Estimating Household Equivalence Scales for South Africa by Bayesian Fuzzy Regression

Hui Feng

(Comments Welcome)

Department of Economics, University of Victoria, B.C., Canada

September 2005

Abstract

In this paper, we use Bayesian fuzzy regression analysis to estimate equivalence scales for South Africa. The Engel curve approach to equivalence scales will be used in this study. The model we use is a single equation Working-Leser model. Bayesian posterior odds analysis is employed to select the best fuzzy model from a number of models based on one to four fuzzy clusters. The equivalence scales from this fuzzy Bayesian regression analysis are consistent with those from previous studies using the semi-parametric index model.

Key Words: Equivalence Scales, Engel Curve, Working-Leser, Fuzzy Cluster, Model Selection, Bayesian Posterior Odds, Fuzzy Bayesian Regression

JEL Classifications:

Author Contact:
Hui Feng
Department of Economics
University of Victoria,
P.O. Box 1700 STN CSC
Victoria, B.C., Canada, V8W 2Y2

E-mail: hui@uvic.ca; FAX (250) 721 6214; Phone: (250) 472 4418

1. Introduction

How much does a household of two adults and two children need to have in order to be equally as well off as a household comprising a single adult? How should tax policy change with changes in the household's demographic composition? How should one family be compensated for losing one family member, or being old? How should child or spousal support be determined in the case of divorce? All of these issues deal with the question of equivalence scales. An equivalence scale is the amount by which a household's consumption expenditure would need to be multiplied to make that household as well off as some reference household (Lewbel, 1997). Usually equivalence scales can be estimated using Engel curves. An Engel curve measure the relationship between the expenditure on a particular good and the total expenditure of the household.

In this paper we estimate equivalence scales for South Africa in 1993 using a Bayesian regression approach. Yatchew, *et al.* (2001) and Yatchew (2003) have studied the same topic using Engel curves and semiparametric specifications. The primary intention of this study is to test the performance of the Bayesian fuzzy regression technique. As a result, in this paper, we will use the same survey data as Yatchew, and employ a similar model for the equivalence scales, but use the Bayesian fuzzy regression technique. By comparing the results with those of Yatchew and his coauthors, we hope to prove the high performance of this Bayesian fuzzy regression methodology.

This paper is organized in the following way. The second section provides an introduction to various methods used for the equivalence scales study; and the third section provides the reader with some basic information about the survey and the data. The fourth section is an introduction for the Bayesian fuzzy multiple regression; the fifth section provides the results of this study, and compares them with those of Yatchew (2003). The last section provides our conclusions, along with some future extensions of this research.

2. Formulation of the Equivalence Scales

Basically, there are two ways to estimate equivalence scales using Engel curves: single equation estimation and complete demand systems estimation. The first type of approach concentrates on one single Engel curve for good *i,* which is isolated from the rest of the demand system. Prais and Houthakker (1955) found that the semi-log Engel curve (Eq. 2.01) is best suited for necessity goods, while the double-log Engel curve (Eq. 2.02) is preferable for luxury goods.

$$e_i = \alpha_i + \beta_i \log(E) + \varepsilon_i \tag{2.01}$$

$$\log(e_i) = \alpha_i + \beta_i \log(E) + \varepsilon_i \tag{2.02}$$

where e_i is the individual expenditure on good i and E is the total expenditure for the household. Engel curves are Marshallian demand curves, holding the price vector constant. One drawback of these two models is the utility function they represent. There are no sensible utility functions behind these two models. The second type of approach is built on a reasonable form of the utility function and concentrates on systems of Engel curves. Examples include the studies of Leser (1963), Bewley (1982), Aasness and Rodseth (1983), and Giles and Hampton (1985). There are different types of models for this system approach and what we consider here are Working-Leser (1963) and addilog models.

The addilog model was proposed by Bewley (1982) and is derived from the work by Leser (1941) and Houthakker (1960):

$$\ln(w_{ij} / \widetilde{w}_j) = \alpha_i + \beta_i \log(E_j) + \varepsilon_{ij} \quad , \tag{2.03}$$

where i denotes the "*i-th*" good, and j denotes the "*j-th*" household. w_{ij} measures good i's share of total expenditure for household j, and $\widetilde{w}_j$ is the geometric mean of the shares for household *j*. E_j is the total expenditure for household *j*. (i=1, 2, 3, …, N and j=1, 2, 3, …, M.)

To satisfy Engel aggregation, which states that the sum of the expenditures for the individual goods has to equal the total expenditure, so that the sum of the expenditure shares equals unity, the following adding-up restrictions are needed:

$$\sum_{i=1}^{N} \alpha_i = \sum_{i=1}^{N} \beta_i = \sum_{i=1} \varepsilon_{ij} = 0 \qquad ; j = 1, \ldots, M \quad . \tag{2.04}$$

Deaton and Muellbauer (1980) introduced the almost ideal demand system and showed that this system collapses to the Working-Leser model once we fix the price level. In this research we are going to use a single equation Working-Leser model to model the Engel curves, and hence to estimate the equivalence scales for South Africa in 1993.

The Working-Leser model (Working,1943) is:

$$w_{ij} = \alpha_i + \beta_i \log(E_j) + \varepsilon_{ij} \quad , \tag{2.05}$$

where the notation is as for equation (2.03).

The adding-up restrictions for the Working-Leser model are:

$$\sum_{i=1}^{N} \alpha_i = 1; \sum_{i=1}^{N} \beta_i = \sum_{i} \varepsilon_{ij} = 0 \qquad ; j = 1, \ldots, M \quad . \tag{2.06}$$

The Engel curve we use is the food Engel curve. The assumption we have to make in order to estimate the equivalence scales using the food Engel curve is that two households are said to be equally well off if their food shares are the same. Under this assumption, using the Engel curve to estimate the equivalence scales is equivalent to measuring the horizontal shift needed to place the food Engel curve for a particular household on top of that for the reference household. Allowing the model to capture the different demographic features of a household, the Working-Leser model has been extended into the following form:

$$y \;= \beta_1 + \beta_2(\log(x\;) - \beta_3 \log(A + \beta_4 K)) + \varepsilon \qquad , \qquad (2.07)$$

where y and x are the food share and the total expenditure for the household, and A, K are the number of adults and children in the household. β_3 measures the effect of scale economies in the household and β_4 measures the effect on the equivalence scale of children relative to adults. Both β_3 and β_4 are restricted to be lie between zero and one. Citro and Micheal (1995, p.176) suggested that both parameters have values around 0.7.

We define the equivalence scale Δ as $\Delta = \frac{x_c}{x_r}$, under the assumption that the two household are equally well off if their food share are the same, and so

$$\delta = \log(\Delta) = \log(\frac{x_c}{x_r}) = \log(x_c) - \log(x_r) \qquad . \qquad (2.08)$$

Let subscript "*r*" denote the reference group, so we then have:

$$\hat{y}_r = \beta_1 + \beta_2(\log(x_r) - \beta_3 \log(A_r + \beta_4 K_r)) \qquad , \qquad (2.09)$$

and for a particular household, we have

$$\hat{y} \;= \beta_1 + \beta_2(\log(x) - \beta_3 \log(A + \beta_4 K\;)) \qquad . \qquad (2.10)$$

If the two household are equally well off, we have $\hat{y}_r = \hat{y}$.

Then,

$$\beta_1 + \beta_2(\ln(x) - \beta_3 \log(A + \beta_4 K)) = \beta_1 + \beta_2(\ln(x_r) - \beta_3 \log(A_r + \beta_4 K_r)) \qquad (2.11)$$

or ,

$$\log(x) - \log(x_r) = \beta_3(\log(A + \beta_4 K) - \log(A_r + \beta_4 K_r)) \qquad (2.12)$$

or, $$\delta = \beta_3 \log(\frac{A + \beta_4 K}{A_r + \beta_4 K_r}) \tag{2.13}$$

so that $$\Delta = (\frac{A + \beta_4 K}{A_r + \beta_4 K_r})^{\beta_3} \quad . \tag{2.14}$$

If the reference household is the single adult only household, the equivalence scale Δ can be written as:

$$\Delta = (A + \beta_4 K)^{\beta_3} \quad . \tag{2.15}$$

Given the restrictions for the two scale parameters, we can rewrite the model as:

$$y = \beta_1 + \beta_2 \times (\log x - \frac{{\beta_3}^2}{({\beta_3}^2 + 1)} \log(A + \frac{{\beta_4}^2}{({\beta_4}^2 + 1)} K)) + \varepsilon \quad . \tag{2.16}$$

Actually, later the results of this study show that β_3 will fall between zero and one every time even without the restriction, but this is not the case of β_4. Our final model will be simplified to:

$$y = \beta_1 + \beta_2 \times (\log x - \beta_3 \log(A + \frac{{\beta_4}^2}{({\beta_4}^2 + 1)} K)) + \varepsilon \quad . \tag{2.17}$$

We are fully aware the fact Engel curves do not provide a perfect approach for the study of equivalence scales, and the main criticism is that households with children are more likely to have a larger food share than households without children, given the same utility level. Detailed critics can be found in Deaton (1997). Yatchew (2003) noticed this and by putting a set of dummy variables for all the 36 households, he avoided this criticism of the Engel curve approach to equivalence scales, and noted "rather than searching for a horizontal shift that superimposes on Engle cure on another, a combined horizontal and vertical shift that achieves the superimposition is sought instead". (Yatchew, 2003, p.148.)

Given the feature of the fuzzy clustering that we employ in this study, we believe our model without the dummy variables for different households is free from this criticism. An important advantage of using the fuzzy clustering is that we realize the differences in the data and cluster them into different groups accordingly. For this multiple regression model, later in the third section, we will discuss how we cluster the data into different groups according to the log monthly total expenditure, the number of adults in the household, and also the number of children in the household. Clustering the data with respect to the number of adults and children in the household is equivalent to clustering the data according to its household type. As a result, the fuzzy clustering technique gives us an advantage when analyzing the equivalence scales household type by type. When we try to measure the equivalence scales, we not only shift the Engel curves horizontally, but also vertically, (implicitly) as well.

3. Data Analysis

The data set we use for this study is from the South Africa Integrated Household Survey in 1993 (nine months before that country's first democratic election), which was funded by the governments of Denmark, the Netherlands and Norway working through the International Bank for Reconstruction and Development (World Bank).

A comprehensive household questionnaire was used in this survey, and the topics include demography, household services, household expenditures, educational status and expenditure, remittances and martial maintenance, land access and use, employment and income, health status and expenditure and anthropometrics, (Children under the age of six were weighed and their heights measured).

The "household" definition was individuals who:

1. Live under this "roof" or within the same compound/household/stand at least 15 days out of the past year, and
2. When they are together they share food from a common source (i.e., they cook and eat together); and
3. Contribute to or share in, a common resource pool.

A two-stage self-weighting design was used for the sampling, in which the first stage units were Census Enumerator Sub-districts (ESDs, or their equivalent) and the second stage was households. The advantage of using such a design is that it provides a representative sample that need not be based on an accurate census population—it can over-sample the poor households. A complete introduction of the survey design can be found at http://www.worldbank.org/lsms/country/za94/docs/za94ovr.txt.

The data we use in this study is the same data set as that used in Yatchew (2003), and Yatchew, Sun, and Deri (2003). A data summary is reported in Table 1. There are over 8794 valid households in this survey. Taking away 5% from both tails and restricting the sample only to those households with no more than 6 adults and no more than 5 children, the sample size reduces to 7358. Four series are going to be used for our Bayesian fuzzy regression analysis. They are log total monthly expenditure for the household (x), monthly food share for the household (Y), number of adults in the household (A) no younger than 18, and number of children less than 18 in the household (K).

Table 1 Data Summary

Name	N	Mean	St.Dev.	Variance	Minimum	Maximum
K	7358	1.4691	1.4786	2.1863	0	5
A	7358	2.6775	1.3833	1.9135	1	6
Y	7358	0.4700	0.1933	0.0374	0.03	0.96
Ln (x)	7358	6.9537	0.7495	0.5618	5.59	8.62

As we can see, in 1993 the poorest household in South Africa spent 96% of their income buying food, while the richest household only spent 3% of their income for food monthly. An average of 47% income was spent for food monthly. Table 2 gives more detailed information about the household types. About 59% of the households have no more than three adults and no more than two children.

Table 2. Distributions of the Household Types

A/K	0	1	2	3	4	5
1	1109	138	126	85	61	14
2	890	526	524	309	144	65
3	373	314	322	233	138	67
4	222	227	230	160	104	66
5	105	117	144	116	66	43
6	50	44	71	78	45	32

4. Bayesian Fuzzy Regression

4.1 Fuzzy Clustering

A classical set is known as a "crisp" set, which has a clear crisp boundary. For example, the set E could be defined as any integer that is greater than 10. The membership of a "crisp" set requires the individual element either be a member or not be a member—the membership equals either zero or one.

$$Mu(x) = \{ \begin{matrix} 1 \\ 0 \end{matrix} \quad \text{if} \quad \begin{matrix} x \in E \\ x \notin E \end{matrix} \quad . \tag{4.01}$$

A zero indicates that the element *"x"* is not a member of the set *E*, and unity indicates the element is a member of the set *E*.

On the other hand, the idea of the fuzzy set is just as the name implies: "without a crisp boundary". The difference between a fuzzy set and a classical set lies in the nature of the membership for each element. Zadeh (1965) redefined the meaning of the membership for fuzzy sets to be a continuous number between zero and one. This means an element could be associated with one or more clusters. Further, all these memberships added together should equal unity. Generally, this association involves different degrees of membership with each of the fuzzy sets. Just as this makes the boundaries of the sets fuzzy, it makes the location of the centroid of the set fuzzy as well.

The "fuzzy c-means" (FCM) algorithm, which was developed and improved by Dunn (1974, 1977), and Bezdek (1973, 1981), is frequently used in pattern recognition. The objective is to partition the data into fuzzy clusters, to locate these clusters, and to quantify the degree of membership of every data-point with every cluster. It is based on minimization of the following objective function:

$$J(U,v)=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^{m}(d_{ik})^{2} \quad , \qquad (4.02)$$

where u_{ik} is the "degree of membership" of data-point "k" in cluster "i", and d_{ik} is the distance between data x_k and the *i-th* cluster center v_i. "c" is the number of clusters presumed, and m is any real number greater than 1 which measures the degree of the fuzziness. If m equals 1, we say the membership is crisp. Usually, 2 is a common choice for m. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ik} and the cluster centers v_i by:

$$u_{ik}=1/\{\sum_{j=1}^{n}[(d_{ik})^{2}/(d_{jk})^{2}]^{1/(m-1)}\} \qquad (4.03)$$

and,

$$v_{i}=[\sum_{k=1}^{n}(u_{ik})^{m}x_{k}]/[\sum_{k=1}^{n}(u_{ik})^{m}] \; ; i = 1, 2, \ldots, c. \qquad (4.04)$$

Then, separate regressions will be fitted over each of the clusters. The fuzzy regression is obtained by combining the results from each cluster using the membership functions as the weights. See Giles and Draeseke (2003) for full details. We have written codes for commands in the SHAZAM (2004) econometric package for this analysis.

Over the years, there have been numerous developments in fuzzy sets studies following the research by Zadeh (1965, 1987). Quite a lot applications focus on the areas of computer science, systems analysis, electrical and electronic engineering, and psychology. In recent years, various applications of fuzzy set in the area of econometrics have been provided by Giles (2005), Draeseke and Giles (2002), Giles and Draeseke (2003), Giles and Mosk (2003), Chen and Giles (2004), Giles (2005), Giles and Feng (2005), and Giles and Stroomer (2004, 2005).

As we have three explanatory variables in this study, we assume that the sample data for each variable has the same number of clusters, and we limit the value for c to be less than 5[1]. Hence, we set up the model space for this research includes four models: *"c=1"*— there is one cluster for each of the variable; *"c=2"*—there are two clusters for each of the variables; *"c=3"*— there are three clusters for each of the variable; and there are four clusters for each of the variable—*"c=4"*. As a result, for the first case, the overall sample would be the only cluster; for the second case, we would expect to see 8 clusters given each variable has two clusters; for the third case, we would expect to see 27 clusters; and we would expect to see 64 clusters in the fourth case. Of course, some of the clusters will be empty set or not have enough observations to provide positive degrees of freedom. Table 3 is a simply an example to illustrate this idea.

Table 3. Hypothetical Clustering: n = 20, k = 3 & c = 2

	Log (x)	A	K
1st Cluster	1,3,4,6,9,11,14,15,19,20	2,4,6,8,13,17,18,20	2,8,10,14,16,17,18,19
2nd Cluster	2,5,7,8,10,12,13,16,17,18	1,3,5,7,9,10,11,12,14,15,16,19	1,3,4,5,6,7,9,11,12,13,15,20
Group 111	Empty	Group 211	2,8,17,18
Group 112	4,6,20	Group 212	13
Group 121	14,19	Group 221	10,16
Group 122	1,3,9,11,15	Group 222	5,7,12

The numbers in the table are the orders of the observations. The first two rows of the table show how the data has been partitioned into two clusters using log (x), A and K respectively. The last four rows of the table show us the result for each of the eight clusters we expect to get. Group *ijk* means the observations that fall into the *i*th cluster according to the variable log(x), *j*th cluster according to the variable A, and *k*th cluster according to the variable K. As we can see one cluster is an empty set, and five clusters have zero or negative degrees of freedom. Two clusters are left with positive degrees of freedom. Thus, we only need fit the model for those two clusters and the fuzzy regression would be the weighted average of these two clusters' regressions. More generally, if the sample size is large, we will only have a few empty sets. In this study, with 7358

---

[1] For all the existing fuzzy clustering studies, the choice of c seldom goes above 4.

observations, for the $c = 2$ case, we have eight clusters with positive degrees of freedom, and for the $c = 3$ case, we have fifteen effective clusters and for the $c = 4$ case, we have sixty-three effective clusters, once empty and infeasible clusters are discarded.

Assuming that the clustering for the three variables is undertaken independently, the final membership function can be calculated as the product of the three membership functions obtained from the fuzzy clustering of the three variables. Because in fuzzy set theory, the union operator is max and the intersection operator is min. a second form of the membership function for each observation can be calculated by using the minimum membership function among the three membership functions. Later in Section 5, we report the Engel curves for each of the households using the product form of these weights, we also will provide equivalence scales calculated from both forms for the weights.

4.2 Bayesian Posterior Odds Analysis

We have four candidate models here, and which one to use is a problem that needs to be solved first. This is equivalent to asking how many clusters we should include in this analysis. There are essentially two ways to due with this model selection problem—the classical approach and the Bayesian approach. For this study, we will use the Bayesian Posterior Odds analysis to select the value for "*c*". This study involves the prior information for the model, the prior information for each parameter and also the choice of a loss function.

4.2.1 Prior Information

Let M be the model space, and M_i denote each individual model. There are m available models altogether. Usually, we assume that m is a finite number.

$$M=\{M_i\}\ i=1,2,3\ldots m. \tag{4.05}$$

Use $p(M_i)$ to denote the prior probability for the *i-th* model where each prior probability lies between 0 and 1. In addition,

$$\sum_{i=1}^{m} p(M_i) = 1 \tag{4.06}$$

if the model space, *M*, is exhaustive.

In this study, we have four candidate models and the prior probability of each model being a true model *p*(*Mi*) is assigned to be 0.25, which is to say that we do not have any bias against or in favor of a particular model being a true model, before we see the data.

As for the prior information for the parameters, we use the usual natural-conjugate (conditional Normal, marginal inverted-gamma) prior:

$$p(\beta_i, \sigma_i) = p(\beta_i \mid \sigma_i) * p(\sigma_i) \tag{4.07}$$

$$p(\beta_i \mid \sigma_i) = \frac{|C_i|^{1/2}}{(2\pi)^{k/2} \sigma_i^{\,k}} \exp[-\frac{1}{2\sigma_i^{\,2}} (\beta_i - \bar{\beta}_i)' C_i (\beta_i - \bar{\beta}_i)] \tag{4.08}$$

$$p(\sigma_i) = \frac{K_i}{\sigma_i^{q_i+1_i}} \exp(-\frac{q_i \bar{s}_i^2}{2\sigma_i^2}) \tag{4.09}$$

where i denotes the ith regressor, and the normalizing constant $K_i = 2(q_i \bar{s}_i^2 / 2)^{q_i/2} / \Gamma(q_i / 2)$. We assume the proper prior pdf for β_i given σ_t with prior mean vector $\bar{\beta}_i$ and covariance matrix $\sigma_i^2 C_i^{-1}$; C_i a $k \times k$ matrix to be assigned by the investigator, is assumed to be a positive definite symmetric matrix. The prior information for σ_i is in the inverted gamma form with parameters q_i and $\bar{s}_i^2$ to be assigned values by the investigator. For this to be a proper prior, we need $0 < q_i$, $\bar{s}_i^2 < \infty$, i=1, 2. To allow our prior information to be relatively vague we assign q_i to be slightly greater than 2.

4.2.2 Loss Function

In the context of the model selection problem, we want to come to the conclusion of "rejecting" or "accepting" one model compared with the other model. This is a *two-action* problem (Zellner, 1971, p.291).

There are two possible states of the world—H_0 true or H_1 true.

If we accept the true model or we reject the false model, we will incur zero loss. However, our loss would be $L(H_1, \hat{H}_0)$ if we accept the false null hypothesis, and our loss would be $L(H_0, \hat{H}_1)$ once we reject the true null hypothesis.

Which model will be accepted dependd on which model will minimize expected loss:

$$\begin{aligned} &\text{If } E(L \mid \hat{H}_0) < E(L \mid \hat{H}_1)\text{, Accept } H_0 \\ &\text{If } E(L \mid \hat{H}_1) < E(L \mid \hat{H}_0)\text{, Accept } H_1, \end{aligned} \tag{4.10}$$

and,

$$\begin{aligned} E(L \mid \hat{H}_0) &= p(H_0 \mid y)L(H_0, \hat{H}_0) + p(H_1 \mid y)L(H_1, \hat{H}_0) = p(H_1 \mid y)L(H_1, \hat{H}_0) \\ E(L \mid \hat{H}_1) &= p(H_0 \mid y)L(H_0, \hat{H}_1) + p(H_1 \mid y)L(H_1, \hat{H}_1) = p(H_0 \mid y)L(H_0, \hat{H}_1) \end{aligned} \tag{4.11}$$

Under a symmetric loss function[2],

$$\begin{aligned} L(H_0, \hat{H}_1) &= L(H_1, \hat{H}_0) \\ L(H_0, \hat{H}_0) &= L(H_1, \hat{H}_1) = 0. \end{aligned} \tag{4.12}$$

Then, the null will be accepted only when $E(L \mid \hat{H}_0) < E(L \mid \hat{H}_1)$, that is:

$$p(H_0 \mid y) > p(H_1 \mid y) \text{ or } \frac{p(H_0 \mid y)}{p(H_1 \mid y)} > 1 \quad . \tag{4.13}$$

So, in the context f our study this means that:

$$M_1 \succ M_2 \text{ if } \frac{p(M_1 \mid y)}{p(M_2 \mid y)} > 1$$

---

[2] We could assume that the loss is not symmetric in our research; however, the idea will be the same, except more complicated.

$$M_1 \succ M_3 \text{ if } \frac{p(M_1 \mid y)}{p(M_3 \mid y)} > 1$$

$$M_1 \succ M_4 \text{ if } \frac{p(M_1 \mid y)}{p(M_4 \mid y)} > 1$$

$$M_2 \succ M_3 \text{ if } \frac{p(M_2 \mid y)}{p(M_3 \mid y)} > 1$$

$$\ldots \ldots \textit{etc.} \tag{4.14}$$

As a result, we do not need to calculate the posterior probability for each model to determine out the selected model. However, according to the value of the BPO, we will know which model will be preferred between any two models, and by using all of the BPOs, we can rank all of the candidate models. Working with the BPO's, rather than the individual posterior probabilities, arrives the need to fully specify the model space.

4.2.3 Bayesian Posterior Odds

Using Bayes' Rule, the posterior probability of model M_i is

$$\begin{aligned} p(M_i \mid y) &= \frac{p(M_i) * p(y \mid M_i)}{p(y)} \\ &\propto p(M_i) * p(y \mid M_i) \end{aligned} \tag{4.15}$$

where

$$p(y) = \sum_{j=1}^{m} p(M_j) * p(y \mid M_j) \quad . \tag{4.16}$$

We could choose the model according to the value of the posterior probabilities for each candidate model, but the posterior probability of M_i will be incorrect if the model space $\boldsymbol{M}$ is not been specified completely. However, even if $\boldsymbol{M}$ is incompletely specified, we can still work out the Bayesian Posterior Odds (BPO) correctly: Let $[P(M_i)/P(M_j)]$ be the prior odds. Then

$$BPO_{ij} = \frac{p(M_i \mid y)}{p(M_j \mid y)} = \{\frac{p(M_i) * p(y \mid M_i) / p(y)}{p(M_j) * p(y \mid M_j) / p(y)}\} = [\frac{p(M_i)}{p(M_j)}] * [\frac{p(y \mid M_i)}{p(y \mid M_j)}] \ . \tag{4.17}$$

The BPO provide a basis for us to compare any two models. Under a symmetric loss function: if BPO is 1, this means the two models are equal; if BPO is less than 1, this means the *j-th* model is preferred to the *i-th* model; and if BPO is greater than 1, this means the *i-th* model is preferred to the *j-th* model. In order to get the BPO, we need to know the likelihood function for each of the models. Employing the prior information and loss function specified above, we can get the BPO between every pair of the fuzzy models and to rank all the possible models will not be a big problem.

$$\begin{aligned}
p(y \mid M_1) &= \iint p(y \mid \beta_1, \sigma_1, M_1) p(\beta_1 \mid \sigma_1) p(\sigma_1) d\beta_1 d\sigma \\
&= \iint \frac{1}{(2\pi)^{n/2}} \frac{1}{{\sigma_1}^{n}} \exp\{\frac{1}{2\sigma_1^2}[vs_1^2 + (\beta_1 - \hat{\beta}_1)' X_1{}' X_1 (\beta_1 - \hat{\beta}_1)]\} \\
&* \frac{|C_1|^{1/2}}{(2\pi)^{k/2} {\sigma_1}^{k}} \exp[-\frac{1}{2{\sigma_1}^{2}} (\beta_1 - \overline{\beta}_1)' C_1 (\beta_1 - \overline{\beta}_1)] * \frac{K_1}{\sigma_1^{q_i + 1_i}} \exp(-\frac{q_1 \overline{s}_1^2}{2\sigma_1^2}) d\beta_1 d\sigma_1 \\
&= \frac{1}{2} K_1 \left\| \frac{|C_1|}{|A_1|} \right\|^{1/2} 2^{\frac{n_1 + q_1}{2}} \Gamma(\frac{n_1 + q_1}{2}) (q_1 \overline{s}_1^2 + v_1 s_1^2 + Q_{1a} + Q_{1b})^{-\frac{n_1 + q_1}{2}}
\end{aligned} \tag{4.18}$$

where $A_1 = C_1 + X_1{}' X_1$

$$\widetilde{\beta}_1 = A_1^{-1} (C_1 \overline{\beta}_1 + X_1{}' X_1 \hat{\beta})$$

$$Q_{1a} = (\overline{\beta}_1 - \widetilde{\beta}_1)' C_1 (\overline{\beta}_1 - \widetilde{\beta}_1)$$

$$Q_{1b} = (\hat{\beta}_1 - \widetilde{\beta}_1) X_1{}' X_1 (\hat{\beta}_1 - \widetilde{\beta}_1)$$

$$K_1 = \frac{2(q_1 \overline{s}_1^2 / 2)^{q_1/2}}{\Gamma(q_1/2)} \ .$$

Above is the likelihood function for model 1 where we have only one cluster, and similar operations would be used to calculate the other three models' likelihood function. Treating the clusters independently when c=2 we get:

$$p(y\,|\,M_2)=\frac{1}{2}K_{21}\left\|\frac{|C_{21}|}{|A_{21}|}\right\|^{1/2}2^{\frac{n_{21}+q_{21}}{2}}\Gamma(\frac{n_{21}+q_{21}}{2})(q_{21}\bar{s}_{21}^2+v_{21}s_{21}^2+Q_{21a}+Q_{21b})^{-\frac{n_{21}+q_{21}}{2}}\,*$$

$$\frac{1}{2}K_{22}\left\|\frac{|C_{22}|}{|A_{22}|}\right\|^{1/2}2^{\frac{n_{22}+q_{22}}{2}}\Gamma(\frac{n_{22}+q_{22}}{2})(q_{22}\bar{s}_{22}^2+v_{22}s_{22}^2+Q_{22a}+Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}$$

$$=\frac{1}{4}K_{21}K_{22}\left[\frac{|C_{21}\|C_{22}|}{|A_{21}\|A_{22}|}\right]^{1/2}\frac{(q_{21}\bar{s}_{21}^2+v_{21}s_{21}^2+Q_{21a}+Q_{21b})^{-\frac{n_{21}+q_{21}}{2}}(q_{22}\bar{s}_{22}^2+v_{22}s_{22}^2+Q_{22a}+Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}}{2^{\frac{n_{21}+q_{21}}{2}}\Gamma(\frac{n_{21}+q_{21}}{2})2^{\frac{n_{22}+q_{22}}{2}}\Gamma(\frac{n_{22}+q_{22}}{2})}$$

(4.19)

The posterior odds between model 1 and model 2 is:

$$BPO_{12}=[\frac{p(M_i)}{p(M_j)}]*[\frac{p(y\,|\,M_i)}{p(y\,|\,M_j)}]=2\frac{K_1}{K_{21}K_{22}}\left[\frac{|C_1\|A_{21}\|A_{22}|}{|A_1\|C_{21}\|C_{22}|}\right]^{1/2}$$

$$\times\frac{2^{\frac{n_1+q_1}{2}}\Gamma(\frac{n_1+q_1}{2})}{2^{\frac{n_{21}+q_{21}}{2}}\Gamma(\frac{n_{21}+q_{21}}{2})2^{\frac{n_{22}+q_{22}}{2}}\Gamma(\frac{n_{22}+q_{22}}{2})}\,* \tag{4.20}$$

$$\times\frac{(q_{21}\bar{s}_{21}^2+v_{21}s_{21}^2+Q_{21a}+Q_{21b})^{-\frac{n_{21}+q_{21}}{2}}(q_{22}\bar{s}_{22}^2+v_{22}s_{22}^2+Q_{22a}+Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}}{(q_1\bar{s}_1^2+v_1s_1^2+Q_{1a}+Q_{1b})^{-\frac{n_1+q_1}{2}}}.$$

Similarly, the Bayesian Posterior Odds between model 1 and model 3 is

$$BPO_{13}=[\frac{p(M_1)}{p(M_3)}]*[\frac{p(y\,|\,M_1)}{p(y\,|\,M_3)}]=2^{3-1}\frac{K_1}{K_{31}K_{32}K_{33}}\left[\frac{|C_1\|A_{31}\|A_{32}\|A_{33}|}{|A_1\|C_{31}\|C_{32}\|C_{33}|}\right]^{1/2}$$

$$\times\frac{2^{\frac{n_1+q_1}{2}}\Gamma(\frac{n_1+q_1}{2})}{2^{\frac{n_{31}+q_{31}}{2}}\Gamma(\frac{n_{31}+q_{31}}{2})2^{\frac{n_{32}+q_{32}}{2}}\Gamma(\frac{n_{32}+q_{32}}{2})2^{\frac{n_{33}+q_{33}}{2}}\Gamma(\frac{n_{33}+q_{33}}{2})} \tag{4.21}$$

$$\times\frac{(q_{31}\bar{s}_{31}^2+v_{31}s_{31}^2+Q_{31a}+Q_{31b})^{-\frac{n_{31}+q_{31}}{2}}(q_{22}\bar{s}_{32}^2+v_{22}s_{32}^2+Q_{32a}+Q_{32b})^{-\frac{n_{32}+q_{32}}{2}}\;)^{-\frac{n_{33}+q_{33}}{2}}}{(q_1\bar{s}_1^2+v_1s_1^2+Q_{1a}+Q_{1b})^{-\frac{n_1+q_1}{2}}}$$

$$\times(q_{33}\bar{s}_{33}^2+v_{33}s_{33}^2+Q_{33a}+Q_{33b}$$

The BPO between model 2 and model3 is

$$
\begin{aligned}
BPO_{23} &= [\frac{p(M_2)}{p(M_3)}]*[\frac{p(y\,|\,M_2)}{p(y\,|\,M_3)}] = 2^{3-2}\,\frac{K_{21}K_{22}}{K_{31}K_{32}K_{33}}\left[\frac{|C_{21}\|C_{22}\|A_{31}\|A_{32}\|A_{33}|}{|A_{21}\|A_{22}\|C_{31}\|C_{32}\|C_{33}|}\right]^{1/2} \\
&\times \frac{2^{\frac{n_{21}+q_{21}}{2}}\Gamma(\frac{n_{21}+q_{21}}{2})2^{\frac{n_{22}+q_{22}}{2}}\Gamma(\frac{n_{22}+q_{22}}{2})}{2^{\frac{n_{31}+q_{31}}{2}}\Gamma(\frac{n_{31}+q_{31}}{2})2^{\frac{n_{32}+q_{32}}{2}}\Gamma(\frac{n_{32}+q_{32}}{2})2^{\frac{n_{33}+q_{33}}{2}}\Gamma(\frac{n_{33}+q_{33}}{2})} \\
&\times \frac{(q_{31}\bar{s}_{31}^2+v_{31}s_{31}^2+Q_{31a}+Q_{31b})^{-\frac{n_{31}+q_{31}}{2}}(q_{22}\bar{s}_{32}^2+v_{22}s_{32}^2+Q_{32a}+Q_{32b})^{-\frac{n_{32}+q_{32}}{2}}}{(q_{21}\bar{s}_{21}^2+v_{21}s_{21}^2+Q_{21a}+Q_{21b})^{-\frac{n_{21}+q_{21}}{2}}(q_{22}\bar{s}_{22}^2+v_{22}s_{22}^2+Q_{22a}+Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}} \\
&\times (q_{33}\bar{s}_{33}^2+v_{33}s_{33}^2+Q_{33a}+Q_{33b})^{-\frac{n_{33}+q_{33}}{2}}
\end{aligned}
\tag{4.22}
$$

where $K_1 = \dfrac{2(q_1\bar{s}_1^2/2)^{q_1/2}}{\Gamma(q_1/2)}$

$$K_{21} = \frac{2(q_{21}\bar{s}_{21}^2/2)^{q_{21}/2}}{\Gamma(q_{21}/2)}$$

$$K_{22} = \frac{2(q_{22}\bar{s}_{22}^2/2)^{q_{22}/2}}{\Gamma(q_{22}/2)}$$

$$K_{31} = \frac{2(q_{31}\bar{s}_{31}^2/2)^{q_{31}/2}}{\Gamma(q_{31}/2)}$$

$$K_{32} = \frac{2(q_{32}\bar{s}_{32}^2/2)^{q_{32}/2}}{\Gamma(q_{32}/2)}$$

$$K_{33} = \frac{2(q_{33}\bar{s}_{33}^2/2)^{q_{33}/2}}{\Gamma(q_{33}/2)}.$$

Further details, included those for the case when c=4 (*i.e.*, M4) are given in Appendix A.

4.2.4 Linearize the Non-linear Model for the BPO Analysis

In the above sections, we set up the Bayesian fuzzy regression technique based on a general linear regression model. However, for this South Africa equivalence scales study, the model we are employing is non-linear. As a result, we will use a Taylor's series approximation to linearize the model first. Our model has been specified in section 2 as:

$$y = \beta_1 + \beta_2 \times (x - \beta_3 \log(A + (\frac{\widehat{\beta}_4^{\,2}}{1+\widehat{\beta}_4^{\,2}}) \times K)) + \varepsilon \quad . \tag{2.14}$$

Applying a Taylor's series approximation, we can write the model into the following form (see Appendix B).

$$y - cons. = b_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 \qquad , \tag{4.23}$$

where:

$$cons. = [\tilde{\beta}_1 + \tilde{\beta}_2(x - \tilde{\beta}_3 \log(A + \tilde{\beta}_4 K))] - \tilde{\beta}_1 - \tilde{\beta}_2 x + 2\tilde{\beta}_2\tilde{\beta}_3 \log(A + \tilde{\beta}_4 K) + \frac{\tilde{\beta}_2\tilde{\beta}_3\tilde{\beta}_4{}^2}{A + \tilde{\beta}_4 K}$$

$$\begin{aligned} b_1 &= \beta_1 \\ b_2 &= \beta_2 \\ b_3 &= -(\beta_2\tilde{\beta}_3 + \beta_3\tilde{\beta}_2) \\ b_4 &= \beta_4. \end{aligned}$$

Here, $\tilde{\beta}$ is the estimator for β from a nonlinear regression for the whole sample set. It should be emphasized that this linearized model is used only for the analysis of the Bayesian Posterior Odds under the symmetric loss function. After we find the preferred value for *"c"* through the Bayesian Posterior Odds analysis, we will use the Bayesian nonlinear model to fit each of the clusters for the Bayesian fuzzy regression. The Bayesian Fuzzy Regression is the weighted average of the Bayesian regressions from each cluster, the weights being determined by the membership functions. Similarly, the estimators for the *βs* will be calculated as the weighted average of all the estimators for the *βs* from the results of all the clusters. The equivalence scales are then obtained using equation (2.13):

$$\Delta = (A + \beta_4 K)^{\beta_3} \tag{2.13}$$

4.2.5 Standard Errors

As the model is highly non-linear, it is difficult to obtain the standard errors for the equivalence scales. We used the bootstrap to approximate these standard errors and found that for a few households they are unexpectedly large compared with Yatchew (2003)[3].

---

[3] We also tried the Delta method (Greene 2003, Papke and Wooldridge, 2004), to calculate the standard errors, but these were huge compared with the one we got using the bootstrap. The local validity of the Delta method appears to be its greatest weakness.

We also constructed a 90% confidence interval for the reference household's Engel curve to check the significance of each equivalence scales. The equivalence scale for a household measures the horizontal shift needed for this household's Engel curve so that it may lie on top of the Engel curve of the reference household which means the welfare of the two households are the same after the adjustment. Hence, we argue that if the Engel curve of a household after the horizontal shift adjusted for its equivalence scale lies with the 90% confidence interval of the Engel curve for the reference household, the equivalence scale for this household is significant at a 10% significant level. Figure 7 to 12 show the relative position of the Engel curves for some household types after that adjustment of the equivalence scales.

5. Result

5.1.1 Engel Curves for Each Household Type

Using the single Working-Leser model in equation (3.07), we know this non-linear model can be simplified to a linear model with only one explanatory variable—the logarithm of total monthly expenditure and an intercept, since for the same household type, the number of children and adults will be the same. The model we use here is equation 2.05:

$$y = \alpha + \beta \ln(x) + \varepsilon \quad . \tag{2.05}$$

According to the family type, we group the sample into thirty-six groups based on the different demographic features of the household—the result is shown in Table 2. We use the logarithm of monthly total expenditure to cluster the sample of very household type into "c" clusters, where we use a Bayesian Posterior Odds analysis to pick the correct value for *"c"*. Since a simple plot of the data reveals that the assumption of one cluster for each of the explanatory variables is not a reasonable assumption, we restrict the value for *c* to be greater than one but no great than four. Table 4 gives the results of the Bayesian Posterior Odds analysis for each of the household types.

Table 4. Bayesian Posterior Odds Results for Each Household

A/K	0	1	2	3	4	5
1	3	2	2	2	3	2
2	3	2	2	2	2	2
3	2	2	2	3	2	2
4	2	2	2	2	2	2
5	2	2	2	2	2	2
6	2	2	2	2	2	2

Note: "A" is the number of adults in the household and "K" is the number of Children in the household; the selected value for *"c"* is reported in the table.

As the prior information for each of the parameters becomes diffuse, the Bayesian estimator collapses to the Maximum Likelihood estimator. The Bayesian Fuzzy Regression will be a combined regression of the results for each of the cluster using the membership function as the weights of each data for each household. Given we have only one explanatory variable in this household Engel Curve analysis, the membership function for each data set will be the membership function of each data point according to the logarithm of monthly total expenditure.

We expect to see a negative relation between the food share and the logarithm of total monthly expenditure, according to the Engel's Law. Figures 1 to 6 are the plots of the Engel Curves for each of the household type. As we can see, the Engel Curves are downward-sloped, except in a few cases where there are not a large number of observations. Then, the Engel Curves have a tendency to go up at the right tail. In Figure 1, given the same income level and the same number of adults in the family, we can see the Engel curve is lower when there are fewer children in the household at the range where the total expenditure is above the mean. This means that the relatively rich household (for which the household total expenditure is above the average level) has a smaller food share as the number of children declines. This does not happen to the poor household (for which the household total expenditure is well below the average level): a household with more children is likely to have less food share. The reason might be that besides the basic food consumption, rents[4] and clothes could be the other two important aspects of a household consumption. When the household is poor, the shares of the rent and clothes, especially rent, are relatively large compared with the rich household. Since the rent has occupied a big share of their monthly expenditure, the poor household is left with small margin to spend on food. Hence, when the household is poor, the bigger the household (or the more the children in a household given the number of adult in the household is the same cross the group), the larger the rent share will be and the smaller the food share will be. The same result can be found in the other five Figures as well. As the monthly total expenditure increases (or as the household becomes rich), the share that rent occupies in the total monthly expenditure will become relatively small. Thus, a bigger family will spend more money on food, and the food share will become larger as well.

5.1.2 Equivalence Scales for South Africa

Using the BPO analysis to choose the number of clusters to estimate the equivalence scales, under the assumption that each of the variables will have the same number of clusters every time, model two (*"c = 2"*) has been selected.

If there are two clusters for each regressor, log(x), A and K; we will potentially have 8 clusters, as is shown in Table 5.

Table 5: Eight Clusters for $k = 3$ & $c = 2$

Cluster 1	LOW income Household with A<= 3 and K<= 2
Cluster 2	HIGH income Household with A<= 3 and K<= 2
Cluster 3	LOW income Household with A<= 3 and K> 2
Cluster 4	HIGH income Household with A<= 3 and K> 2
Cluster 5	LOW income Household with A> 3 and K<= 2
Cluster 6	HIGH income Household with A> 3 and K<= 2
Cluster 7	LOW income Household with A> 3 and K> 2
Cluster 8	HIGH income Household with A> 3 and K> 2

---

[4] We realize that for some households living in a rural area, rent may not exist. However, other consumption as education, agriculture/farming expenditures, and etc., will occupy a large part of the household expenditure.

Using the linearized model in section 4.2.4 equation (4.20), our estimate of β_3 which is the estimator for the effect of the scale economies, is 0.58, and the estimator for β_4 the children's effect on the household equivalence scales is 0.75, compared with 0.59 and 0.74 respectively from Yatchew (2003).

The equivalence scales for each household type are reported in Table 6. The numbers in the parentheses are the standard errors obtained by bootstrapping.

Table 6. The Equivalence Scales for South Africa in 1993

Adult\Kid	0	1	2	3	4	5
1	1.0000 (0.0000)	1.3843 (0.2837)	1.7030 (0.2690)	1.9833 (0.5406)	2.2376 (0.4164)	2.4724 (0.3065)
2	1.4956 (0.0460)	1.7996 (0.2188)	2.0703 (0.2224)	2.3175 (0.2550)	2.5470 (0.5970)	2.7624 (1.1180)
3	1.8927 (0.1360)	2.1548 (0.1756)	2.3956 (0.1361)	2.6200 (1.1961)	2.8313 (1.9255)	3.0319 (0.5384)
4	2.2369 (0.0374)	2.4718 (0.1184)	2.6916 (2.2143)	2.8991 (0.3812)	3.0964 (1.5610)	3.2850 (0.4833)
5	2.5463 (0.2049)	2.7618 (0.5209)	2.9657 (0.5309)	3.1600 (0.6437)	3.3460 (0.5819)	3.5248 (0.7823)
6	2.8307 (0.0358)	3.0313 (0.1572)	3.2227 (0.4288)	3.4062 (2.0784)	3.5828 (0.7941)	3.7534 (6.3839)

Note: Bootstrap standard errors appear in parentheses.

For example, the equivalence scales for the household with two adults is 1.4956, which means that in order for support a two adult household at the same welfare level as the one adult only household, we need to scale the total monthly expenditure of the one adult only household by 1.4956. For a household of four adults and one child to be at the same level of the welfare as the one adult only household, the total monthly expenditure of this household have to be 2.4718 times of that of the one adult only household. Noticed the standard errors from the bootstrap indicate that the equivalence scales of five types of households are not significant which deviate a little bit from the results of Yatchew (2003) where all the equivalence scales are significant.

Table 7 reports the comparison of the equivalence scales from the two possible approaches for calculating the weight for each of the observations, using the membership functions for the fuzzy clusters.

Table 7. Comparison of the Equivalence Scales between the Product and Min form of the Weight for Each Observation

Adult\Kid	0	1	2	3	4	5
1	1.00	1.38	1.70	1.98	2.24	2.47
	1.00	*1.38*	*1.69*	*1.97*	*2.22*	*2.45*
2	1.50	1.80	2.07	2.32	2.55	2.76
	1.49	*1.79*	*2.06*	*2.30*	*2.53*	*2.74*
3	1.89	2.15	2.40	2.62	2.83	3.03
	1.88	*2.14*	*2.38*	*2.60*	*2.81*	*3.00*
4	2.24	2.47	2.69	2.90	3.10	3.29
	2.22	*2.45*	*2.67*	*2.87*	*3.07*	*3.25*
5	2.55	2.76	2.97	3.16	3.35	3.52
	2.53	*2.74*	*2.94*	*3.13*	*3.31*	*3.49*
6	2.83	3.03	3.22	3.41	3.58	3.75
	2.81	*3.01*	*3.19*	*3.37*	*3.55*	*3.71*

The number reported in every first row of the equivalence scales is from the product form of the membership function and the number in italics in every second row is the equivalence scales estimated using the minimum form of the membership function. We find that the results are very similar even though different forms of the membership function are used. This may well indicate that the fuzzy regression results are relatively robust.

Table 8 gives us a comparison between the equivalence scales from this Bayesian fuzzy regression analysis and the semiparametric index model of Yatchew (2003). We can see that the results are very similar.

Table 8. A Comparison for the Equivalence Scales with Yatchew's Result

Adult\Kid	0	1	2	3	4	5
1	1.00	1.38	1.70	1.98	2.24	2.47
	1.00	*1.39*	*1.71*	*1.99*	*2.25*	*2.49*
2	1.50	1.80	2.07	2.32	2.55	2.76
	1.51	*1.81*	*2.09*	*2.34*	*2.57*	*2.79*
3	1.89	2.15	2.40	2.62	2.83	3.03
	1.91	*2.18*	*2.42*	*2.65*	*2.87*	*3.07*
4	2.24	2.47	2.69	2.90	3.10	3.29
	2.27	*2.50*	*2.73*	*2.94*	*3.14*	*3.33*
5	2.55	2.76	2.97	3.16	3.35	3.52
	2.58	*2.80*	*3.01*	*3.21*	*3.40*	*3.58*
6	2.83	3.03	3.22	3.41	3.58	3.75
	2.88	*3.08*	*3.28*	*3.47*	*3.65*	*3.82*

Note: The results in italics are from Yatchew (2003)

The construction of confidence intervals can be found from Figure 7 to 12. These 90% confidence intervals are for the base group. An equivalence scale is measured by the horizontal shift needed for the Engel curve of a household in order that it will coincide with the Engel curve from the base group. Here, if we see the Engel curve for a household falls into the confidence interval of that of the base group after the horizontal shift according to it's equivalence scales get, then we say that the equivalence scale for this type of household is significant at the 10% significant level. Figures from 7 to 12 show that the equivalence scales we get for many of the household types are significant at the 10% significant level and also, there is not a single Engel curve that falls outside the confidence interval after the adjustment for all the income range. This result is similar to Yatchew (2003) where he found that all the equivalence scales are significant.

6. Conclusions and Future Suggestions

In this study, we have shown the highly satisfactory performance of the Bayesian fuzzy regression in this study of the equivalence scale estimation for the South Africa. A single-equation Working-Leser model allowing for different demographic features of a household is used. Assuming each explanatory variable has the same number of clusters, between one and four, the BPO analysis picks up two to be the number of clusters for each explanatory variable. Given the three variables employed in the model: the logarithm monthly total expenditure, the number of adults and children in a household, we get eight valid clusters with positive degrees of freedom. The equivalence scale we got using the Bayesian fuzzy regression analysis is very close to what Yatchew (2003) found using a single index model with a semi-parametric approach. Besides the usual advantages of any Bayesian approach—unity of the approach to all types of inferences, prior information incorporated explicitly and flexibly, admissibility of the estimator *etc*.—the fast computational time seems to be another gain of this Bayesian fuzzy regression. By employing two forms of the membership function—product form and min form, we see the equivalence scales are similar, which suggests that the fuzzy regression is fairly robust to the different choice of the membership function form.

One possible future extension of this study would be to relax the restriction that the number of clusters for each explanatory variable is the same. If we can allow that the number of clusters for each variable to be different, for instance, one variable may have two clusters and another variable may have one cluster and the third could have three clusters, even when we restrict the number of clusters less or equal to four, we might end up with lots of combination of clusters for this equivalence scales study. In particular, due to the nature of the data in this study—the number of children runs from zero to five, and the number of adults runs from one to six, an assumption that there are up to six clusters for the A and K variables would be very reasonable. Thus, we would end up analyzing four cases: logarithm monthly total expenditure has one, two, three, four clusters while A, K have six clusters each.

This research takes the first initiative in using the Bayesian approach of model selection in determining the number of clusters in a fuzzy regression. This Bayesian approach has significantly improved the efficiency of the fuzzy regression analysis. Our research findings on equivalence scales based on this Bayesian fuzzy regression approach are very encouraging. We expect to see more applications of this methodology in other areas of econometric studies in the future.

Reference:

Aasness, J. and A. Rodseth (1983), "Engel Curves and Systems of Demand Functions." *European Economic Review,* 20, 95-121.

Bewley, R. (1982), "On the functional form of Engel curves: the Australian household expenditure survey 1975-76", *Economic Record*, 58, 82-91.

Bezbek, J. C. (1973), Fuzzy Mathematics in Pattern Classification, Ph.D. Thesis, Applied Mathematics Center, Cornell University, Ithaca, NY.

Bezbek, J. C. (1981), Pattern Recognition With Fuzzy Objective Function Algorithms, Plenum Press, New York.

Citro, C., and Michael, R. T. (eds.) (1995), *Measuring Poverty—A New Approach,* Washington DC: National Academy Press.

Dunn, J. C. (1974), "Well Separated Clusters and Optimal Fuzzy Partitions", *Journal of Cybernetics*, 4, 95-104.

Dunn, J. C. (1977), "Indices of Partition Fuzziness and the Detection of Clusters in Large Data Sets", in M. Gupta and G. Seridis (eds.), *Fuzzy Automata and Decision Processes*, Elsevier, New York.

Deaton, A. S. and J. Muellbauer (1980), "An almost ideal demand system", *American Economic Review*, 70, 312-326.

Deaton, A. (1997), *The Analysis of Household Survey: A Microeconomic Approach to Development Policy,* Baltimore: Johns Hopkins University Press for the World Bank.

Draeseke, R. and D.E.A. Giles (2002), "Modelling the New Zealand Underground Economy Using Fuzzy Logic Techniques", *Mathematics and Computers in Simulation*, 59, 115-123

Lewbel, A. (1997), "Consumer Demand Systems and Household Equivalence Scales," in M.H. Pesaran and P. Schmidt(eds.), Handbook of Applied Econometrics, Volume II: Microeconomics, Oxford: Blackwell Handbooks in Economics, 166-201.

Leser, C.E.V. (1941), Family budget data and price-elasticities of demand, *Review of Economic Studies*, 9, pp. 40-57.

Leser, C.E.V. (1963), "Forms of Engel Functions," *Econometrica*, 31, 694-703.

Giles, D. E. A. (2005), “Output convergence and international trade: Time-series and fuzzy clustering evidence for New Zealand and her trading partners, 1950-1992”, *Journal of International Trade and Economic Development,* 14, 93-114.

Giles, D. E. A. and R. Draeseke (2003), “Econometric modeling using fuzzy pattern recognition via the fuzzy c-means algorithm”, in D. E. A. Giles (ed.), Computer Aided Econometrics, Marcel Dekker, New York, 407-450.

Giles, D.E.A. and Hui Feng (2005), "Output and Well-Being in Industrialized Nations in the Second Half of the 20th Century: Testing for Convergence Using Fuzzy Clustering Analysis", *Structural Change & Economic Dynamics*, 2005, 16, 285-308.

Giles, D. E. A. and P. Hampton (1985), "An Engel curve analysis of household expenditure in New Zealand", *Economic Record*, 61, 450-462.

Giles, D. E. A. and C. A. Mosk (2003), “A long-run environmental Kuznets curve for enteric CH4 emissions in New Zealand: A fuzzy regression analysis”, Econometrics Working Paper EWP0307, Department of Economics, University of Victoria, in preparation.

Giles, D.E.A. and C. Stroomer (2004), “Identifying the Cycle of a Macroeconomic Time-Series Using Fuzzy Filtering”, Econometrics Working Paper EWP0406, Department of Economics, University of Victoria.

Giles, D.E.A. and C. Stroomer (2005), “Does Trade Openness Affect the Speed of Output Convergence? Some Empirical Evidence”, Econometrics Working Paper EWP0509, Department of Economics, University of Victoria.

Greene, W. H. (2003), *Econometric Analysis* 5th Edition, Prentice Hall

Houthakker, H.S. (1960), Additive preferences, *Econometrica*, 28, pp. 244-257.

MacQueen, J. (1967), “Some Methods for Classification and Analysis of Multivariate Observations”, in J. M. Le Cam and J. Neyman (eds.), *Proceedings of the 5th Berkeley Symposium in Mathematical Statistics and Probability*, University of California Press, Berkeley CA, 281-297.

Papke L.E., Wooldridge J.W. (2004), “A Computational Trick for Delta-method Standard Errors” *Economic Letters*, 2004

Ruspini, E. (1970), "Numerical Methods for Fuzzy Clustering", *Information Science*, 2, 319-350.

Prais, S. J. and H.S. Houthakker (1955), *The Analysis of Family Budgets*, Cambridge: Cambridge University Press.

SHAZAM (2001), SHAZAM Econometrics Package, User's Guide, Version 9, Northwest Econometrics, Vancouver, B.C..

Shepherd, D. and F. K. C. Shi (1998), "Economic modelling with fuzzy logic", paper presented at the CEFES '98 Conference, Cambridge, U.K..

Working, H. (1943) "Statistical Laws of family expenditure", *Journal of the American Statistical Association*, 38, 43-56.

Yatchew, A., Y. Sun and C. Deri (2001), "Efficient Estimation of Semi-parametric Equivalence Scales with Evidence from South Africa", *Journal of Business and Economic Statistics*, **21**, 245-257.

Yatchew, A., (2003), "Index Models and Other Semiparametric Specifications", *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press, page 138-150

Zadeh, L. A. (1965), "Fuzzy Sets". *Information and Control*, 8, 338-353.

Zadeh, L.A.(1987), "Fuzzy sets". In R.R. Yager, S. Ovchinnikov, R.M. Tong, and H.T. Nguyen, editors, *Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh*, pp. 29–44. Wiley & Sons. Originally published in *Information and Control*, 8, 1965, pp. 338–353.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics,* New York: John Wiley & Sons.

Figure 1. The Engel Curves for 1 Adult Household: Kids from 0 to 5

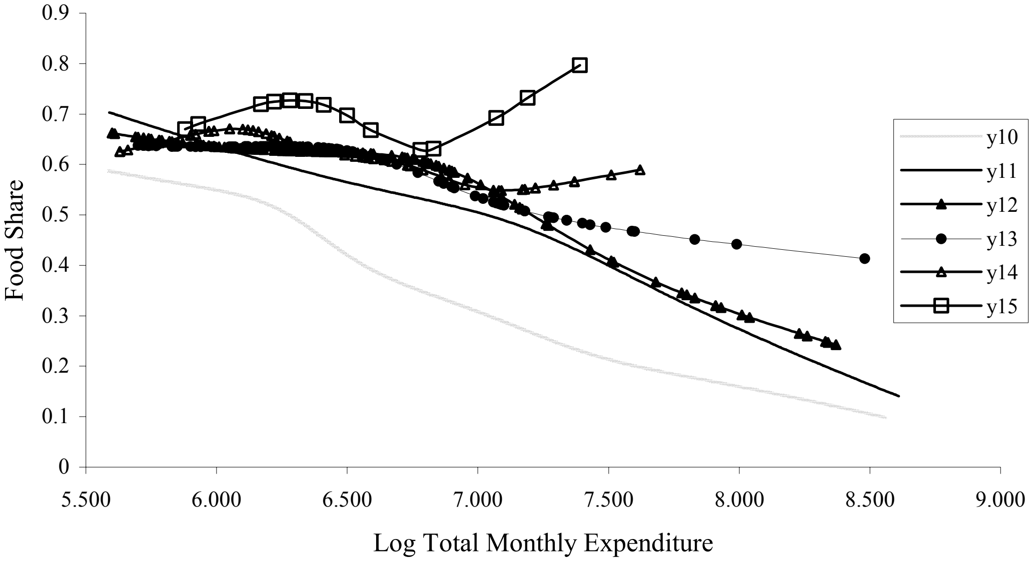


Figure 2. Engel Curve 2 Adults Families: Kids runs fron 0 to 5

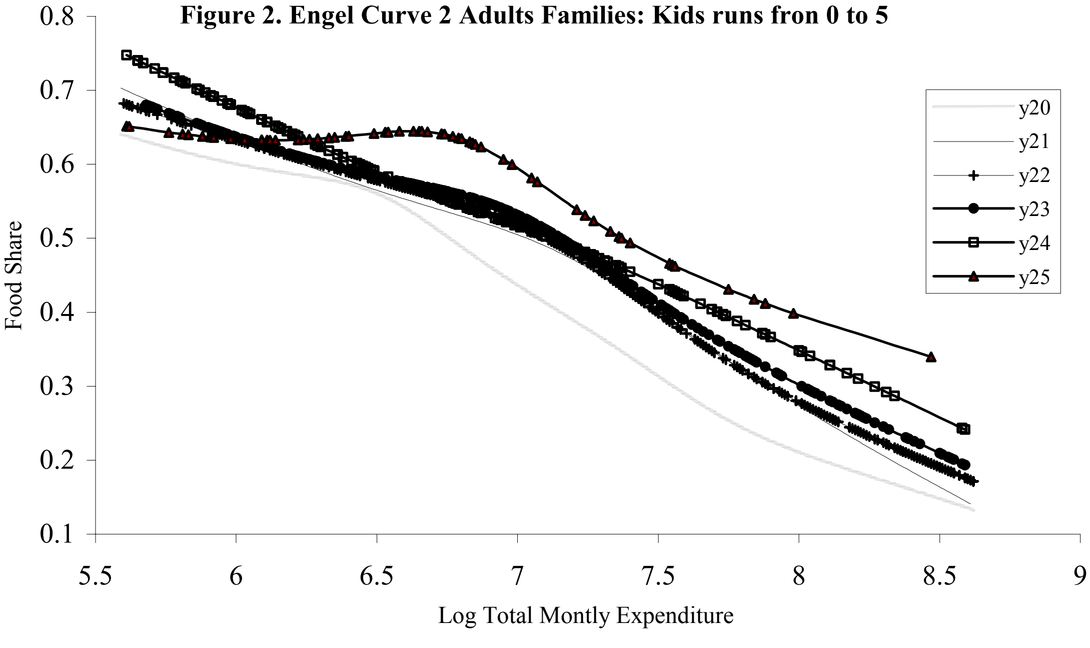


Figure 3. Engel Curves 3 Adults Household: Kids runs from 0 to 5

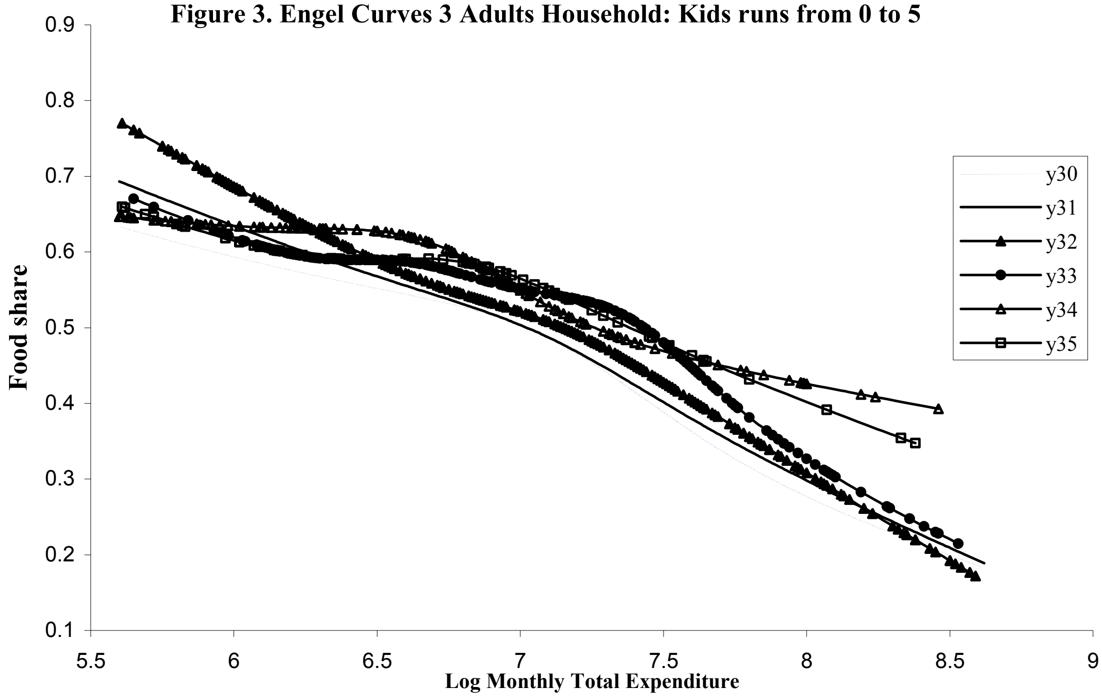


Figure 4. Engel Curves for 4 Adults Household: Kids runs from 0 to 5

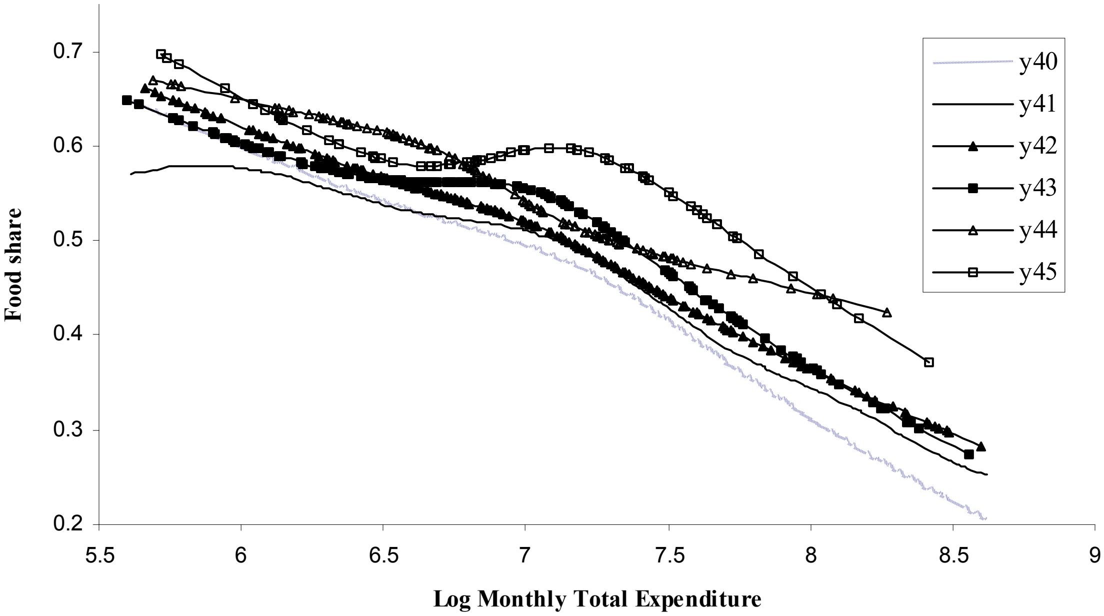


Figure 5. Engel Curves for 5 Adults Households: Kids runs from 0 to 5

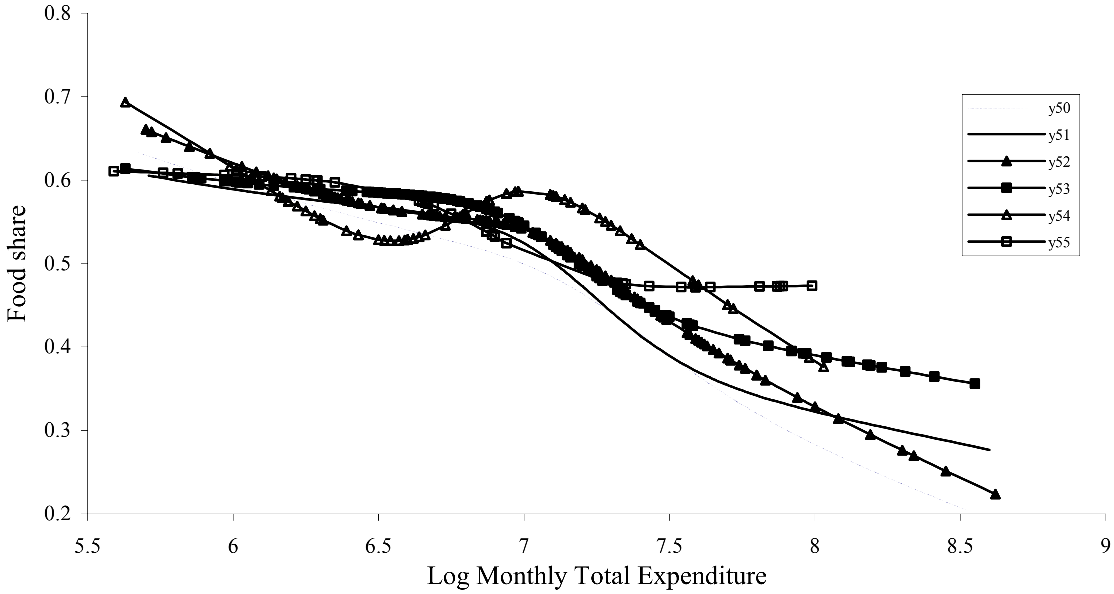


Figure 6. Engel Curves for 6 Adults Households: Kids runs from 0 to 5

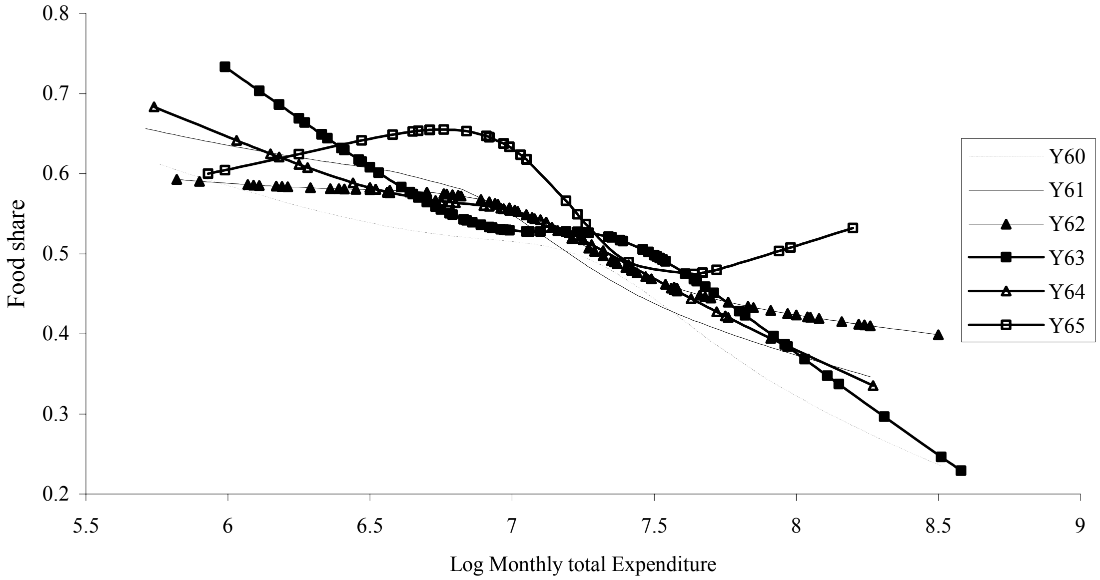


Figure 7. Engel Curves adjusted for the Equivalence Scales for 1 Adults Households

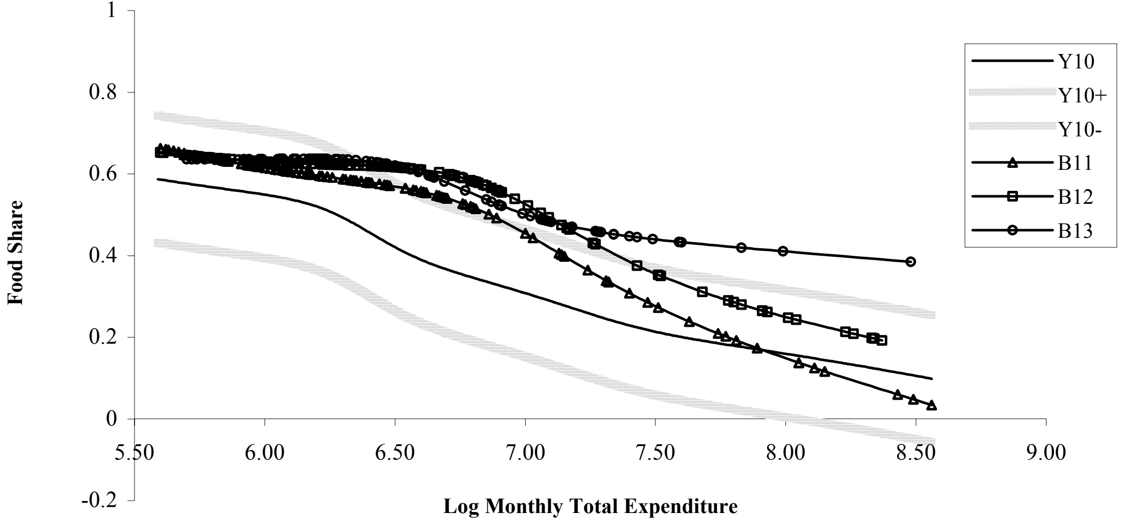


Figure 8 Engel Curves Adjusted for the Equivalence Scalse for 2 Adults Households

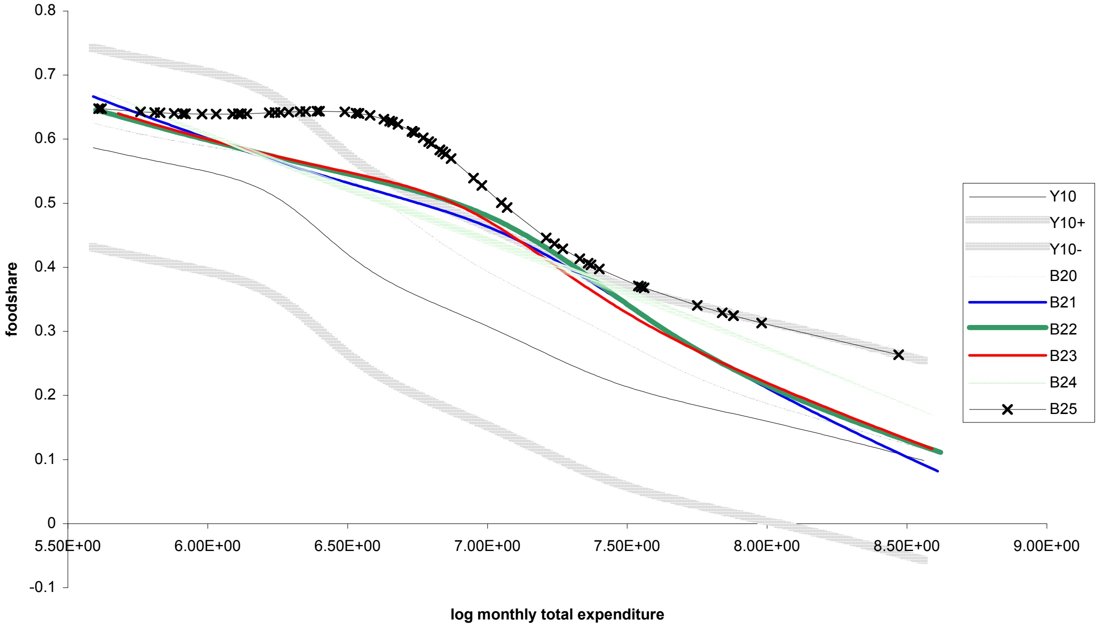


Figure 9. Engel Curves Adjusted for The Equivalence Scales for 3 Adults Households

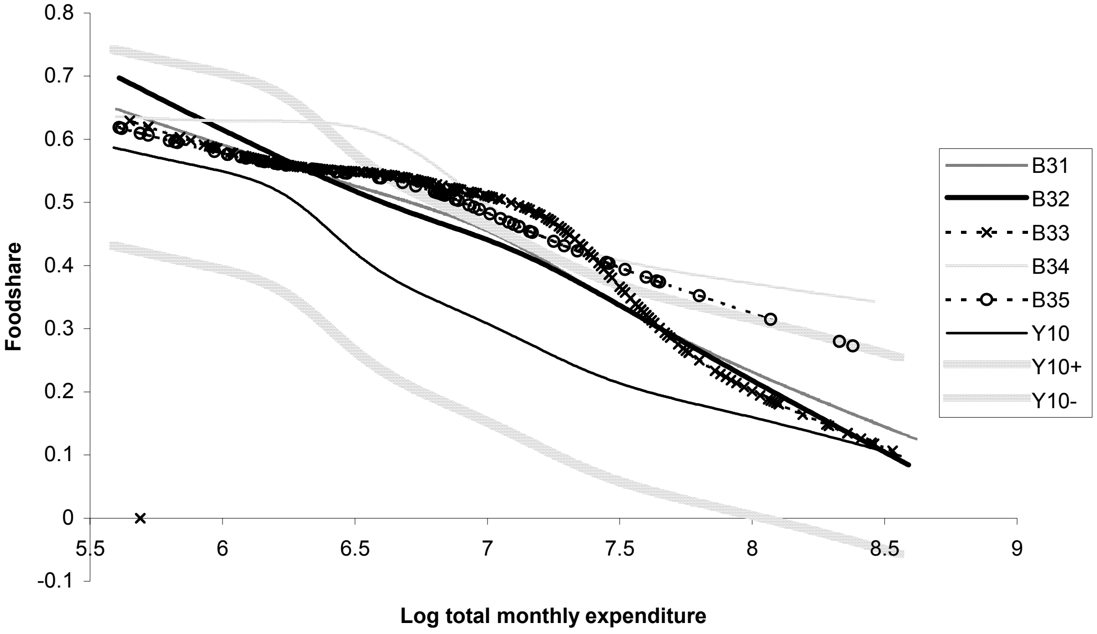


Figure 10. Engel Curves for 4 Adults Households Adjusted for the Equivalence Scales

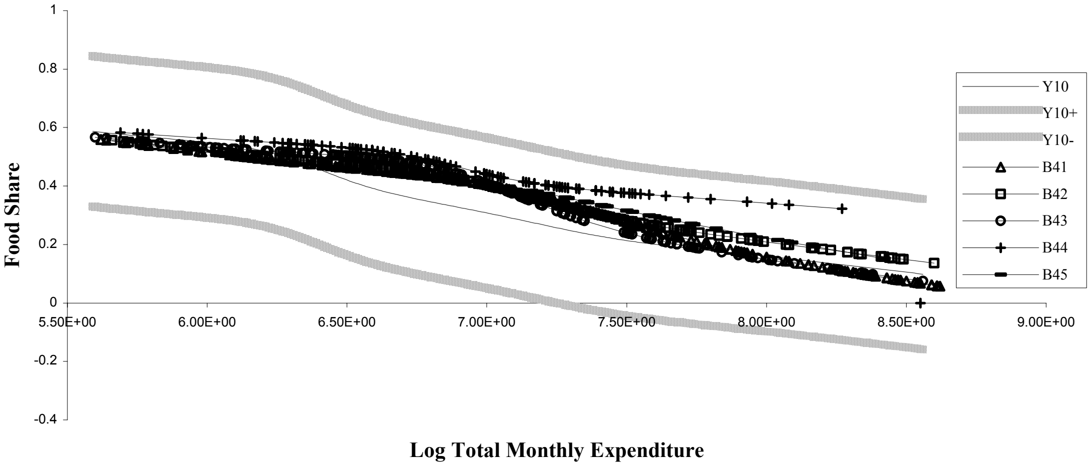


Figure 11. Engel Curves for 5 Adults Households Adjusted for the Equivalence Scales

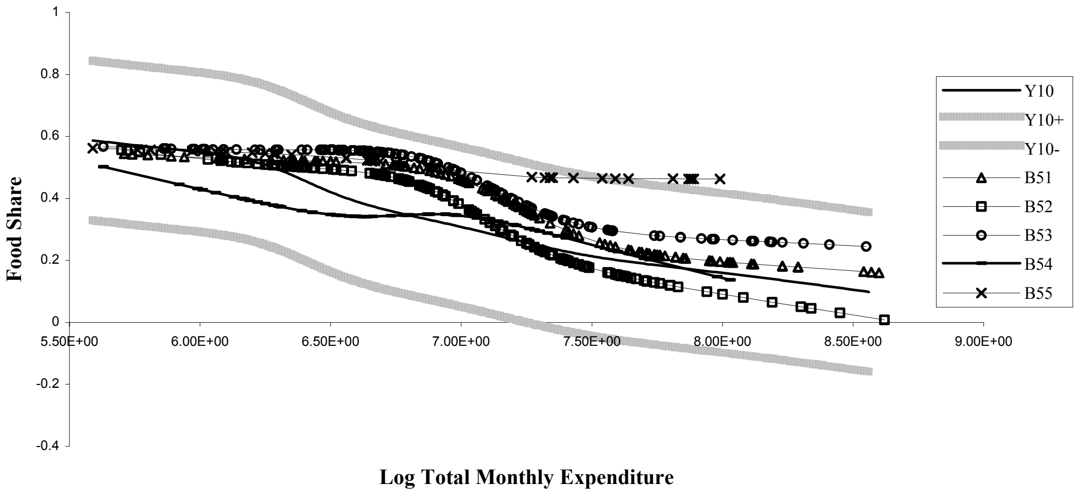
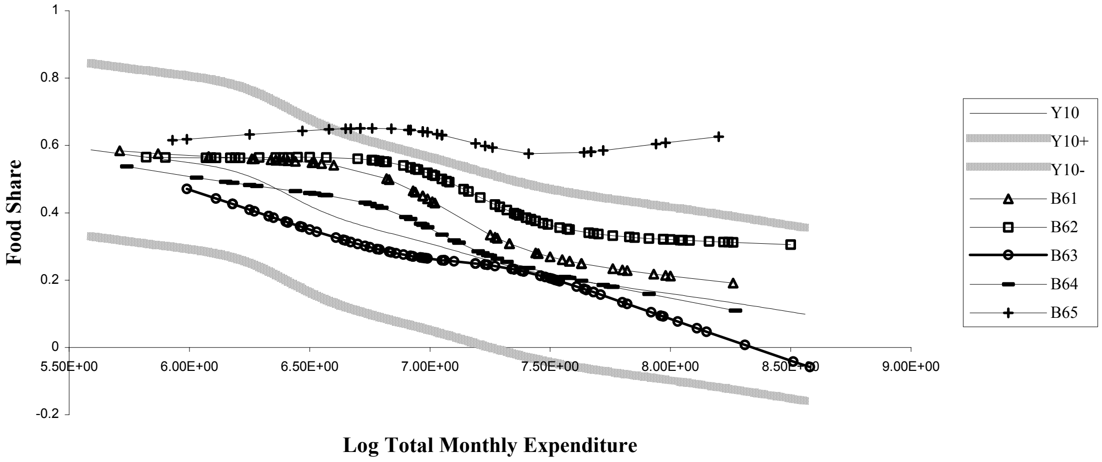


Figure 12. Engel Curves for for 6 Adults Households Adjusted for the Equivalence Scales



Appendix A: Bayesian Posterior Odds Under the Natural Conjugate prior under the Diffuse Case for Bayesian Fuzzy Models as "*c*" Runs from 1 to 4

$$BPO_{14} = [\frac{p(M_1)}{p(M_4)}]*[\frac{p(y|M_1)}{p(y|M_4)}] = 2^{4-1}\frac{K_1}{K_{41}K_{42}K_{43}K_{44}}\left[\frac{|C_1||A_{41}||A_{42}||A_{43}||A_{44}|}{|A_1||C_{41}||C_{42}||C_{43}||C_{43}|}\right]^{1/2}$$

$$\times\frac{2^{\frac{n_1+q_1}{2}}\Gamma(\frac{n_1+q_1}{2})}{2^{\frac{n_{41}+q_{41}}{2}}\Gamma(\frac{n_{41}+q_{41}}{2})2^{\frac{n_{42}+q_{42}}{2}}\Gamma(\frac{n_{42}+q_{42}}{2})2^{\frac{n_{43}+q_{43}}{2}}\Gamma(\frac{n_{43}+q_{43}}{2})2^{\frac{n_{44}+q_{44}}{2}}\Gamma(\frac{n_{44}+q_{44}}{2})}$$

$$\times\frac{(q_{41}\bar{s}_{41}^2+v_{41}s_{41}^2+Q_{41a}+Q_{41b})^{-\frac{n_{41}+q_{41}}{2}}(q_{42}\bar{s}_{42}^2+v_{42}s_{42}^2+Q_{42a}+Q_{42b})^{-\frac{n_{42}+q_{42}}{2}}}{(q_1\bar{s}_1^2+v_1s_1^2+Q_{1a}+Q_{1b})^{-\frac{n_1+q_1}{2}}}$$

$$\times(q_{43}\bar{s}_{43}^2+v_{43}s_{43}^2+Q_{43a}+Q_{43b})^{-\frac{n_{43}+q_{43}}{2}}(q_{43}\bar{s}_{43}^2+v_{43}s_{43}^2+Q_{43a}+Q_{43b})^{-\frac{n_{43}+q_{43}}{2}}$$

$$BPO_{24} = [\frac{p(M_2)}{p(M_4)}]*[\frac{p(y|M_2)}{p(y|M_4)}] = 2^{4-2}\frac{K_2}{K_{41}K_{42}K_{43}K_{44}}\left[\frac{|C_{21}||C_{22}||A_{41}||A_{42}||A_{43}||A_{44}|}{|A_{21}||A_{22}||C_{41}||C_{42}||C_{43}||C_{43}|}\right]^{1/2}$$

$$\times\frac{2^{\frac{n_{21}+q_{21}}{2}}\Gamma(\frac{n_{21}+q_{21}}{2})2^{\frac{n_{22}+q_{22}}{2}}\Gamma(\frac{n_{22}+q_{22}}{2})}{2^{\frac{n_{41}+q_{41}}{2}}\Gamma(\frac{n_{41}+q_{41}}{2})2^{\frac{n_{42}+q_{42}}{2}}\Gamma(\frac{n_{42}+q_{42}}{2})2^{\frac{n_{43}+q_{43}}{2}}\Gamma(\frac{n_{43}+q_{43}}{2})2^{\frac{n_{44}+q_{44}}{2}}\Gamma(\frac{n_{44}+q_{44}}{2})}$$

$$\times\frac{(q_{41}\bar{s}_{41}^2+v_{41}s_{41}^2+Q_{41a}+Q_{41b})^{-\frac{n_{41}+q_{41}}{2}}(q_{42}\bar{s}_{42}^2+v_{42}s_{42}^2+Q_{42a}+Q_{42b})^{-\frac{n_{42}+q_{42}}{2}}}{(q_{21}\bar{s}_{21}^2+v_{21}s_{21}^2+Q_{21a}+Q_{21b})^{-\frac{n_{21}+q_{21}}{2}}(q_{22}\bar{s}_{22}^2+v_{22}s_{22}^2+Q_{22a}+Q_{22b})^{-\frac{n_{22}+q_{22}}{2}}}$$

$$\times(q_{43}\bar{s}_{43}^2+v_{43}s_{43}^2+Q_{43a}+Q_{43b})^{-\frac{n_{43}+q_{43}}{2}}(q_{43}\bar{s}_{43}^2+v_{43}s_{43}^2+Q_{43a}+Q_{43b})^{-\frac{n_{43}+q_{43}}{2}}$$

$$BPO_{34} = [\frac{p(M_3)}{p(M_4)}]*[\frac{p(y\,|\,M_3)}{p(y\,|\,M_4)}] = 2^{4-3}\,\frac{K_{31}K_{32}K_{33}}{K_{41}K_{42}K_{43}K_{44}}\left[\frac{|C_{31}\|C_{32}\|C_{33}\|A_{41}\|A_{42}\|A_{43}\|A_{44}|}{|A_{31}\|A_{32}\|A_{33}\|C_{41}\|C_{42}\|C_{43}\|C_{43}|}\right]^{1/2}$$

$$\times\frac{2^{\frac{n_{31}+q_{31}}{2}}\Gamma(\frac{n_{31}+q_{31}}{2})2^{\frac{n_{32}+q_{32}}{2}}\Gamma(\frac{n_{32}+q_{32}}{2})2^{\frac{n_{33}+q_{33}}{2}}\Gamma(\frac{n_{33}+q_{33}}{2})}{2^{\frac{n_{41}+q_{41}}{2}}\Gamma(\frac{n_{41}+q_{41}}{2})2^{\frac{n_{42}+q_{42}}{2}}\Gamma(\frac{n_{42}+q_{42}}{2})2^{\frac{n_{43}+q_{43}}{2}}\Gamma(\frac{n_{43}+q_{43}}{2})2^{\frac{n_{44}+q_{44}}{2}}\Gamma(\frac{n_{44}+q_{44}}{2})}$$

$$\times\frac{(q_{41}\bar{s}_{41}^2+v_{41}s_{41}^2+Q_{41a}+Q_{41b})^{-\frac{n_{41}+q_{41}}{2}}(q_{42}\bar{s}_{42}^2+v_{42}s_{42}^2+Q_{42a}+Q_{42b})^{-\frac{n_{42}+q_{42}}{2}}}{(q_{31}\bar{s}_{31}^2+v_{31}s_{31}^2+Q_{31a}+Q_{31b})^{-\frac{n_{31}+q_{31}}{2}}(q_{32}\bar{s}_{32}^2+v_{32}s_{32}^2+Q_{32a}+Q_{32b})^{-\frac{n_{32}+q_{32}}{2}}}$$

$$\times\frac{(q_{43}\bar{s}_{43}^2+v_{43}s_{43}^2+Q_{43a}+Q_{43b})^{-\frac{n_{43}+q_{43}}{2}}(q_{44}\bar{s}_{44}^2+v_{44}s_{44}^2+Q_{44a}+Q_{44b})^{-\frac{n_{44}+q_{44}}{2}}}{(q_{33}\bar{s}_{33}^2+v_{33}s_{33}^2+Q_{33a}+Q_{33b})^{-\frac{n_{33}+q_{33}}{2}}}$$

Appendix B. Linearize the Working-Leser Model using the Taylor's Series

Let $\frac{\widehat{\beta}_4^{\,2}}{1+\widehat{\beta}_4^{\,2}} = \beta_4 \in [0,1]$

$$\begin{aligned}
y &= \beta_1 + \beta_2 \times (x - \beta_3 \log(A + \beta_4 K)) \\
&= \beta_1 + \beta_2 x - \beta_2\beta_3 \log(A(1 + \frac{\beta_4}{A}K)) \\
&= \beta_1 + \beta_2 x - \beta_2\beta_3 \log(A) - \beta_2\beta_3 \log((1 + \frac{\beta_4}{A}K))
\end{aligned}$$

Apply Taylor's Series of the first order, we get

$$\begin{aligned}
y &= [\tilde{\beta}_1 + \tilde{\beta}_2 x - \tilde{\beta}_2\tilde{\beta}_3 \log(A) - \tilde{\beta}_2\tilde{\beta}_3 \log(1 + \tilde{\beta}_4 \frac{K}{A})] + \frac{(\underset{\sim}{\beta} - \underset{\sim}{\tilde{\beta}})' H(\underset{\sim}{\beta})(\underset{\sim}{\beta} - \underset{\sim}{\tilde{\beta}})}{2!} + \ldots\ldots \\
&= [\tilde{\beta}_1 + \tilde{\beta}_2 x - \tilde{\beta}_2\tilde{\beta}_3 \log(A) - \tilde{\beta}_2\tilde{\beta}_3 \log(1 + \tilde{\beta}_4 \frac{K}{A})] + \begin{bmatrix} \beta_1 - \tilde{\beta}_1 \\ \beta_2 - \tilde{\beta}_2 \\ \beta_3 - \tilde{\beta}_3 \\ \beta_4 - \tilde{\beta}_4 \end{bmatrix}' \\
&\times \begin{bmatrix} 1 \\ x - \tilde{\beta}_3 \log(A) - \tilde{\beta}_3 \log(1 + \tilde{\beta}_4 (K/A)) \\ -\tilde{\beta}_2 \log(A) - \tilde{\beta}_3 \log(1 + \tilde{\beta}_4 (K/A)) \\ -\tilde{\beta}_2\tilde{\beta}_3\tilde{\beta}_4 (\frac{K/A}{(1 + \tilde{\beta}_4 (K/A))}) \end{bmatrix} \\
&= \ldots\ldots \\
&\ldots \\
&= Cons. + \beta_2 x - (\beta_2\tilde{\beta}_3 + \beta_3\tilde{\beta}_2)(\log(A) + \log(1 + \tilde{\beta}_4 \frac{K}{A})) + \beta_4 (\frac{\tilde{\beta}_2\tilde{\beta}_3\tilde{\beta}_4 (K/A)}{1 + \tilde{\beta}_4 (K/A)})
\end{aligned}$$

where

$$cons. = [\tilde{\beta}_1 + \tilde{\beta}_2 (x - \tilde{\beta}_3 \log(A + \tilde{\beta}_4 K))] - \tilde{\beta}_1 - \tilde{\beta}_2 x + 2\tilde{\beta}_2\tilde{\beta}_3 \log(A + \tilde{\beta}_4 K) + \frac{\tilde{\beta}_2\tilde{\beta}_3\tilde{\beta}_4^{\,2}}{A + \tilde{\beta}_4 K}$$

$x_2 = x$

let $x_3 = \log(A + \tilde{\beta}_4 K)$

$x_4 = \tilde{\beta}_2\tilde{\beta}_3 K / (A + \tilde{\beta}_4 K)$

we could write the linearized model into the following form

$$y - cons. = b_1 + b_2 x_2 + b_3 x_3 + b_4 x_4$$

where:

$$\begin{aligned} b_1 &= \beta_1 \\ b_2 &= \beta_2 \\ b_3 &= -(\beta_2 \tilde{\beta}_3 + \beta_3 \tilde{\beta}_2) \\ b_4 &= \beta_4 \end{aligned}$$