Modelling Zeroes in Micro Data

Jane M. Fry^{*}, Tim R. L. Fry[#], Keith R. McLaren[#] and Tanya N. Smith^{#1}

^{*} Centre of Policy Studies

[#] Department of Econometrics and Business Statistics

Monash University, Clayton, Victoria 3168

Australia

12 May 1998

Abstract: Although the literature contains a number of suggestions for dealing with problems caused by a preponderance of zero expenditure observations that frequently occur in micro level budget studies, in general, these suggestions seem to be either empirically intractable or theoretically unappealing. In this paper it is argued that a natural theoretical specification can be motivated by duality theory and that the statistical technique of compositional data analysis provides a corresponding complementary stochastic specification. The resulting model is a consistent theoretical and stochastic specification for handling the possibility of a zero demand over a range of expenditures and/or prices. The model is then applied to the 1988/89 Australian Household Expenditure Survey.

Keywords: Demand systems, Compositional Data Analysis, stochastic specification, corner solutions, micro level budget studies.

J. E. L. Classification: D12, C30.

Correspondence to: Jane M. Fry, Centre of Policy Studies, Monash University, Wellington Road, Clayton, Victoria 3168, Australia.

¹ Part of this research was funded by an Australian Research Council Grant.

1. Introduction

Zero expenditures are a common occurrence in micro data sets and pose both economic theoretic and estimation problems. Previous approaches to modelling data sets with zeroes have included the use of Kuhn-Tucker conditions in the model specification and the use of censoring methods in estimation. In this paper it is argued that a natural theoretical specification can be motivated by duality theory and that the statistical technique of compositional data analysis provides a corresponding complementary stochastic specification. The resulting model is a consistent theoretical and stochastic specification for handling data sets with a preponderance of zeroes and has the desirable property that the number of commodities consumed may expand (or contract) as real expenditure and/or relative prices change.

2. Background Theory

Let Ω^N represent the non-negative orthant, and let Ω_+^N represent the strictly positive orthant. Let $q \in \Omega^N$ represent an N vector of commodities, $p \in \Omega_+^N$ the corresponding N -vector of prices, and $c \in \Omega_+^1$ a level of expenditure. Preferences are represented by the (direct) utility function $u=U(q)$ where the notation is indicative of the notation to be used in general. That is, upper case letters will be used to denote functions, to distinguish these functions from the corresponding values of variables, to be denoted by lower case letters.

The direct utility function is assumed to satisfy the following regularity conditions RU:

$$\text{RU1: } U: \Omega^N \rightarrow R$$

RU2: U is continuous

RU3: U is non-decreasing and strictly increasing in at least one argument

RU4: U is strictly quasi-concave.

The Marshallian demand equations $Q^M(c, p)$ are defined as solutions to the maximisation problem

$$\max_q \{U(q) : p'q \leq c, q \in \text{domain}(U)\}. \quad (2.1)$$

The system of Marshallian demands will inherit regularity properties from the properties RU and from the fact that they are solutions of the constrained optimisation problem (2.1).

Dual to $U(q)$ is the indirect utility function defined by

$$U^M(c, p) = U(Q^M(c, p)) \quad (2.2)$$

the maximised value of utility conditional on given expenditure and prices. Because of this conditioning, the superscript M for Marshallian is used to distinguish the indirect utility function from the direct utility function, because the demand equations with the same conditioning are usually referred to as Marshallian demands. The indirect utility function will inherit the regularity conditions RIU:

$$\text{RIU1: } U^M : \Omega_+^{N+1} \rightarrow R$$

$$\text{RIU2: } U^M \text{ is continuous}$$

$$\text{RIU3: } U^M \text{ is homogeneous of degree zero in } (c, p)$$

$$\text{RIU4: } U^M \text{ is non-increasing in } p \text{ and strictly decreasing in at least one price}$$

$$\text{RIU5: } U^M \text{ is increasing in } c$$

$$\text{RIU6: } U^M \text{ is quasi-convex in } p.$$

The Marshallian demand equations are related to the indirect utility function via Roy's Identity

$$Q_i^M(c, p) = -\frac{U_{p_i}^M(c, p)}{U_c^M(c, p)}. \quad (2.3)$$

Also dual to both the direct utility function and the indirect utility function is the cost function, defined by

$$\begin{aligned} c &= C(u, p) \\ &= \min_q \{p'q : U(q) \geq u, q \in \Omega^N\} \\ &= \min_c \{c : U^M(c, p) \geq u, c \in \Omega^1\} \\ &= p'Q^H(u, p) \end{aligned} \quad (2.4)$$

where the $Q^H(u, p)$ are the Hicksian or utility compensated demand equations. The cost function will inherit the set of regularity conditions RC:

$$\text{RC1: } C: R \times \Omega^N \rightarrow \Omega^1$$

RC2: C is continuous

RC3: C is homogeneous of degree one in p

RC4: C is non-decreasing in p and strictly increasing in at least one price

RC5: C is increasing in u

RC6: C is concave in p .

The Hicksian demand equations are related to the cost function and by Shephard's Lemma can be derived by differentiation with respect to prices

$$Q_i^H(u, p) = C_{p_i}(u, p). \quad (2.5)$$

Duality theory is concerned with the fact that preferences may be represented equivalently by a direct utility function satisfying RU, an indirect utility function satisfying RIU or a cost function satisfying RC. It is usually argued that starting with a direct utility

function is infeasible for all but the simplest of cases because of the need to derive the analytical solution to an optimisation problem, and that if the aim of the exercise is to derive a system of regular Marshallian demands, then one should begin with an indirect utility function satisfying RIU and apply Roy's Identity. Alternatively, if the aim of the exercise is to derive a system of regular Hicksian demand equations, then one should begin with a cost function satisfying RC and apply Shephard's Lemma. Since in most applied work it is expenditure and prices that are treated as exogenous, it is natural to search for general specifications of indirect utility functions that satisfy RIU. Cooper and McLaren (1992, 1996) are examples of this approach.

3. The Problem of Zeroes

3.1. Introduction

In practice, regardless of whether demand systems are derived from direct utility functions satisfying RU, indirect utility functions satisfying RIU or cost functions satisfying RC, they typically have the property that demands for all commodities are strictly positive for all admissible values of expenditure and prices. This is presumably a consequence of the need to find mathematical functions satisfying the regularity conditions, and in the case of the direct utility function the need to explicitly solve the first order conditions. This leads to the use of functions such as logarithmic functions which only allow zeroes as limiting solutions, in order to avoid violation of the regularity conditions. While this is not a problem with highly aggregated time series data sets, it does present a conceptual problem when working with disaggregated micro level data, which are usually characterised by a preponderance of 'zero' observations. The econometrics literature has attempted to solve this problem in one of three ways (see Pudney (1989), pp. 153-158).

One approach is to relax the regularity conditions on the direct utility function and then to impose the non-negativity condition as part of the defining optimisation problem. Condition RU1 is relaxed, allowing the direct utility function to be defined over potentially negative values but the constrained optimisation is replaced with

$$\max_q \{U(q): p'q \leq c, q \in \Omega^N\} . \quad (3.1)$$

While this will in principle achieve the desired result, the optimisation problem (3.1) leads to a set of Kuhn-Tucker conditions which is too complex for empirical work in all but the simplest of cases. An example of this approach is Wales and Woodland (1983). However, it could be argued that it is precisely in the avoidance of such problems that duality theory should come to the fore as a theoretical tool. The second approach, characterised by Lee and Pitt (1986), is cast in terms of using a dual representation to find the shadow prices which turn negative demands into zero demands and specifying a set of switching regressions. Again, the order of complexity in estimation is similar to that of the approach of Wales and Woodland (1983). A third approach is to accommodate the problem of observed zeroes at the estimation phase, by the use of censored regression or other related methods. See, for example, Heien and Wessells (1990). The two main disadvantages with this approach are that censoring will almost certainly destroy the adding up (and possibly other) conditions, and that the estimation method may again become too complex for all but the simplest of cases.

The approach to be proposed in this paper is a combination of an appropriate specification of an indirect utility function, which is derived in the next subsection, together with an appropriate stochastic specification which is discussed in Section 4.

3.2. The Specification of the Deterministic Component

By Shephard's Lemma, or by Roy's Identity as appropriate, zero demands for the i th commodity correspond to a zero derivative of the cost function, or of the indirect utility function, with respect to the price of the i th commodity. Provided that the indirect utility function is increasing in expenditure, as is assumed in RIU5 (or equivalently that the cost function is increasing in utility, as is assumed in RC5), then the identity

$$U_{p_i}^M(C(u, p), p) = -U_c^M(C(u, p), p)C_{p_i}(u, p) \quad (3.2)$$

ensures that these zero demand conditions are consistent for the two preference specifications. Zero demands over an open set in price-utility or price-expenditure space, which is

presumably what is meant by a ‘corner’, will in addition require a zero second derivative over this open set. Most functional forms that are available for cost or indirect utility functions either give zero at a point, as the functional form crosses the boundary of the regular region and moves into a region of negative demands, or else, if they are regular, by the nature of the conditions that are required for their regularity, only achieve zero derivatives, if at all, at limiting values.

The approach pursued here will be to begin with functional forms which have other attractive properties (such as the empirical attractiveness of for example the Almost Ideal Demand System of Deaton and Muellbauer (1980), the global curvature of models such as proposed by Diewert and Wales (1987), or effective global regularity as proposed by Cooper and McLaren (1992, 1996)), but that potentially violate monotonicity in prices, either over the whole parameter space, or in regions of the parameter space that are excluded by regularity conditions. Beginning with a dual functional form that would allow negative demands, the procedure is then to ‘regularise’ this functional form to achieve a legitimate dual representation of preferences.

To motivate the suggested approach, begin with a specification in terms of the cost function and assume that the specified cost function $\hat{C}$ violates RC4 but satisfies the other regularity conditions, so that the corresponding indirect utility function $\hat{U}^M$ violates RIU4 but satisfies the remaining regularity conditions. Now, construct the **approximate** (the reason for the qualifier approximate will be elaborated below) partial derivatives of the underlying cost function in terms of the partial derivatives of the given function

$$D_{p_i}(u, p) \equiv \max\{0, \hat{C}_{p_i}(u, p)\}. \quad (3.3)$$

Provided the function $\hat{C}$ was twice continuously differentiable, then the functions defined in (3.3) will be differentiable almost everywhere. About each of these points in (u, p) space at which this function is differentiable, there will be a region within which the set of goods $\{1, 2, \dots, N\}$ can be decomposed into the subset P^H for which the corresponding Hicksian

demands are strictly positive, and the subset Z^H (possibly empty) for which the Hicksian demands are identically zero.

While it is relatively easy, by Euler's theorem, to relate a cost function to its underlying partial derivatives, so that (3.3) is an appropriate starting point, what is needed to derive Marshallian demands is an indirect utility function. Since the indirect utility function is homogeneous of degree zero rather than one, the corresponding relation between the function and its derivatives is not available. However, just as C is dual to U^M , then so are the derivatives of C dual to the derivatives of an implied U^M . Thus U^M corresponding to the cost function defined by (3.3) is implicitly defined by the identity

$$C(U^M(c, p), p) \equiv c. \quad (3.4)$$

While the existence of such a U^M is assured, its explicit solution is not necessarily available from the derivatives of the cost function. However, the derivatives of U^M are directly available, and the important point is that it is only these derivatives which are needed explicitly in order to derive Marshallian demands. Since

$$U_{p_i}^M(c, p) = 0 \text{ iff } C_{p_i}(c, p) = 0, \quad (3.5)$$

it is legitimate to define **approximate** derivatives of the indirect utility function by

$$V_{p_i}^M(c, p) = \min\{0, \hat{U}_{p_i}^M(c, p)\} \quad (3.6)$$

and hence, using the required homogeneity of degree zero of U^M , to define

$$V_c^M(c, p) \equiv -\left(\frac{1}{c}\right) \sum_{j=1}^N p_j V_{p_j}^M(c, p). \quad (3.7)$$

The Marshallian demands are now defined solely in terms of these derivatives as

$$Q_i^M(c, p) = -\frac{V_{p_i}^M(c, p)}{V_c^M(c, p)}. \quad (3.8)$$

From a practical point of view, the numerator of (3.8) is relatively simple, but because of its definition via (3.7) the denominator is a potentially complex expression which could cause numerical problems. Fortunately, as shown in the next section, an appropriate specification of the stochastic components of such demand equations can lead to a considerable simplification of the specification of the deterministic component that follows from (3.8). For Engel curve analysis, the above construction is quite attractive and will be illustrated in Sections 5 and 6. But if price responses are of interest then the issue of Slutsky symmetry must be addressed.

The derivatives defined above were referred to as approximate derivatives because of the following reasoning. If a true cost function corresponding to (3.3) were to exist, then it would be defined as the sum of prices times Hicksian demands

$$C(u, p) \equiv \sum_{i=1}^N p_i D_{p_i}(u, p). \quad (3.9)$$

By construction, this C is homogeneous of degree one, non-negative and non-decreasing in prices. However,

$$C_{p_i} \neq D_{p_i}.$$

This is due to the fact that, although

$$\sum_i p_i \hat{C}_{p_i p_j} = 0,$$

the above method of construction means that it is not true that

$$\sum_i p_i D_{p_i p_j} = 0.$$

This is presumably reflecting the fact that the demands so constructed violate integrability in the region of zero demands. While Slutsky Symmetry is satisfied for pairs i, j where $i, j \in P^H$, or $i, j \in Z^H$, it is not satisfied if $i \in P^H$, $j \in Z^H$ or $i \in Z^H$, $j \in P^H$. This is the cost in terms of theoretical specification that has been paid in order to use this method of construction to generate demands that are always non-negative, but potentially zero over a region.

It is an interesting question as to whether there may be an alternative method of ‘regularising’ either the cost function or the indirect utility function which avoids this problem. The approach of Blackorby and Diewert (1979) may be relevant to this possibility. On the other hand, it is not necessarily clear whether Slutsky symmetry is a property which is expected to hold at corners.

To elaborate further, it is necessary to address the question of duality between specifications in terms of first derivatives. While it is true that Shephard’s Lemma and Roy’s Identity only require the derivatives of the corresponding dual functions, regularity conditions are normally specified in terms of the underlying cost or indirect utility functions. Analogous regularity conditions would be required if the derivatives were to be used as the primitive constructs. Corresponding to the regularity conditions RC on the cost function would be the regularity conditions RDC on the partial derivatives (Hicksian demands):

$$\text{RDC1: } D_{p_i} : R \times \Omega^N \rightarrow \Omega^1$$

$$\text{RDC2: } D_{p_i p_j} = D_{p_j p_i}$$

$$\text{RDC3: } D_{p_i} \text{ is homogeneous of degree zero in } p$$

$$\text{RDC4: } \sum_i p_i D_{p_i p_j} = 0$$

$$\text{RDC5: } \sum_i p_i D_{p_i} \text{ is increasing in } u$$

$$\text{RDC6: } D_{p_i p_j} \text{ is a negative semi-definite matrix.}$$

At those points at which Z^H is empty, then clearly RDC1 - RDC6 are satisfied. If Z^H is not empty then using the construction of (3.3) RDC2 will be violated if $i \in P^H$ and $j \in Z^H$ and thus Slutsky symmetry is violated, as is RDC4. RDC6 is not applicable, and it is unclear whether or not RDC5 is satisfied. Nevertheless, for the subset of goods for which all $i, j \in P^H$, Slutsky symmetry and negative definiteness are satisfied and the corresponding

model is potentially quite attractive - particularly for the analysis of Engel responses. The demands generated by this construction will always be non-negative, and potentially can be identically zero (i.e. ‘corner solutions’ are possible on an open region in (u, p) space), and the model would allow the set of zero solutions to alter as utility and/or prices change.

Just as there are a set of derivative conditions on C corresponding to RC, so there are a set of conditions on the derivatives of U^M corresponding to RIU:

$$\text{RDIU1: } V_{p_i}^M, V_c^M : \Omega_+^{N+1} \rightarrow R$$

$$\text{RDIU2: } V_{p_i p_j}^M = V_{p_j p_i}^M, V_{p_i c}^M = V_{c p_i}^M$$

$$\text{RDIU3: } V_{p_i}^M, V_c^M \text{ are homogeneous of degree -1 in } (c, p)$$

$$\text{RDIU4: } V_{p_i}^M \leq 0$$

$$\text{RDIU5: } V_c^M \geq 0$$

$$\text{RDIU6: The determinants of the principal minors of the bordered Hessian of } U^M \text{ formed from the } V_{p_i}^M \text{ are non-positive.}$$

In general, by (3.2) the regularity conditions RDIU are dual to the regularity conditions RDC which have already been discussed. Defining P^M and Z^M as the Marshallian analogues of P^H and Z^H , the problem from the construction (3.6) is again the non-equality of the cross derivatives for pairs of commodities for which one is in the set P^M and one is in the set Z^M . However, RDIU5 is now satisfied by construction, which resolves the question over RDC5.

4. Stochastic Specification

The “traditional” approach to the stochastic specification of a system of budget share equations determined from economic theory is to append an $n(=N-1)$ -variate normal

distribution with a mean of 0 and an $n \times n$ positive definite (constant) covariance matrix Γ to the deterministic component, $W_i(z, \beta)$. Thus,

$$w_i = W_i(z, \beta) + u_i . \quad (4.1)$$

In Fry, Fry and McLaren (1996a) it is shown that, as a statistical model, (4.1) is deficient. Regularity conditions imply that the $W_i(z, \beta)$ are restricted to the closed unit interval $[0,1]$ and use of multivariate normality for u leads to a non-zero probability that the modelled shares will fall outside the $[0,1]$ interval. If the convenience of the normal distribution is to be maintained, an alternative transformation is necessary, and one such transformation is provided by the field of compositional data analysis (CODA). The transformation is defined by

$$y_i = \log(w_i / w_N) \quad (4.2)$$

with Jacobian, $\text{jac}(y|w^{(n)}) = (w_1 \dots w_n)^{-1}$. If the logratio composition y has an n -dimensional normal distribution ($N^n(\mu, \Sigma)$) then the composition w will have an additive logistic normal distribution ($L^n(\mu, \Sigma)$) where μ is the mean of the logratios and Σ denotes the logratio covariance matrix. A direct (statistical) application of the CODA approach would involve the modelling of a logratio transformation of the observed shares, via a multivariate regression model, in terms of the parameters μ and Σ (see Fry, Fry and McLaren (1996a) for details and McLaren, Fry and Fry (1995) for an empirical example).

In principle the CODA model can handle the problem of a ‘pile-up’ of observations at zero. Numerically, there is a problem as the logratio transformation requires us to find the logarithm of zero. These infinite values would cause a problem, but the statistical literature has modified the CODA methodology to deal with the occurrence of zeroes (see Aitchison (1986)). This approach has been investigated in Fry, Fry and McLaren (1996b) where experiments with this “zero replacement” strategy show that it performs very well with economic data.

A second reason why the CODA approach is attractive in the case of corner solutions is that, from the model in Section 3.2, the functional form of the modelled share is

$$W_i(c, p) = \frac{p_i \min\{0, \hat{U}_{p_i}^M(c, p)\}}{\sum_{j=1}^N p_j \min\{0, \hat{U}_{p_j}^M(c, p)\}} \quad (4.3)$$

in which the denominator will potentially be a relatively complex expression, due to its definition as given in (3.7). In the CODA approach, however, it is the *ratios* of these shares that is modelled, and this ratio simply consists of the ratio of two expressions of the relatively simple form of (3.6)

$$\frac{W_i(c, p)}{W_N(c, p)} = \frac{p_i \min\{0, \hat{U}_{p_i}^M(c, p)\}}{p_N \min\{0, \hat{U}_{p_N}^M(c, p)\}}. \quad (4.4)$$

In most empirical work it is quite likely that there is at least one commodity whose budget share is strictly positive for every observation in the sample. In that case (4.4) can be replaced by the relatively simpler expression

$$\frac{W_i(c, p)}{W_N(c, p)} = \frac{p_i \min\{0, \hat{U}_{p_i}^M(c, p)\}}{p_N \hat{U}_{p_N}^M(c, p)}. \quad (4.5)$$

It is this functional form specification that is to be used in the CODA approach to give both a deterministic and stochastic specification that is completely consistent with the observed phenomenon of zeroes in budget share data.

Intuitively, the homogeneity of degree zero of the indirect utility function corresponds to the budget constraint, and hence the adjustment of (4.3) is the adjustment of the budget shares as non-negativity of expenditures is imposed. But this adjustment is such as to maintain relative budget shares, precisely the approach to zero replacement that was motivated by numerical considerations in Fry, Fry and McLaren (1996b).

5. An Illustration Based on the Almost Ideal Demand System

From Cooper and McLaren (1992), the indirect utility function corresponding to the PIGLOG cost function can be written as

$$\hat{U}^M(c, p) = \frac{\log(c / P1)}{P2} \quad (5.1)$$

where $P1$ and $P2$ are two functions of prices. The Almost Ideal Demand System results if $P1$ is specified as the Translog functional form, and $P2$ is specified as a (homogeneous of degree zero) Cobb-Douglas functional form.

It is well-known that the corresponding budget shares are

$$W_i = E1_i + E2_i M \quad (5.2)$$

where

$$M = \log(c/P1) \quad (5.3)$$

and $E k_i$ is the elasticity of Pk with respect to p_i .

Clearly expenditures on commodities can be recovered by multiplying shares by total expenditure c , and quantities are positive if and only if expenditures are positive. Therefore the ratios of shares corresponding to (4.4) are

$$\frac{W_i}{W_N} = \frac{\max\{0, E1_i + E2_i M\}}{\max\{0, E1_N + E2_N M\}} \quad (5.4)$$

In the case where it is known that one commodity, to be designated the N th, is always positive, the equation corresponding to (4.5) is

$$\frac{W_i}{W_N} = \frac{\max\{0, E1_i + E2_i M\}}{E1_N + E2_N M} \quad (5.5)$$

This functional form, together with the stochastic specification in the CODA approach, would potentially give both a deterministic and stochastic specification that is capable of modelling the observed phenomenon of zeroes in budget share data.

6. An Empirical Application

6.1. The Model

While the Almost Ideal Demand System was useful in Section 5 as an illustration of the methodology, it is not ideal for an empirical application. Cooper and McLaren (1992) show that the indirect utility function (5.1) will violate RIU4, RIU5 and RIU6 for c , and hence M , large enough. They demonstrate that a modified version of (5.1), called MPIGLOG, will satisfy all of these regularity conditions for all values of M greater than a lower limit, given sufficient conditions on the parameters of the functions $P1$ and $P2$. Relaxation of some of these sufficient conditions then gives a functional form for the indirect utility function that will be used for the empirical application.

From Cooper and McLaren (1992), the indirect utility function corresponding to the MPIGLOG cost function is

$$\hat{U}^M(c, p) = \log(c/P1) \left[\frac{c^\eta}{P2} \right] \quad (6.1)$$

where $P1$ and $P2$ are again two functions of prices, but now $P2$ is homogeneous of degree η instead of zero. The Modified AIDS (MAIDS) System results if $P1$ is specified as a Translog functional form, and $P2$ is specified as a (homogeneous of degree η) Cobb-Douglas functional form.

The corresponding budget shares are

$$\hat{W}_i = \frac{E1_i + E2_i M}{\sum_j E1_j + \sum_j E2_j M} \quad (6.2)$$

where now

$$M = \log(c/\kappa P1) \quad (6.3)$$

with κ a parameter determining the range of c values which achieve regularity.

The ratios of shares corresponding to (4.4) are again

$$\frac{W_i}{W_N} = \frac{\max\{0, E1_i + E2_i M\}}{\max\{0, E1_N + E2_N M\}} \quad (6.4)$$

where now the only difference from (5.4) is the relaxation of the constraint on $\sum_j E2_j$. Again,

in the case where it is known that one commodity, to be designated the N th, is always positive, the equation corresponding to (4.5) is

$$\frac{W_i}{W_N} = \frac{\max\{0, E1_i + E2_i M\}}{E1_N + E2_N M} . \quad (6.5)$$

This functional form, together with the stochastic specification in the CODA approach, gives both a deterministic and stochastic specification that is capable of modelling the observed phenomenon of the ‘pile up’ of zeroes that is common with microeconomic studies of budget share data.

6.2. The Data

This methodology has been applied to a subset of the data from the 1988/89 Australian Household Expenditure Survey (HES). The HES is a periodic survey by the Australian Bureau of Statistics which has a sampling methodology similar to that of the UK Family Expenditure Survey (FES). The HES collects, among other information, household income and expenditure data on 421 commodities. These 421 ‘fine level’ commodities were aggregated

into 18 commodity groups (further details of this aggregation are available, upon request). The data subset used below avoids problems due to demographic heterogeneity by including only two adult, single or dual income households with no children and positive after tax incomes. The data used is "zero-replaced", using a modified Aitchison zero replacement technique, as described in Fry, Fry & McLaren (1996b).

6.3. The Model to be Estimated

Because the HES contains no price variation, the MAIDS specification is equivalent to a Cobb-Douglas specification for both $P1$ and $P2$. This still allows the desirable Engel response properties of the model, which is the main purpose of the illustration. The logratio form used for estimation is therefore

$$y_i = \log \left[\frac{\text{POS}(\alpha_i + \beta_i \log c) + \tau}{\text{POS}(\alpha_N + \beta_N \log c) + \tau} \right] + v_i \quad (6.6)$$

where $\text{POS}(\cdot)$ is the POS function in TSP which returns the argument if it is positive, and zero otherwise. The symbol τ is the small constant that is used in zero replacement, set as a very small number (5.18175×10^{-6}), representing one cent as a proportion of the maximum total expenditure in the sample.² The use of τ avoids the numerical problem of taking the log of zero and also the possibility of division by zero. Since η and κ are not separately identified, κ has been normalised to the minimum observed value of c in the sample. In the case of an AIDS specification, which corresponds to the constraint $\eta = 0$, κ may be absorbed into the α term and is not separately identifiable even though η is specified. One consequence of the use of the POS function is that parameter constraints alone no longer guarantee that the adding up condition ($\sum W_i = 1$) is satisfied, so it is necessary to achieve this by the alternative normalisation of programming the denominator in (6.2) as the sum of the numerators. No

² Refer to Fry, Fry & McLaren (1996b).

constraints on the alphas or betas are then necessary. The eighteenth category was used for normalisation.³ Aside from its theoretical attractiveness, the use of the POS function avoids a problem that is especially inherent in estimation of the Almost Ideal Demand System using the CODA methodology, which is that some β values are sure to be negative, allowing the possibility that $\alpha_i + \beta_i \ln c$ will be negative over some expenditure range and/or at some point during the iterative procedure, causing numerical errors.⁴ For this reason a simple AIDS model appeared to be impossible to estimate.

6.4. Estimation Results

Estimation of the 17 equation system yielded the results reported in Table 1. The critical expenditure level is the point at which the POS function switches from being active to inactive (or vice versa). The total expenditure range in the data set is from \$79.08 per week to \$1929.85 per week. Five of the categories have their respective POS functions becoming binding within this range - alcohol; clothing and footwear; furniture and other household goods; public transport; and leisure goods and services. For the other categories the POS function is not active over any part of the sample expenditure range. For illustrative purposes, the resulting shape of the budget share curves is illustrated for three of these five categories is shown in Figure 1.

For this model the classification of a good as a luxury or necessity is established by reference to its expenditure elasticity, which is given by⁵

³ The system is invariant to the choice of a normalisation category.

⁴ It is also possible that some alpha values will be negative which can lead to the same problem. However, as

$\sum \alpha_i = 1$ and $\sum \beta_i = 0$ in AIDS, the problem tends to occur more with the beta values.

⁵ Refer to Cooper and McLaren (1992). Note that this formula collapses to the well-known result for AIDS when $\sum \beta_i = 0$.

$$E_i = 1 + \frac{\left(\beta_i - w_i \sum_j \beta_j \right)}{\left(w_i \left(\sum_j \alpha_j + \log c \sum_j \beta_j \right) \right)} . \quad (6.7)$$

This formula is only valid over an expenditure range where no POS functions are active. However, calculation at an expenditure level where one or more POS functions are active is achieved by simply removing the alpha and beta terms of those commodities whose budget share is zero at that expenditure level from the summation terms in the elasticity equation. Of course, the expenditure elasticity for a commodity over the expenditure range where the POS function is active must be zero.

Using the first quartile, average and maximum levels of expenditure, the elasticities are as reported in Table 2. Luxuries are denoted by bold and necessities by italics. Although the elasticities vary over the expenditure level, none alter their status as luxuries or necessities between the first quartile and the maximum expenditure level in the sample. However, as shown in Figure 2, at a low level of total expenditure Privately Purchased Health is a luxury whereas at higher levels of expenditure it becomes a necessity. As people on lower incomes would generally not purchase private health insurance, but those at higher income levels would spend a fairly stable amount this result seems sensible. The other luxury/necessity distinctions seem reasonable. Perhaps the only surprise is that the group Cigarettes & Tobacco is seen as a necessity, however, if the addiction factor is considered⁶ this result is perhaps reasonable. Overall the elasticities are in line with expectations and the results appear sensible.

⁶ As noted in Fry, Fry & McLaren (1996b).

7. Conclusions

The literature contains a number of suggestions for dealing with the fact that many micro level budget studies contain a ‘pile up’ of zero observations. In general, these suggestions seem to be either empirically intractable or theoretically unappealing. This paper has used a dual approach to specify a model which allows for the possibility of a zero demand over a range of expenditures and/or prices. It has also been argued that the compositional data analysis method of estimation provides a complementary stochastic specification, providing a feasible modelling strategy. An application has shown that it is possible to obtain reasonable parameter estimates and associated elasticities using this modelling strategy.

Table 1. Commodity Groupings and Parameter Estimates

Category	Estimated α	Estimated β	Critical c Level (\$)*
1. Bread Cereal & Grain Products	3.785×10^{-3} (9.116×10^{-4})	-3.846×10^{-4} (1.365×10^{-4})	18,784
2. Meat & Fish	1.806×10^{-3} (1.748×10^{-3})	6.861×10^{-5} (2.995×10^{-4})	0
3. Dairy, Eggs & Oil Products	4.678×10^{-3} (9.115×10^{-4})	-5.581×10^{-4} (1.288×10^{-4})	4,366
4. Sugar, Preserves & Confectionary	5.643×10^{-4} (4.104×10^{-4})	1.843×10^{-4} (7.755×10^{-5})	21
5. Fruit & Vegetables	3.098×10^{-3} (1.284×10^{-3})	-1.986×10^{-4} (2.116×10^{-4})	5,951,777
6. Other Food & Non-Alcoholic Drinks	-7.952×10^{-3} (3.266×10^{-3})	2.385×10^{-3} (6.763×10^{-4})	28
7. Alcohol	-2.203×10^{-3} (4.979×10^{-4})	4.398×10^{-4} (9.756×10^{-5})	150
8. Cigarettes & Tobacco	1.325×10^{-5} (2.036×10^{-5})	-1.169×10^{-6} (3.419×10^{-6})	83,591
9. Clothing & Footwear	-3.245×10^{-3} (6.945×10^{-4})	6.272×10^{-4} (1.337×10^{-4})	176
10. Housing Expenditure (current)	-4.128×10^{-3} (7.801×10^{-4})	9.445×10^{-4} (1.784×10^{-4})	79
11. Fuel (not including motor vehicle)	6.470×10^{-3} (1.479×10^{-3})	-7.361×10^{-4} (2.183×10^{-4})	6,565
12. Furniture & Other Household Goods	-6.564×10^{-3} (1.212×10^{-3})	1.387×10^{-3} (2.527×10^{-4})	114
13. Private Transport	-8.685×10^{-3} (4.444×10^{-3})	2.643×10^{-3} (8.682×10^{-4})	27
14. Public Transport	-8.213×10^{-5} (4.329×10^{-5})	1.408×10^{-5} (7.241×10^{-6})	341
15. Leisure Goods & Services	-4.226×10^{-3} (9.328×10^{-4})	8.096×10^{-4} (1.779×10^{-4})	185
16. Other Goods	-1.682×10^{-3} (2.428×10^{-3})	1.138×10^{-3} (4.812×10^{-4})	4
17. Other Services	-1.109×10^{-2} (3.854×10^{-3})	2.964×10^{-3} (7.865×10^{-4})	42
18. Privately Purchased Health	-2.015×10^{-3} (2.282×10^{-3})	8.328×10^{-4} (4.223×10^{-4})	11

System log-likelihood: -34722.2

Note: standard errors in parentheses.

* The point at which the POS function switches from being active to inactive (or vice versa).

Table 2. Expenditure Elasticities at Selected Expenditure Levels

Category [#]	Expenditure Elasticity at:		
	First Quartile <i>c</i>	Average <i>c</i>	Maximum <i>c</i>
<i>Bread Cereal & Grain Products</i>	0.432	0.446	0.909
<i>Meat & Fish</i>	0.699	0.737	0.976
<i>Dairy, Eggs & Oil Products</i>	0.308	0.287	0.800
Sugar, Preserves & Confectionary	1.058	1.044	1.003
<i>Fruit & Vegetables</i>	0.567	0.602	0.954
Other Food & Non-Alcoholic Drinks	1.110	1.082	1.005
Alcohol	2.399	1.708	1.027
<i>Cigarettes & Tobacco</i>	0.569	0.604	0.954
Clothing & Footwear	2.987	1.909	1.031
Housing Expenditure (current)	1.481	1.320	1.016
<i>Fuel (not including motor vehicle)</i>	0.354	0.348	0.856
Furniture & Other Household Goods	1.822	1.496	1.022
Private Transport	1.101	1.075	1.005
Public Transport	0.000*	2.583	1.039
Leisure Goods & Services	3.285	1.985	1.032
<i>Other Goods</i>	0.910	0.928	0.995
Other Services	1.207	1.150	1.009
<i>Privately Purchased Health</i>	0.982	0.986	0.999

* At this level of expenditure the POS function is active for this category.

Luxuries are denoted by bold and necessities by italics.

Figure 1. Illustrative Budget Shares

Figure 2. Budget Share for Health

References

- Aitchison J** (1986) *The Statistical Analysis of Compositional Data*. Chapman and Hall, New York.
- Blackorby C, Diewert WE** (1979) Expenditure Functions, Local Duality and Second Order Approximations. *Econometrica* 47/3: 579-601.
- Cooper RJ, McLaren KR** (1992) An Empirically Oriented Demand System with Improved Regularity Properties. *Canadian Journal of Economics* 25: 652-668.
- Cooper RJ, McLaren KR** (1996) A System of Demand Equations Satisfying Effectively Global Regularity Conditions. *Review of Economics and Statistics* 78/2: 359-364.
- Deaton A, Muellbauer J** (1980) An Almost Ideal Demand System. *American Economic Review* 70/3: 312-326.
- Diewert WE, Wales TJ** (1987) Flexible Functional Forms and Global Curvature Conditions. *Econometrica* 55/1: 43-68.
- Fry JM, Fry TRL, McLaren KR** (1996a) The Stochastic Specification of Demand Share Equations: Restricting Budget Shares to the Unit Simplex. *Journal of Econometrics* 73/2: 377-385. (An extended version available as Working Paper 3/93, Department of Econometrics, Monash University.)
- Fry JM, Fry TRL, McLaren KR** (1996b) Compositional Data Analysis and Zeroes in Micro Data. In M McAleer, P Miller and C Ong (eds) *Econometric Society Australasian Meeting 1996 Proceedings Volume 4: Microeconometrics*. The University of Western Australia.
- Heien D, Wessells CR** (1990) Demand Systems Estimation with Microdata: A Censored Regression Approach. *Journal of Business and Economic Statistics* 8/3: 365-371.
- Lee L-F, Pitt MM** (1986) Microeconomic Demand Systems with Binding Non-negativity Constraints. *Econometrica* 54/5: 1237-1242.
- McLaren KR, Fry JM, Fry TRL** (1995) A Simple Nested Test of the Almost Ideal Demand System. *Empirical Economics* 20/1: 149-161.
- Pudney S** (1989) *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes*. Oxford, Blackwell.

Wales TJ, Woodland AD (1983) Estimation of Consumer Demand Systems with Binding Non-negativity Constraints. *Journal of Econometrics* 21: 263-285.