomes more costly to provide milk to the household, in this example, because the addition of an infant increases the number of equivalent adults who consume that good in the household. The birth of a child increases $p_2^* = m_2 p_2 = (1 + m_2 \eta) p_2$ while $p_1^* = m_1 p_1 = (1 + 0\eta) p_1$ is constant, so the budget line shifts as shown. In order to achieve the

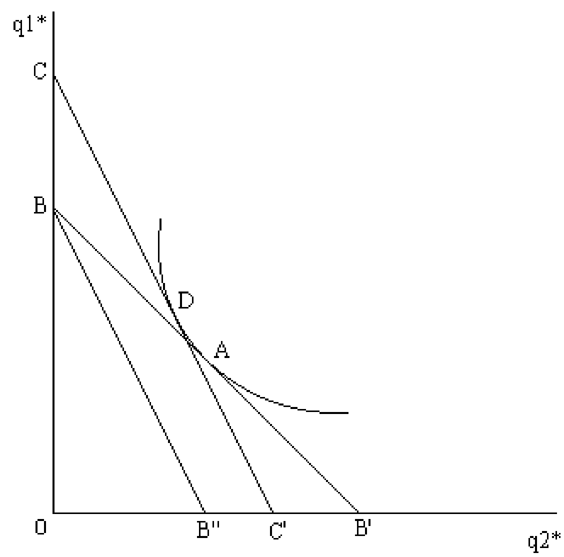
---

[7] In a two good case, both goods must be normal. Therefore the marginal budget shares are both positive. The β's are also assumed to be positive.

[8] That is $p_1(\partial q_1 / \partial \eta) + p_2(\partial q_2 / \partial \eta) = 0$.

original level of utility, the household would need sufficient income to shift the new budget line to CC′. At the new point of tangency D, the household is as well off as before the change, but the consumption bundle has changed. The consumption of the relatively more expensive good (q_2^*) decreases as q_1^* is substituted for it. Muellbauer (1977) calls this the quasi-price effect. The equivalence scale m_2 can be read from the vertical axis as OC/OB. Demographic scaling allows a household to adjust its consumption by substitution as its composition changes.

Figure 1



These two effects, the increase in the consumption of q_2 and the decrease in the consumption of q_2^* when h increases, appear contradictory. The difference between them is that the decrease in q_2^* refers to Household *h*'s adult equivalent of good 2 after the household has been compensated for the change in the effective price. That is, it is the pure substitution effect which remains after the income effect has been removed. The effect of a change in h is to increase the relative price of good 2 and so cause a substitution away from it. However the household actually consumes q_2 not q_2^* and faces prices p_1 and p_2. The change in η has a quasi-income effect on household h, since the budget constraint is still binding.[9] The effect of the change is to decrease the income of the household and consumption of both goods, but the consumption of q_1 falls by more than q_2. m_2 measures the cost of the increase in income needed to raise Household h back to its original level of utility.

Deaton (1986, pg. 1809) states "In my view, the problem of dealing appropriately with zero expenditures is currently one of the most pressing in applied demand analysis." Deaton notes that a Tobit model is one means of handling this problem. A problem with the Tobit model is that the maximum likelihood estimator is inconsistent when heteroscedasticity is present in the data (Greene 1993). Peterson and Waldman (1984) use a model of multiplicative hetero-scedasticity to modify the maximum likelihood equation.[10] Multiplicative heteroscedasticity allows the variance to be an exponential function of the explanatory variables and is fairly general with homoscedasticity as a special case.

The full model is:[11]

$$q_i = \begin{cases} m_i^i(\eta)\beta_i^i + \alpha_i/p_i\,(\mu - \sum_k m_k^i(\eta)\,p_k\,\beta_k^i) + \varepsilon_i & \text{if} \quad m_i^i(\eta)\beta_i^i + \alpha_i/p_i\,(\mu - \sum_k m_k^i(\eta)\,p_k\,\beta_k^i) + \varepsilon_i > 0 \\ 0 & \text{if} \quad m_i^i(\eta)\beta_i^i + \alpha_i/p_i\,(\mu - \sum_k m_k^i(\eta)\,p_k\,\beta_k^i) + \varepsilon_i < 0 \end{cases} \tag{7}$$

$$m_i^i(\eta) = 1 + \sum \gamma_{ij}^i \eta_j \quad i = 1-10 \quad j = 1-7$$

$$\varepsilon_i \sim N(0, \sigma_i^2) \quad \sigma_i^2 = \exp(\sum \delta_i^i z_i)$$

where α, β, γ, σ^2, and δ are vectors of unknown parameters to be estimated and the z_i are the heteroscedastic variables. The superscript indexes which equation the betas, gammas, and deltas

---

[9] If q_i is the quantity of good consumed by the reference household, then household h's budget constraint is: $p_1 q_1 m_1 + p_2 q_2 m_2 = \mu$. When q_1 equals zero then $q_2 = \mu / p_2\, m_2$, which obviously falls when η rises.

[10] Harvey (1976) is cited as a source for this specification.

[11] See Appendix 1 for expanded estimation and parameter equations.

are derived from, as all of these parameters can be calculated from any one of the equations. For example γ^{2}_{12} refers to the gamma parameter for the first good and the second demographic variable, cereal and toddlers respectively, estimated in the second demand equation (Tubers). The marginal budget shares, α_i, are unique to each equation and therefore do not need the superscript. The ***z***'s, as I explain below, are the goods prices and expenditure on food. Homoscedasticity occurs when $\delta_1 = \ldots = \delta_{10} = 0$. Additionally, other regional factors are controlled for by the inclusion of twenty-six dummy variables to control for a household's province.

At the risk of insulting the reader's intelligence, I also want to point out that the equations estimated are not the same as those in Equation 7 due to collecting terms. All of the parameters referred to above, except for the alphas (marginal budget shares), must be derived from the estimated coefficients. For the rest of the chapter I will use the term coefficients to refer to the estimated values and parameters to refer to the estimated values of the model (the betas, gammas, and deltas) derived from these coefficients. As an example, collecting the terms for the observable variables for equation 5 yields:

$$\text{(8)} \quad \begin{aligned} q_1 &= (1-\alpha_1)\beta_1 + (1-\alpha_1)\beta_1\gamma_1\eta + \alpha_1 \mu/p_1 - \alpha_1\beta_2\, p_2/p_1 - \alpha_1\beta_2\gamma_2\, p_2/p_1\, \eta \\ q_2 &= (1-\alpha_2)\beta_2 + (1-\alpha_2)\beta_2\gamma_2\eta + \alpha_2 \mu/p_2 - \alpha_2\beta_1\, p_1/p_2 - \alpha_2\beta_1\gamma_1\, p_1/p_2\, \eta \end{aligned}$$

which, with the addition of an error term, are the estimation equations:

$$\text{(9)} \quad \begin{aligned} q_1 &= \zeta^1_0 + \zeta^1_1\eta + \zeta^1_2\, \mu/p_1 + \zeta^1_3\, p_2/p_1 + \zeta^1_4\, p_2/p_1\, \eta + \varepsilon_1 \\ q_2 &= \zeta^2_0 + \zeta^2_1\eta + \zeta^2_2\, \mu/p_1 + \zeta^2_3\, p_2/p_1 + \zeta^2_4\, p_2/p_1\, \eta + \varepsilon_2 \end{aligned}$$

It follows that the β and γ parameters of the model are calculated from the ζ coefficients estimated in equation 9 as:

$$
(10)\qquad \begin{array}{llll} b_1^1 = \dfrac{z_0^1}{1 - z_2^1} & b_2^1 = \dfrac{z_3^1}{z_2^1} & g_1^1 = \dfrac{z_1^1}{z_0^1} & g_2^1 = \dfrac{z_4^1}{z_3^1} \\ b_1^2 = \dfrac{z_3^2}{z_2^2} & b_2^2 = \dfrac{z_0^2}{1 - z_2^2} & g_1^2 = \dfrac{z_1^2}{z_0^2} & g_2^2 = \dfrac{z_4^2}{z_3^2} \end{array}
$$

The parameters for the second equation are calculated the same with the appropriate changes in the super and subscripts. Appendix 1 gives an example of the model and the estimation equations for a two goods and two demographic variables and the equations relating the parameters to the estimated coefficients.

2. Price Variation

Cross-sectional data does not generally include sufficient variation in prices to allow the estimation of price effects in demand equations. The market price is generally assumed to be the same for all households in the cross-section, hence there is no price variation. However, as Pollak and Wales (1978, 1992) show, two household budget studies (generally from different times) are sufficient to provide enough price variation to determine all the coefficients for a linear expenditure system. The Susenas survey is a large cross-sectional survey of household consumption in Indonesia. In place of price variation over time, the Susenas survey offers variation over geography. There is ample variation in the household specific price of goods across provinces to allow the estimation of the price and income effects specified in the LES.

Table 1 gives the means and standard deviations of the variables used in the estimation of the demand equations of equation 7. The units of the quantities are 10 grams. However, even ten gram units are somewhat suspect for some of the goods. The quantity of condiments consumed, for example, is undoubtedly too high if measured in 10 gram units, given that this good is an aggregation of cooking oils and spices. The household specific price for each aggregate good is calculated by taking the average of the individual good’s weighted prices in a category for a household.[12] The weights are the share of the individual good's expenditure to the total category expenditure for that household. Weighting the prices by the expenditure share accomplishes two things; the aggregate price for each category will be less skewed by the inclusion of higher priced "luxury" goods that are less frequently purchased by the household, and the prices of goods that the household does not consume are zeroed out by the expenditure share. The demographic variables are defined as the number of household members in an age group: Infants 0-1 year, Toddlers 1-5 years, Children (School Age) 6-15 years, Teenagers 16-21 years, Young Adults 22-40 years, Older Adults 40-60 years, and Elderly 60+ years.

Table 1

Goods, Prices, and Demographic Variables Means

Quantities	Mean	Std Dev	Prices	Mean	Std Dev	Demographics	Mean	Std Dev
CEREALS	1167.42	633.88	P1	4.51	1.52	Infants	0.08	0.27
TUBERS	248.39	476.14	P2	1.66	1.27	Toddlers	0.54	0.73
FISH	361.48	379.63	P3	5.39	5.16	Children	1.20	1.26
MEAT	32.98	73.63	P4	28.13	9.25	Teenagers	0.58	0.84
DAIRY	129.71	384.50	P5	15.49	10.57	Young Adults	1.31	0.93
VEGETABLES	994.49	670.68	P6	0.92	0.74	Older Adults	0.70	0.79
FRUIT	271.83	344.89	P7	2.86	2.22	Elderly	0.22	0.50
CONDIMENTS	1050.83	1041.55	P8	1.75	1.24			
PREPARED	1008.59	840.85	P9	0.68	0.84			
TOBACCO	361.71	428.57	P10	3.59	1.76			

Units for means are ten grams for most goods. Prices are in Indonesian Rupiah

The variation in the household specific prices due to the geographic dispersion of the survey allows the identification of their parameters. A concern is that the price parameters are not identifying only the impact of different prices on demand, but are also capturing other factors such as cultural or religious differences. The dummy variables for a household's province are included

[12] An individual good's price for each household is calculated using the reported expenditure and quantities consumed and is treated as a market price. See Grewe 1998 for a detailed discussion of the data.

in the regression equations to control for these factors. The effect of household composition is controlled for by the demographic variables. Least squares regressions of price on the provincial dummies and demographic variables and simple correlations between price and the demographic variables confirm that the price parameters are capturing the effect of different prices on demand and not some conflation of factors. A detailed discussion can be found in Chapter Four of Grewe 1998.

3. Estimation Results

3.1 Significance of Demographic Variables

The results of the estimation of the individual equations presented below support the hypothesis that the extended demographic variables are significant. The R^2s of the OLS estimations, used by the program for the initial values for the maximum likelihood calculations, ranged from a low of .26 to a high of .58, a reasonable range for cross-sectional data. A test for the presence of heteroscedasticity shows that it is present in the data. The Tobit likelihood function specified with multiplicative heteroscedasticity corrects for this problem and the parameters are therefore consistent.

I use a likelihood ratio test to examine whether the coefficients on the demographic variables (including the price/demographic interaction variables) are jointly significant in the estimations. The test is the maximum likelihood counterpart to an F test of joint significance in ordinary least squares. As can be seen in Table 2, the null hypothesis that the demographic variables are not significant is firmly rejected. The critical value for the test is 104.21 at the .005% level of significance with 70 restrictions. The estimated intercepts (ζ^3_0 and ζ^7_0 respectively) for Fish and Fruit are significant at the 1% level in the Heteroscedastic Tobit equations without the demographic variables but are not significant when the demographic variables are included. I include the intercepts as example of the effect including the demographic variables has on the estimated coefficients.

Table 2

Heteroscedastic Tobit Intercepts w and w/o Demographics

Likelihood Ratio Test that all γ=0

	w/o Demographics		w/Demographics		Likelihood
Variable	Het. Tobit	Std Error	Het. Tobit	Std Error	Ratio Test
Cereals	0.4108*	0.0154	-0.3139*	0.0129	28516
Tubers	39.625*	2.3541	41.959*	3.9881	1159
Fish	0.5983*	0.1011	0.0276	0.1080	637
Meat	-9.1209*	0.4500	-8.8505*	0.5492	867
Dairy	-1.3442*	0.1176	-1.5755*	0.1235	1599
Vegetables	6.1609*	0.1907	2.7004*	0.1987	2291
Fruit	0.2215*	0.0935	0.0680	0.1038	817
Condiments	0.4311*	0.0270	0.2658*	0.0294	560
Prepared Fds	0.8010*	0.0191	0.2291*	0.0206	5728
Tobacco	-0.5328*	0.1753	-2.1933*	0.1810	1932

* Significant at the 1% level. Critical value for the likelihood ratio test is 104.21 at the .005% significance level.

The bias which results from the omission of the demographic variables is, generally, to cause the estimated coefficient to move farther from zero. The consequence of this for the model can be seen in one of two ways. The first can be seen by calculating the βs from the estimated intercepts (see equation 10) and interpreting these as the necessary quantities of the goods consumed by a household. The result of the bias is to overstate the minimum quantities of goods the household requires. Alternatively, the impact of the bias can be seen in the equation for the own-price elasticity: $e_i^i = (b_i(1-a_i)/q_i) - 1$. The first term is ζ^i_0, so the result of the bias seen in Table 2 will be to increase the own-price elasticity of good *i*, at least for this data.

The g_{ij}^{j}s are the parameters of the demographic variables in each equation. Table 3 presents the gammas from the individual equations. These gammas are derived from the coefficients of the demographic variables of the ith equation. That is, γ^1_{1j} is calculated from the coefficients on η_j from the estimated Cereals equation, γ^2_{2j} from the Tubers equation, and so on. These are the marginal effects of the demographic variable *j* on good *i*.

Table 3

Single Equation Gammas

Variable	i=	Cereals	Tubers	Fish	Meat	Dairy	Vegetables	Fruit	Condiments	Prepared Foods	Tobacco
Infants	γi1	-0.193	0	NA	0	-0.270	0.302	NA	0.249	0.359	-0.204
Toddlers	γi2	-0.452	0.077	NA	0.090	-0.047	0.213	NA	0.122	0.293	-0.045
Children	γi3	-0.610	0	NA	0.056	0.047	0.247	NA	0.133	0.399	0
Teenagers	γi4	-0.654	0	NA	0	-0.053	0.323	NA	0.177	0.686	-0.132
Young Adults	γi5	-0.614	0	NA	-0.103	-0.098	0.384	NA	0.218	0.860	-0.430
Older Adults	γi6	-0.659	0	NA	0	0	0.415	NA	0.180	0.766	-0.271
Elderly	γi7	-0.570	0	NA	0	0.032	0.261	NA	0.128	0.785	-0.049
Total Scaling	Mi	-1.7766	1.0409	NA	0.9791	0.8566	2.4793	NA	1.7906	3.9179	0.1162

All non-zero gammas are significant at the 10% level or better. Mi is calculated using average demographic values.

All of the gammas reported in Table 3 are significant at the 10% level or better, with many significant at the .01% level. The zero values represent gammas are where the estimated coefficients are not significant in a particular equation. The gammas for Fish (i=3) and Fruit (i=7) can not be calculated due to the zero intercepts for those equations.[13] The coefficients for the η_2-h_6 demographic variables are significant at the .1% level in the Fish equation, while the coefficients for h_1, h_4, and h_5, are significant at the .1% level for the Fruit equation.[14] At least some of the demographic variables affect the demand for these goods (Fish and Fruit), even though the size of the effect can not be calculated from the estimated coefficients.

---

[13] $g^3_{31} = (1-a_3)b^3_3 g^3_{31} / (1-a_3)b^3_3$ for example, where the numerator is the estimated coefficient of η_1 and the denominator is the equation intercept. See Equation 10 or Appendix 1 for an example of calculating the parameters from the estimation coefficients.

[14] The complete set of estimated coefficients are not presented for reasons of space. They can be found in Grewe 1998.

It is particularly worth noting that γ'_{i5}- γ'_{i7}, the marginal effects of the adult demographic variables, are all significant for five out of the ten foods. γ'_{i5} is significant for Meat and Dairy as well and γ'_{i7} is significant for Dairy. All of the equations except Tubers have at least one significant adult parameter (or coefficient in the case of Fish and Fruit). The inclusion of adult demographic variables and accounting for age differences between adults are both important to the estimation of the demand for food.

The patterns discernible in Table 3 are intuitively appealing for at least some of the goods. In general, the size of the effect of a demographic variable on demand for a good increases with the age of the household member, up to the elderly. As working adults consume more than children, this is an appealing result. The negative effect of the demographic variables on the demand for Tobacco also makes sense if it is a relative luxuries in the household budget. The negative effect of the demographic variables on the demand for cereal is somewhat surprising given that these are a basic staple of consumption. I would expect that consumption of rice would increase with the addition of a Young Adult, for example. The Tobit method corrects for the truncation of the distribution at zero and so I interpret these as indicating that an additional member causes a surplus in this good which the household desires to dispose of. As the households in this data are seventy percent rural and own-production is a significant component of the goods consumed, this seems a reasonable interpretation.

The M_i's of the last row of Table 3 are the total scaling effect on the demand for a particular good calculated for the average demographic variables ($m_i(\eta) = 1 + \sum \gamma_{ij}\bar{\eta}_j$ i=1-10, j=1-7). These are the total effect household composition has on the demand for a given good and are the adult equivalent scales which deflate the good for comparison to a single adult household. The overall effect on demand for all the goods is positive, except for Cereals. Household composition has the greatest effect on Prepared foods, Vegetables, Condiments (cooking oils primarily), and Cereals, increasing the quantity demand for the first three and reducing the demand for Cereals. The demand for the other goods relative to a single adult household are reduced by the household composition, especially Tobacco.

While the scaling functions discussed above are calculated with average demographic values, I can calculate these values for all of the households in the data set. Table 4 shows the values of mean scaling functions calculated using single equation gammas (M_i) for the total sample and the five most common household types. For space considerations, only three goods are reported. The goods; Cereals, Vegetables, and Prepared Foods, have the largest average share of food expenditure. Complete tables for each household type are available in Grewe 1998.

Table 4

Scaling Functions for Major Household Types and Primary Goods

	Household Size	N	Percent of Households	Cereals M_i	Vegetables M_i	Prepared Foods M_i
Average Full Sample	4.6		100%	-1.76	2.47	3.90
m5=2 m2=1	3	1750	3.80%	-0.68	1.98	3.01
m5=2 m2=1 m3=1	4	1591	3.45%	-1.29	2.23	3.41
m5=2 m2=1 m3=2	5	1110	2.41%	-1.90	2.48	3.81
m5=2 m3=2	4	1074	2.33%	-1.45	2.26	3.52
m5=2 m3=1	3	951	2.06%	-0.84	2.02	3.12

M1=Infants, M2=Toddlers, M3=Children, M4=Teenagers, M5=Young Adults, M6=Older Adults, M7=Elderly

Table 4 highlights an important feature of the scaling functions and their underlying gammas. The gammas capture the effect on demand of both the number of members of a given category and their age. The importance the age effect has on the scaling function depends on the good being considered. The scaling function (M_i) for a three member household increases as the age of the child increases for all three goods, but the difference is greater for Cereals than Prepared Foods or Vegetables.

The adult equivalent model presented in Section 1 frames the household's consumption problem in terms of q^* and p^*, the adult equivalent quantities of the goods it will consume and their effective price. By using q^* and p^*, all of the households are facing the same consumer problem, regardless of their composition. The theoretical reference household, where the scaling factor M_i is equal to one, has a single member whose demand is unaffected by their age. I calculate the predicted $\hat{q}_i^*$ for each household by dividing the quantity consumed of q_i by the household scaling factor M_i. $\hat{p}_i^*$ is calculated by multiplying p_i by M_i. Table 5 sets out the mean values for q, p, $\hat{q}_i^*$, and $\hat{p}_i^*$ for the full data set and the five most common household types. As with Table 4, I limit the goods to Cereals, Vegetables, and Prepared Foods for reasons of space.

Table 5

Mean Predicted Values of q* and p* for Major Household Types and Primary Goods

	Cereals		Vegetables		Prepared Foods	
	q	q*	q	q*	q	q*
Average Full Sample	1167.42	-839.09	994.49	398.60	1008.59	254.24
m5=2 m2=1	759.93	-1117.55	818.36	413.11	771.96	256.21
m5=2 m2=1 m3=1	973.16	-754.39	920.80	413.28	877.52	257.19
m5=2 m2=1 m3=2	1214.08	-638.99	1025.25	414.24	1013.68	265.99
m5=2 m3=2	1027.78	-709.80	965.32	426.75	892.04	253.56
m5=2 m3=1	790.02	-942.74	814.10	404.02	738.99	236.93
	p	p*	p	p*	p	p*
Average Full Sample	4.53	-7.87	0.97	2.38	1.08	4.11
m5=2 m2=1	4.53	-3.08	0.98	1.94	1.13	3.41
m5=2 m2=1 m3=1	4.53	-5.85	1.00	2.23	1.08	3.69
m5=2 m2=1 m3=2	4.51	-8.57	0.95	2.35	1.10	4.18
m5=2 m3=2	4.45	-6.44	0.94	2.13	1.07	3.77
m5=2 m3=1	4.53	-3.80	1.00	2.02	1.14	3.55

M1=Infants, M2=Toddlers, M3=Children, M4=Teenagers M5=Young Adults, M6=Older Adults, M7=Elderly

As I would expect, the unscaled quantities q for all three goods increase as the number and age of children in the household increase. The reasoning underlying the adult equivalent model, that the households are facing the same consumer problem in terms of q^* and p^*, leads me to expect that $\hat{q}_i^*$ for different households should be close to a constant if the model is doing a good job of explaining demand for the good. For the three goods presented, $\hat{q}_i^*$ is fairly close for the different household types. There are some anomalies, but the mean standard deviation for the five household types is smaller for $\hat{q}_i^*$. By the same logic, the values of $\hat{p}_i^*$ should be larger that that of p_i as should the standard deviations. As with $\hat{q}_i^*$, this is what is observed.

3.2 Elasticities

The statistical significance of the gammas is confirmation that the addition of adult demographic variables is valid for the estimation of the demand for food. However, the size of the coefficients tells us little about the magnitude of the total effect the individual variables have on demand, while the scaling function gives the effect of household composition on the demand for a good. Equation 11 shows that the effect of a change in a demographic variable on the demand for a good is the combination of the direct effect of the change on the minimum necessary quantity of the good (β_i) less the sum of the effects of the change on all the goods necessary quantities, weighted by the good's marginal budget share.

$$(11) \qquad \frac{\partial q_i}{\partial \eta_r} = \beta_i^i \gamma_{ir}^i (1-\alpha_i) - \frac{\alpha_i}{p_i} \sum_k \beta_k^i \gamma_{kr}^i p_k \qquad k \neq i$$

Calculating the elasticity for a change in a variable gives a better sense of the real impact the demographic variables have on the demand for a good. Table 6 shows the calculated elasticities (at average prices and demographic variables) for the gammas from Table 3. The elasticity is calculated as:

$$(12) \qquad \varepsilon_{\eta r} = \frac{\partial q_i}{\partial \eta_r} \frac{\bar{\eta}_r}{\bar{q}_i}$$

where $\bar{\eta}_r$ is the mean value for η_r and $\bar{q}_i$ is the mean value of q_i. The range for the gamma elasticities is minus to plus infinity, since the effect of the demographic variables can be infinitely large or small.

Table 6

Single equation Gamma Elasticities											
Variable	i=	Cereals	Tubers	Fish	Meat	Dairy	Vegetables	Fruit	Condiments	Prepared Foods	Tobacco
Infants	γi1	0.0011	0.0006	0.0000	0.0000	0.0025	0.0097	0.0019	0.0054	0.0053	0.0124
Toddlers	γi2	0.0607	0.0263	0.0915	-0.0138	0.0032	0.0450	0.0198	0.0165	0.0474	0.0324
Children	γi3	0.1979	-0.1182	0.1120	-0.0070	-0.0030	0.1135	0.0476	0.0527	0.1362	0.0552
Teenagers	γi4	0.1086	-0.0343	0.0505	0.0110	0.0062	0.0725	0.0340	0.0227	0.1129	0.0966
Young Adults	γi5	0.2382	-0.1591	0.4057	0.0837	0.0184	0.1753	0.0836	0.0973	0.3110	0.4648
Older Adults	γi6	0.1273	-0.0879	0.1789	0.0128	0.0004	0.1066	0.0398	0.0474	0.1395	0.1646
Elderly	γi7	0.0308	-0.0156	0.0475	0.0059	-0.0008	0.0227	0.0116	0.0090	0.0472	0.0190
Total Scaling	Mi	2.0760	-1.8712	-5.2723	-1.8989	10.0516	-0.5798	0.6028	-6.0094	-6.0094	0.6479

As can be seen from table 6, the size of the individual elasticities vary a great deal by demographic variable and by good. The good most sensitive to a change in any demographic variable is cereals. The elasticity of Tobacco to a change in the number of Young Adults is the most elastic of any good. Vegetables and Prepared Foods are the next most sensitive to a change in the demographic variables while Fruit and Meat are the least sensitive. The positive elasticities of cereals and vegetables demand to a change in the demographic variables makes sense in light of their role as staple goods. The relatively larger elasticities for fish compared to meat suggest that households increase the former as they grow, relative to the latter.

The last row is the scaling function ($m_i(\eta)$) elasticity for the total effect of the average household's composition on the demand for a good. I calculated these elasticities by calculating the predicted quantities demanded at the average prices and demographic variables, taking the difference between the average quantities and the predicted values demanded, and then calculating the elasticity as the difference times the average household size over the average quantity demanded ($e_{Mi} = (\bar{q}_i - \hat{q}_i) * \bar{J} / \bar{q}_i$ where J is the average household size). The elasticity will be negative if the predicted quantity is greater than the average quantity for a good, as is the case for Tubers, Fish, Meat, Vegetables, Condiments, and Prepared Foods. As with the individual elasticities, the main lesson conveyed by these Total Scaling elasticities is that changes in a household's composition will have an important effect on the demand for food goods.

While the coefficients in a standard LES are easily interpreted, demographic scaling does complicate the interpretation of the effect of marginal changes in market prices on demand for a good. The complications follow from the interaction of prices and demographic variables as shown in equation 13. Let $p_i^* = m_i p_i$ and $q_i^* = g_i(c(u,p^*),p^*) = h_i(u,p^*)$, where g_i is the Marshallian demand function and h_i is the Hicksian demand function. The Slutsky equation then, after some manipulation, is:

$$(13) \qquad \frac{\partial q_i^*}{\partial p_j} = \frac{\partial g_i}{\partial p_j^*} m_j + q_j \frac{\partial g_i}{\partial \mu}$$

That is, the pure substitution effect of an adult equivalent good for a change in the goods market price is modified by the demographic variables.

Own, Cross-price, and expenditure elasticities (the change in actual quantity demanded q_i for a change in the market prices p_i, p_j, or total expenditure on food m) are shown in Table 7. The elasticities are calculated per the formula in Pollak and Wales (1992, pp.5, equation 14 below) modified for the quantity and demographic scaled form.

$$(14) \qquad \begin{aligned} \varepsilon_{ii} &= \frac{(1-\alpha_i)(\beta_i + \sum_r \beta_i \gamma_{ir} \eta_r)}{q_i} - 1 \\ \varepsilon_{ij} &= -\frac{\alpha_i(\beta_j + \sum_r \beta_j \gamma_{jr} \eta_r)}{p_i} \frac{p_j}{q_i} \\ \varepsilon_{i\mu} &= \frac{\alpha_i}{p_i} \frac{\mu}{q_i} \end{aligned}$$

Table 7

Own and Cross-price Elasticities										
	Cereals	Tubers	Fish	Meat	Dairy	Vegetables	Fruit	Condiments	Prepared Foods	Tobacco
P1	**-0.5223**	0.2267	-0.0302	-0.0547	-0.0032	-0.0522	-0.0904	0.0225	-0.0606	-0.1010
P2	0.0035	**0.7583**	0.0086	-0.0285	0.0002	-0.0166	-0.0392	-0.0166	0.0122	0.0869
P3	0.0102	0.0484	**-0.8190**	-0.0296	0.0005	0.0027	-0.0074	0.0300	0.0217	-0.0028
P4	0.0480	-0.2443	-0.0239	**-1.2627**	-0.0081	-0.0536	-0.1434	-0.0583	-0.1246	-0.1583
P5	-0.0744	0.3547	-0.0470	-0.0100	**-1.1040**	0.0000	-0.0058	0.0194	-0.0157	-0.0566
P6	-0.0077	0.1090	-0.0173	-0.0155	-0.0001	**-0.3268**	0.0269	0.0021	0.0033	0.0415
P7	0.0052	-0.0468	0.0000	0.0025	0.0003	-0.0121	**-0.9072**	0.0456	0.0195	0.0116
P8	0.0225	0.1419	0.0346	0.0242	-0.0007	-0.0122	-0.0266	**-0.5470**	0.0090	-0.0217
P9	-0.0030	0.1013	-0.0099	0.0096	-0.0003	0.0037	0.0271	-0.0139	**-0.1100**	-0.0244
P10	-0.0400	-0.1467	-0.0162	-0.0199	0.0009	-0.0001	-0.0072	0.0181	0.0602	**-1.0705**
Income	0.1429	0.1216	3.3924	1.7255	0.0535	0.3075	0.5711	4.1007	0.1867	0.5375

Calculated using single equation coefficients, average price and demographic variables, quantities, and expenditure.

Except for Tubers, all the own-price elasticities have the expected sign. Cereals, Vegetables, and Prepared Foods are the most inelastic, as you would expect for staple goods. Similarly, it is not surprising to see that Meat is the most elastic good, followed by Dairy and Tobacco. The own price elasticity for Tubers is positive since the ratio of the change in quantity demanded for a change in price over the average household quantity demanded is greater than one. It is possible that Tubers (composed primarily of subsistence goods such as Cassava, Sweet Potatoes, and Taro) is an inferior good, at least for the average household composition and expenditure. The own-price elasticity for Tubers, calculated for the households in the top quintile of food expenditure, is -.219, supporting the hypothesis that this is an inferior good for low income households. The cross-price elasticities are all small except for ε_{21} (the elasticity of Tuber with respect to a change in Cereals price).

The income elasticities also conform to expectations, for the most part. As expected Cereals, Tubers, Vegetables, and Prepared Foods can be ranked as necessary goods, with income elasticities below .5. Dairy also has a low income elasticity, which is surprising given its high own price elasticity and its low share of food expenditure. Prepared Foods rank as a necessity, but this may be due to the wide range of goods included in this category. Fruits and Tobacco are both also income inelastic. Fish, Meat, and Condiments are all very income elastic, which seems reasonable.

The primary goal for estimating these equations, examining the impact of an extended set of demographic variables on demand, is satisfied by these results. The likelihood ratio test of the significance of the demographic variables shows that these variables are important in these estimations. The elasticities presented in Tables 6 and 7 show that the demographic variables are not only significant, but also have a substantial effect on the demand for goods. The differences between the elasticities of Young Adults and Older Adults suggest that the age of adults is an important demographic characteristic.

4. Household Equivalence Scales and Cost of Living Indices

A principal application of demand estimations which incorporate demographic variables is the comparison of welfare of different households. A household equivalence index is the tool used for this task. The effect of children on the consumption of households has been recognized as important since Engel. The question has been commonly phrased as the effect an additional child has on the welfare of the household. Since welfare is not directly measurable, the question is stated in terms of the cost of an additional child. Engel proposed that if the shares two different households spend on a good are the same then the households are achieving the same level of satisfaction from the goods they are consuming. If the two households have different composition and different incomes, then the ratio of the household’s incomes gives a measure of the cost of the differences between the households.

Engel’s insight into the measurement of the cost of children or, for that matter, any other difference between households, has been the focus of a great deal of research. Barten (1964) developed his method of demographic scaling to provide a better means of allowing changes in composition to effect a household's demand for goods. Muellbauer (1974) extends Barten's and Diewert's work to show that the calculation of the expenditure ratios used for welfare comparisons is the same as the calculation of True Cost of Living indices of the effect of price changes on a household.

A True Cost of Living index is the constant utility ratio of expenditure functions which calculates the cost for an individual or household to reach a reference level of utility. If the reference level of utility is that of the initial period then the index is a Laspeyres-Konus and an index using the second period is a Paasche-Konus.[15] If c(p,u) is the expenditure function then the respective indexes are:

$$(14) \qquad \frac{c(p_1, u_0)}{c(p_0, u_0)} \quad \text{and} \quad \frac{c(p_1, u_1)}{c(p_0, u_1)}$$

where 1 and 2 are the time period and u is the reference level of utility. The True cost of Living indexes form the lower and upper bounds to the commonly calculated Laspeyres and Paasche indexes.

If m_1 is the scaling function for a reference household and m_2 the scaling function for any other household with a different composition, then we can calculate an index (which I will refer to as a Household Equivalence Index) that gives the relative cost of achieving a reference level of utility for the two households. As with the True Cost of Living indexes, the Household Equivalence Index depends on whether the reference utility used is the reference household's or the comparison household's. Ease of calculation leads me to use the Laspeyres-Konus index in the table presented below. That is, I use the reference household's utility in the calculation of both expenditure functions. The index is formally:

$$(15) \qquad \frac{c(m_2 p, u_1)}{c(m_1 p, u_1)}$$

As I noted in the chapter introduction, the LES has the advantage of having well defined equations for both the indirect utility and expenditure functions. These functions (adjusted for demographic scaling) are:

$$(16) \qquad \begin{array}{ll} \text{Indirect Utility :} & u^* = (\mu - \sum \beta_i m_i p_i) \prod \left(\frac{\alpha_i}{m_i p_i} \right)^{\alpha_i} \\ \text{Expenditure Function :} & c(mp, u) = \sum p_i m_i \beta_i + u^* \prod (p_i m_i)^{\alpha_i} \end{array}$$

Equation 1 is derived from the indirect utility function by the application of Roy's identity. The terms in the indirect utility function and the expenditure function are the same as those defined in Section 1.

Both of these functions have straightforward interpretations. Indirect utility is based on the amount of goods the household can consume after accounting for the minimum expenditure on necessities, weighted by the marginal cost of goods (i.e., the real supernumerary goods consumed).[16] The expenditure function is the minimum cost of buying the necessary goods plus the cost of the supernumerary expenditure represented by u^*. The indirect utility function and expenditure function are directly effected by the magnitude and sign of the gamma parameters via their effect on the scaling function, which in turn effects the quantity of the households necessities and the marginal cost of living.

The Household Equivalence Indices in Table 8 calculate the cost of an additional household member in the seven age groups for six different representative reference households. The reference households are among the most common compositions. A common feature in the

---

[15] The terminology follows Konus (1939). Diewert (1981) is an excellent survey of the theory of index numbers.

literature of equivalence scales is also observed here. The cost of children increases with age, as can be seen from the change in the household equivalence scale for Toddlers and School age children. The cost of each additional child or adult in a given age category is the same as the initial addition. That is, if one school age child increases household expenditure by 15%, then two children will cost an additional 30% of the reference expenditure. This is due to the linearity of the scaling function (see equation 3). The cost of one Elderly is between that of a Toddler and a School Age Child in all six tables. Adults are one to thirteen percent more costly than teenagers.

Table 8

Household Equivalence Tables - Laspeyres-Konus
Calculated using Single Equation Gammas

Reference Household; M5=2		Reference Household; M6=2		Reference Household; M6=1	
Infant	1.16	Infant	1.14	Infant	1.20
Toddler	1.13	Toddler	1.15	Toddler	1.18
School age	1.17	School age	1.17	School age	1.21
Teen	1.11	Teen	1.20	Teen	1.25
Young Adult	1.24	Young Adult	1.15	Young Adult	1.26
Mature Adult	1.23	Mature Adult	1.20	Mature Adult	1.26
Elderly	1.14	Elderly	1.16	Elderly	1.19
Reference Household; M5=2, M2=1		**Reference Household; M5=2 M2=1 M3=2**		**Reference Household; M5=1 M6=2**	
Infant	1.15	Infant	1.11	Infant	1.14
Toddler	1.11	Toddler	1.09	Toddler	1.13
School age	1.15	School age	1.13	School age	1.15
Teen	1.19	Teen	1.16	Teen	1.19
Young Adult	1.22	Young Adult	1.18	Young Adult	1.22
Mature Adult	1.21	Mature Adult	1.17	Mature Adult	1.20
Elderly	1.12	Elderly	1.10	Elderly	1.14

M1=Infants, M2=Toddlers, M3=Children, M4=Teenagers
M5=Young Adults, M6=Older Adults, M7=Elderly

The composition of the reference household effects the cost of adding an additional member of any age group. For instance, the cost of adding a School Age child to a household is 6% less for a three adults household than for a single Mature adult. Part of this is because the

[16] Deaton and Muellbauer (1986, pp. 65) state "Since the β's add to unity, this last term can be thought of as a weighted geometric mean of the prices and can thus be thought of as a price index representing the marginal cost of living." Their β's are the same as my α's.

reference expenditure for the three adult household is larger than that of one adult. If the increase in expenditure due to the addition of a child is the same for both households then the percentage increase in household expenditure will be smaller for the larger household because of the larger base. However this accounts for only about one percent of the absolute difference in the equivalence scales between the larger and smaller households. The remainder of the difference is the effect the different household compositions have on the expenditure for necessary levels of goods and the marginal cost of living. Note that average total food expenditure (μ in the indirect utility function) is constant across all households and that utility (u^*) is held constant at the reference household's level for a given set of equivalence scales.

The Household Equivalence indices of Table 8 demonstrate one of the economic rationales for larger households. The cost of children and elderly are relatively lower for these households than for smaller households, at least for expenditure on food. Moreover the formation of extended households, by adding adult members, will increase the expenditure on food by only 17-26%. If the new adult members bring additional income into the household which exceeds the additional expenditure of the household, it is reasonable to expect that the household will be better off, assuming that the supernumerary income is at least partially shared with the rest of the household.

The Household Equivalence scales presented in Table 8 are notably lower than the adult equivalence scales presented in the literature, although not unreasonably so. Browning (1992, pp.1444) compiles a table of some of the estimated Adult Equivalence Scales from a variety of sources. The equivalence scales for the cost of a child to a two adult household range from a low of 12 to a high of 100, with a mean of 51 for the nine papers cited. Muellbauer (1977), using a PIGLOG specification of the functional form, Barten's model of demographic scaling, and UK expenditure data, gets equivalence scales for a household with two adults and 1 child 0-5 years of 1.25 for food and 1.47 for a household with children 5-16 years. [17]

The differences between the equivalence scales in Table 8 and those in the literature stem from several causes: 1) The equivalence scales in Table 8 are only for food expenditure, while most of the equivalence scales cited above are for general adult equivalence. When the effect of additional children on a households total expenditure is accounted for, it is possible that the equivalence scale will be higher than that of food only. 2) The model used in estimating the equivalence scales presented in this paper allows explicitly for the substitution between goods for households of different composition. As the two good example in Section 1 demonstrates, the cost of reaching a given level of utility is lower when the household can substitute away from the relatively more expensive goods. Much of the literature on cost of children has not taken into account the substitution between goods and so the calculated equivalence scales should be higher than those in Table 8.

The last cause of the difference between the average equivalence scale in the literature and those in Table 8 is the explicit inclusion of adult demographic variables in the model. While the data sets used in most studies of the cost of children are restricted to only those households with two adults, no account of the age of the adults is taken. As a result, I suspect that the adult equivalence scales are partially inflated by the differences between younger and older adult households. In turn, this causes the cost of children to be overstated.

5. Conclusion

[17] A brief description of this paper and a table of his results can be found in Deaton and Muellbauer (1986) pages 202-205.

The main purpose of this paper is to extend the set of demographic variables used to modify the estimation of demand equations. The standard set of demographic variables are the number of children in a range of age groups. The equations estimated in this paper add three adult age groups to allow for the possibility of extended families and differences in demand patterns between adult age groups. The model used is a Linear Expenditure System modified by Barten's method of demographic scaling. The equations are estimated individually using a heteroscedastic corrected Tobit on Indonesian consumption data from the 1990 Susenas survey. The model allows for substitution between goods as household composition changes, which avoids overstating the cost of an additional member to a reference household.

The significance of household composition in estimating demand is confirmed by the results presented above. The significance of controlling for the presence of children when estimating household demand is well established in the applied demand literature. My results show that allowing for extended family structures, effectively the inclusion of more than two adults and different age groups, is equally significant in the estimation of demand, at least in countries where extended families are an meaningful demographic feature. While the marginal effect of a demographic variable on a given good is generally small, the cumulative effect across goods on the household's expenditure is sizable as can be seen in the table (Table 4.10) of the scaling functions values.

The impact on a household of the addition of another member is shown by the Household Equivalence Index. The key role played by the household's composition is established by the differences in the cost of reaching a given level of utility for an additional member for different reference households. Larger households can add another child for relatively less cost than a smaller household. More generally, the Household Equivalence indices show the economic rational for extended families. The cost of an additional adult ranges from 15% to 22% more for food expenditure. As most adults will contribute income or labor to the household they are joining, the enlarged household can be expected to be better off if the new members contribution exceeds the additional cost.

Despite the restrictiveness of the Linear Expenditure functional form and the compromises required by the constraints of the data, the estimations presented above perform reasonably well. Nevertheless there are some troubling problems with the results. In particular, the presence of negative betas suggest that the range of goods is too broad for their demand to be estimated using linear expenditure equations. The suspicion that the demand equations estimated above are not representing the true underlying preferences casts doubt on the conclusions about the role played by the demographic variables. Muellbauer (1977) finds that the substitution allowed by demographic scaling is overstated and leads to household equivalence indexes which are implausibly low. Despite these problems, the conclusion that adult variables are important to the estimation of demand is clearly demonstrated.

References

Amemiya, Takeshi 1974, “Multivariate Regression and Simultaneous Equation Models when the Dependent Variables are Truncated Normal”, Econometrica, V. 42 No. 6 pp. 999-1012

Barten, Anton P. 1964, “Family Composition, Prices and Expenditure Patterns” in *Econometric Analysis for National Economic Planning: 16th Symposium of the Colston Society*, ed. Peter Hart, Gordon Mills, and John K. Whitaker, Butterworth, London

Browning, Martin 1992, “Children and Household Economic Behavior”, Journal of Economic Literature, V30 pp. 1434-1475

Deaton, Angus 1986, “Demand Analysis”, in *Handbook of Econometrics, Vol. III*, ed. by Z. Griliches and M.D. Intriligator, Elsevier Science Publishers, Amsterdam

Deaton, Angus and Muellbauer, John 1986, "On Measuring Child Costs: With Applications to Poor Countries", Journal of Political Economy, V94 No. 4 pp.720-744

Deaton, Angus and Muellbauer, John 1989, *Economics and Consumer Behavior*, Cambridge University Press, Cambridge

Diewert, W. E. 1981, "The Economic Theory of Index Numbers: A Survey" in *The Theory and Measurement of Consumer Behaviour* ed. by Angus Deaton, Cambridge University Press, Cambridge, London, New York

Gorman, W. M. 1976, "Tricks with Utility Functions", in *Essays in Economic Analysis: Proceedings of the 1975 AUTE Conference, Sheffield*, ed. M. J. Artis and A. R. Nobay, Cambridge University Press, Cambridge

Greene, William H. 1993, *Econometric Analysis second edition*, Macmillan Publishing Company, New York

Muellbauer, John 1974, "Household Composition, Engel Curves and Welfare Comparisons Between Households: A Duality Approach", European Economic Review V5 pp.103-122

Muellbauer, John 1977, "Testing the Barten Model of Household Composition Effects and the Cost of Children", The Economic Journal, V87 pp. 460-487

Peterson, David R. and Donald M. Waldman 1984, "A Model of Heterogeneous Expectations as a Determinant of Short Sales", The Journal of Financial Research, V7 No1 pp.1-15

Pollak, Robert A. and Terrence J. Wales 1978, "Estimation of Complete Demand Systems from Household Budget Data: The Linear and Quadratic Expenditure Systems", American Economic Review, V68 No.3 pp. 348-359

Pollak, Robert A. and Terrence J. Wales 1980, "Comparison of the Quadratic Expenditure System and Translog Demand Systems with Alternative Specifications of Demographic Effects", Econometrica, V48 No.3 pp. 595-612

Pollak, Robert A. and Terrence J. Wales 1981, "Demographic Variables in Demand Analysis", Econometrica, V49 No.6 pp.1533-1551

Pollak, Robert A. and Terrence J. Wales 1992, *Demand System Specification and Estimation*, Oxford University Press, Oxford

Appendix 1

The equations are for two goods and two demographic variables. The extension to ten goods and seven demographic variables easily follows.

$$q_1 = b_1(1+g_{11}h_1+g_{12}h_2)+\frac{a_1}{p_1}[m- p_1b_1(1+g_{11}h_1+g_{12}h_2)-p_2b_2(1+g_{21}h_1+g_{22}h_2)]$$

$$q_2 = b_2(1+g_{21}h_1+g_{22}h_2)+\frac{a_2}{p_2}[m- p_1b_1(1+g_{11}h_1+g_{12}h_2)-p_2b_2(1+g_{21}h_1+g_{22}h_2)]$$

Expanding and collecting terms gives :

$$q_1 = (1-a_1)b_1+(1-a_1)b_1g_{11}h_1+(1-a_1)b_1g_{12}h_2+a_1\frac{m}{p_1}-a_1b_2\frac{p_2}{p_1}-a_1b_2g_{21}\frac{p_2}{p_1}h_1-a_1b_2g_{22}\frac{p_2}{p_1}h_2$$

$$q_2 = (1-a_2)b_2+(1-a_2)b_2g_{21}h_1+(1-a_2)b_2g_{22}h_2+a_2\frac{m}{p_2}-a_2b_1\frac{p_1}{p_2}-a_2b_1g_{11}\frac{p_1}{p_2}h_1-a_2b_1g_{12}\frac{p_1}{p_2}h_2$$

Letting z_{ij} represent the coefficient on the observed variables, the estimation equations are :

$$q_1 = z_{10}+z_{11}h_1+z_{12}h_2+z_{13}\frac{m}{p_1}+z_{14}\frac{p_2}{p_1}+z_{15}\frac{p_2}{p_1}h_1+z_{16}\frac{p_2}{p_1}h_2+z_{17}d_1+\cdots z_{133}d_{33}+e_1$$

$$q_2 = z_{20}+z_{21}h_1+z_{22}h_2+z_{23}\frac{m}{p_2}+z_{24}\frac{p_1}{p_2}+z_{25}\frac{p_1}{p_2}h_1+z_{26}\frac{p_1}{p_2}h_2+z_{27}d_1+\cdots z_{233}d_{33}+e_2$$

The parameters a_i, b_i, and g_{ij} are then derived from the estimated coefficients z_{ij} (superscripts on the parameters indicate which equation's coefficients are being used). d_i are the dummies variables for province. e is the error term. :

$$b_1^1=\frac{z_{10}}{1-z_{13}}\quad b_2^1=\frac{z_{14}}{z_{13}}\quad g_{11}^1=\frac{z_{11}}{z_{10}}\quad g_{12}^1=\frac{z_{12}}{z_{10}}\quad g_{21}^1=\frac{z_{15}}{z_{14}}\quad g_{22}^1=\frac{z_{16}}{z_{14}}\quad a_1=z_{13}$$

$$b_2^2=\frac{z_{20}}{1-z_{23}}\quad b_1^2=\frac{z_{24}}{z_{23}}\quad g_{21}^2=\frac{z_{21}}{z_{20}}\quad g_{22}^2=\frac{z_{22}}{z_{20}}\quad g_{11}^2=\frac{z_{25}}{z_{24}}\quad g_{12}^2=\frac{z_{26}}{z_{24}}\quad a_2=z_{23}$$