University of Pretoria

The Aid and Maid System: South African Household Data Pitfalls

Steve F Koch[1]

Working Paper Number 24

[1] Department of Economics, University of Pretoria

THE AID AND MAID SYSTEM: SOUTH AFRICAN HOUSEHOLD DATA PITFALLS*

STEVEN F. KOCH$^{\diamond}$

ABSTRACT. This paper presents research on South African household expenditure share behaviour. The research examines whether or not a theoretical and empirical model, which has been successful in explaining expenditure shares in Australia, is valid when applied to South African data. The primary conclusion of the research is that expenditure shares in South Africa do not conform to the assumptions set out in the model. Although there are many potential reasons for non-conformity, this paper provides evidence that the estimates produced within the AID System and the MAID System suffer from heteroskedasticity and non-normality. Therefore, in order to improve the understanding of spending behaviour by South African households, models will have to be specifically developed to deal with the idiosyncrasies of South African data.

JEL Classification: C31, D12.

1. Introduction

For South African academics, it is often tempting to use economic or econometric models previously used for the analysis of data from developed countries. Although the reasons are many, the most likely explanation for the aforementioned temptation comes down to the fact that few economic theories are originated locally, while econometric software has been geared toward the developed economy academic audience. Unfortunately, the majority of developed economy researchers do not face the same set of application and data problems that are faced by developing economy academics, and, therefore, the assumptions in their models may not be applicable in developing economies, or, at least in South Africa, which is where we apply these models. Since the assumptions from models outside South Africa could be orthogonal to the actual situation on the ground, it is essential that South African researchers carefully consider whether or not the models to be applied, indeed, make sense.

The temptation to apply outside models exists at all levels of economic analysis, including the analysis of consumer behaviour, which is the focus of this paper. Although an economic analysis of consumer behaviour, regardless of commodity, is best achieved through the derivation, specification and estimation of demand equations, applying the correct empirical specification is subject to debate. Importantly, however, those demand equations should follow from consumer theory such that the basic economic assumptions can be tested, while the empirical results can be logically interpreted and used for valid policy formulation. Theoretically, these demand curves are underpinned by individually rational behaviour, which is influenced by the presence of resource constraints. Marshallian demand curves are just one example of rational decision-making behaviour at the individual level.

Although the derivation of demand functions is well established, and the estimation of these demand functions goes back many years, some empirical and theoretical issues remain. The main theoretical and empirical issue to be discussed in this paper is whether South African household data conforms to the predictions

of the Almost Ideal Demand System (AIDS), or to the predictions of a variant of that system, referred to as the Modified Almost Ideal Demand System (MAIDS). The empirical analysis suggests that the application of these models to the South African situation is less than ideal, most likely due to a varied set of data problems, including, but not limited to measurement error and misspecification.

Using data from the 2000 South African Income and Expenditure Survey (SAIES), systematic estimates of expenditure patterns for single-person households are estimated. The results of those estimations are reported in this paper. During the course of the investigation, a number of interesting problems arose and limited the study. Initially, two-person households were used in the analysis, because there were a large number of households in the two race groups studied. However, the results of the analysis suggested that two-person households did not conform well to consumer theory, even in the simple models estimated here, although the violation is not necessarily surprising if each individual in the household maximizes their own utility, while the data captures household level decisions.[1]

Due to the difficulties encountered while studying two-person households, single-person households are used in the analysis presented here. Interestingly, single-person households do not always conform to the simple theory examined in this paper, either. The primary reason for that result may result from the inability of the empirical model to distinguish between potentially heterogeneous populations. For example, another problem that arose in the analysis was the existence of bimodal error distributions. Although the level of aggregation used in the analysis was minimal, in order to limit the occurrence of zeroes in the data, the estimation procedure (suggested by researchers in developed economies) based on share ratios and the replacement of zeroes in the data, resulted in obvious bimodality. The resulting violation of the error distribution assumptions suggests that the technique suggested for developed economy data cannot be applied to South African data with impunity.

---

[1] I thank Duncan Thomas and Chris Udry for pointing out this potential problem.

The paper is presented in five additional sections. In Section 2, a brief history of demand system analysis is presented. As there are few published papers using South African data within a demand system, there will not be a separate South African history. The theoretical and empirical underpinnings are presented in Section 3. The data source and summary statistics of the data are discussed in Section 4, while the presentation of the analysis results is in Section 5. Concluding comments and remarks are provided in Section 6.

2. Demand Systems: A Brief History

Empirically, there are two diverse demand estimation approaches, which can be distinguished by the number of equations to be estimated simultaneously. Single equation estimates, as can be garnered from the name, involves the estimation of a single equation; although there could be a large number of equations to be estimated, each equation will be estimated alone. System estimates, by contrast, make use of the systemic nature of individual behaviour while estimating all equations simultaneously.

The initial methodology employed in applied demand analysis, dating back to Moore (1914), was based on a single equation model; Stone's (1954a) investigation of non-durable goods expenditure patterns in the UK represents the classic application of the single-equation approach. However, Stone (1954b) did provide the precursor to estimating a complete system of demand equations.

As is often the case in applied econometrics, estimation of single-demand equations was severely hampered by data limitations. Microeconomic theory ascertains that a multitude of prices affect the demand for a particular commodity. To include all such prices in an econometric estimation is cumbersome, if not impossible, given limited degrees of freedom. Therefore, many potential explanatory variables must be eliminated from the equation; however, those restrictions could render the system unrelated to economic theory, or, worse, lead to estimates that are biased

because certain necessary variables were not included in the estimation. Fortunately, within an equation system, derived from consumer theory, there exists a series of cross-equation restrictions, which can significantly reduce the number of parameters to be estimated. In addition, provided these restrictions are valid, their imposition improves the efficiency of the estimation procedures employed and allows for more precise estimates of the parameters of the demand equations to be obtained (Thomas, 1987). Thus, estimation of a system of demand functions enables the econometrician to obtain estimates for each single equation that are better than if each equation were estimated in isolation, i.e., if each equation was estimated without all of the relevant information.

As already mentioned, Stone (1954b) conducted the pioneering demand system study. His Linear Expenditure System (LES), derived from a particular utility function, ensured that the restrictions from consumer theory were met in the empirical specification; unfortunately, it was not possible to test the restrictions, due to the fact that the utility function was not general. The Rotterdam model, developed by Theil (1965), answered the preceding criticism to the LES. In some sense, Theil's approach erred too far in the opposite direction. Although he did not use an explicit utility function, the collection of demand equations, which may or may not satisfy the theoretical restrictions, the true functional forms of the demand equations remain unknown, Thomas (1987).

Each of the preceding problems was finally eliminated through the use of duality. The advantages associated with duality are numerous, but for the purposes of evaluating consumer behaviour, the primary advantages are due to Roy's Identity and Shepard's Lemma.[2] Houthakker's (1960) Indirect Addilog model was the first to employ the techniques from duality. Although Houthakker's specification highlighted the opportunities associated with duality theory, the model failed to significantly address the disadvantages of the LES and Rotterdam models Thomas

---

[2] As shown below, these two duality results can be used to specify any number of rather intuitive demand equations, from the simple specification of a solution to a consumer's rational decision-making problem.

(1987). However, a combination of the LES and Rotterdam models was more successful.

The need to overcome the LES and Rotterdam model disadvantages prompted a new search for a solution. The search attempted to combine the disparate approaches of the LES and Rotterdam models, and resulted in the development of specifications that approximated indirect utility and cost functions with flexible functional forms, Cooper & McLaren (1992). The most notable of these are Diewert's (1971) Generalised Leontief and the Translog model of Christensen, Jorgenson & Lau (1975), which use a flexible functional form to approximate any indirect utility function. While these models were more general and did produce meaningful parameter estimates, they possess limited regularity properties, i.e., some of the restrictions from consumer theory could still be violated Cooper & McLaren (1992); furthermore, the models can only be used for the consideration of convex consumer preference orderings Thomas (1987).

The most recent development used flexible forms within the expenditure function rather than in the indirect utility functions, Deaton & Muellbauer (1980); the result of Deaton and Muellbauer's work is referred to as the Almost Ideal Demand System (AIDS). Based on the duality of consumer decision-making, their model yields an arbitrary approximation of any system of demand equations (even non-convex consumer preferences) without violating any axioms of consumer choice. More importantly, the model perfectly aggregates any number of consumers without relying on the assumption of parallel Engel curves.[3] The primary benefit of the AID System is the simplicity of estimation, as it largely avoids the need for non-linear estimation.[4] In addition, the functional forms, which describe the AIDS model, lend themselves to estimation with household budget data. However, since the Deaton

---

[3] This refers to the aggregation of individual household budget share equations into a single budget share equation. This single budget share equation was, essentially, derived from the cost function of some 'representative' household that has maximized its utility (Thomas, 1987). The data used here is for individuals in single-person households; therefore, this aggregation property is irrelevant, although it is extremely important when dealing with aggregated time series data.

[4] Although the AID System is not exactly linear, the non-linearity in the price aggregator function can be eliminated through the use of a price index. For further discussion, see Deaton & Muellbauer (1992).

and Muellbauer system relies on a Translog specification to describe one of the functions (specifically, the aggregator function) it inherits the regularity problems of the Translog. As Cooper & McLaren (1992) demonstrate, demand equations based on the AIDS model are guaranteed to exhibit irregularity, especially with regard to the negativity restrictions, as real expenditure rises.

Therefore, in an effort to exploit all the desirable properties of the AID System in the estimation of demand equations, Cooper & McLaren (1992) developed an extension, which they (and others) applied to Australian data. Their model, termed the Modified Almost Ideal Demand System (MAIDS), enjoys improved regularity properties over a wider expenditure-price space than the AIDS model. Thus, the likelihood of the concavity, or negativity, restrictions being violated is profoundly reduced, although not eliminated. A discussion of their model is provided in the following section.

3. Demand Systems: A Theoretical and Empirical Approach to Analysis

3.1. **Theoretical Underpinnings.** The following analysis is based on the assumption that households are homogeneous. Although South African households are far from identical, the empirical analysis will use homogenised data, where household samples are as similar as is feasible.[5] The proposed model, based upon a set of models used to analyse Australian household data, abstracts from within household allocations, partly due to expedience, as data is not available on within-household expenditures and partly due to the fact that the household data used in the analysis contains only one individual.[6]

Consider a household, an individual in this case, that minimises expenditures on all possible commodities (denoted by j) subject to receiving a specific level of utility from that consumption. The expenditure function, resulting from the cost

---

[5] The data in the empirical analysis is from single-person households, which are likely to be more homogeneous than other household formations; the single-person data can be further homogenized by race, employment status, and home location.

[6] However, Browning & Chiappori (1998) provide an excellent recent example of intra-household allocations.

minimisation, for each household, is assumed to be:

$$\ln\ c\left(r,u\right) = \alpha\left(r\right) + u \cdot \frac{\beta(r)}{(c(r,u))^{\eta}}$$

where

(1)
$$\alpha\left(r\right) = \alpha_0 + \sum_j \alpha_j \ln r_j + \frac{1}{2} \sum_k \sum_\ell \alpha_{k\ell} \ln r_k \ln r_\ell$$

and

$$\beta\left(r\right) = \beta_0 \prod_j r_j^{\beta_j}.$$

In the above specification, c represents the expenditure function, which is assumed to be a function of the prices of all consumed products (denoted by r) and a specific level of utility (denoted by u). All Greek letters denote parameters of the model. The Almost Ideal Demand System (AIDS) model proposed by Deaton & Muellbauer (1980) assumes $\eta = 0$, while Cooper & McLaren's (1992) model assumes $0 < \eta \leqslant 1$.

Duality theory makes it possible to determine the indirect utility function for the household's alcohol and tobacco consumption by rearranging equation (1). Denoting actual expenditure on commodities with Y, the indirect utility function can be written as:

(2)
$$V(r,Y) = \left[\ln Y - \alpha\left(r\right)\right] \cdot \frac{Y^{\eta}}{\beta\left(r\right)}.$$

From the indirect utility function, by applying Roy's Identity[7] and the definition of the expenditure share[8], the expenditure share for the jth commodity (denoted by w_j) will be:

(3)
$$w_j = \frac{\alpha_j + \beta_j \left[\ln Y - \alpha\left(r\right)\right]}{1 + \eta \left[\ln Y - \alpha\left(r\right)\right]}.$$

This implied system of share equations is referred to as the Modified Almost Ideal Demand System (MAIDS), since it is a modification to the Deaton & Muellbauer (1980) Almost Ideal Demand System. The AID System was originally developed to

---

[7] With the Deaton and Mullbauer model, it is easier to leave the solution as a cost/expenditure function, and create demands using Shephard's Lemma.

[8] In the case of indirect utility, applying an elasticity version of Roy's Identity (the opposite of the elasticity of indirect utility with respect to price over the elasticity of indirect utility with respect to income) yields the expenditure shares (see Appendix A.1). Exact details of the derivation are available from the authors upon request.

put appropriate theoretical structure to the Working (1943) and Leser (1963) share functions. The AID System, or Working-Leser share functions can be written as, because $\eta = 0$ is assumed:

$$(4) \qquad w_j = \alpha_j + \beta_j \left[\ln Y - \alpha(r) \right].$$

Since, by construction, all expenditure shares must sum to one. Aggregation in the MAID System occurs if

$$(5) \qquad \sum_j \alpha_j = 1 \qquad \text{and} \qquad \sum_j \beta_j = \eta.$$

Under the AID System, however, the expenditure shares sum to unity if

$$(6) \qquad \sum_j \alpha_j = 1 \qquad \text{and} \qquad \sum_j \beta_j = 0.$$

As can be seen by the similarity of restrictions, the modification is not extensive. Furthermore, with repeated cross-sectional data or time series data, it is possible to estimate and identify the value of η. Currently, however, the focus of this research is only on the applicability of these demand systems for South African data. For that reason, no attempt is made to estimate η;[9] rather, specific values are used (see discussion below).

3.2. **Empirical Specification.** The system of equations represented by (3) is the basis for the empirical analyses reported in this paper.[10] However, the data employed in the analysis does not contain any prices.[11] For that reason, the analysis will include a normalising quantity, taken as minimum total expenditure, in line

---

[9] Within a cross-section, a system of demand equations cannot be estimated (identified) unless one of the equations is first removed the system. The remaining parameters are then 'estimated' from the exclusion restriction; in this case, the restriction is based on the adding-up properties of the system. Attempting to estimate η is based on partial exclusion; only part of the additional equation is removed from the system, and, therefore, the estimates are not reliable within a single cross-section (see Section 3.2).

[10] In addition, system (4) is estimated as a special case by assuming $\eta = 0$.

[11] Although prices are available, since the Income and Expenditure Surveys are used for the creation of the South African Consumer Price Index, the prices are only representative in urban areas.

with Fry *et al* (2000), which is mathematically isomorphic to the price function contained in equation (3). Replacing the price function with the normalising quantity, $\ln K$, and adding a stochastic component leads to:

$$(7) \qquad w'_j = \frac{\alpha_j + \beta_j \left[\ln Y - \ln K\right]}{1 + \eta \left[\ln Y - \ln K\right]} + \omega_j, \ j = \{1, 2, ..., N\}.$$

Unfortunately there are problems with the stochastic terms in (7). Estimating (7) as a non-linear system is not possible, because the system of equations is singular. Due to the fact that the shares sum to unity, the covariance matrix will also be singular.[12] The typical solution to the singularity problem is the elimination of one of the equations in the system, i.e., estimating the remaining equations as a system using the adding up restrictions in (5) to create estimates for the eliminated equation.

In addition to the problems surrounding singularity, it is important to recall that these shares are expected to remain within the unit simplex.[13] For that reason, each stochastic element must also remain within the unit simplex. Fortunately, as long as the adding-up restrictions are true, the equations in (7) will force each individual share to fall within the unit simplex, and, therefore, the stochastic terms will also fall inside the unit simplex. However, econometric models rarely assume that the stochastic component of the model lies within the unit simplex.

In the empirical analysis conducted in this paper, rather than eliminating an equation in the empirical analysis, the equation that would have been dropped is used as a reference equation, thus creating a ratio of shares; although the equation is not dropped, there is still one less estimable equation in the system. A further modification, taking the log of the share ratio, eliminates the unit simplex stochastic

---

[12] Further discussion of the problems associated with the estimation of demand systems can be found in Greene (2003).

[13] A share cannot be negative, since expenditure on any commodity cannot be negative. In addition, a share cannot exceed one, since expenditure for any one item cannot possibly exceed total expenditure.

component. The MAID System of equations to be estimated becomes:

$$(8) \qquad \ln\left(\frac{w_i}{w_N}\right) = \ln\left(\frac{\alpha_i + \beta_i \ln\left(\frac{Y}{K}\right)}{\alpha_N + \beta_N \ln\left(\frac{Y}{K}\right)}\right) + \nu_i, \ i = 1, 2, ..., N-1,$$

N denotes the commodity that would have been dropped from the system, while $\alpha_N = 1 - \sum_{i=1}^{N} \alpha_i$ and $\beta_N = \eta - \sum_{i=1}^{N} \beta_i$ are imposed in order to eliminate the singularity problem. The ratios of the deterministic components stemming from consumer choice theory are also logged. The preceding specification has been developed to deal with the issue of singularity and also assumes that v_i is part of a multivariate log-normal stochastic distribution, as suggested by Fry *et al* (2000). Equation (8) has been applied with some success over Australian data;[14] however, the validity of the assumption underlying the stochastic component was not discussed in those analyses. In the case of South African data, the assumption may not be valid, which has profound effects on the estimates and the interpretation of these estimates.

Although equation (8) has been successfully implemented with data from developed economies, those implementations required one additional modification, which is also made here. Importantly, there is no reason to believe that all individuals will choose a positive level of expenditure on every product. For that reason, a large number of zeroes will exist in the data set. Given the log-ratio form of the dependent variable, there are two problems: it is not possible to take the natural log of zero, nor is it possible to divide by zero. Therefore, another approach must be followed.

Economically, zero shares contain information. In the estimating equation, however, these zeros cause problems. The removal of a zero share from the estimating sample is tantamount to ignoring relevant information regarding those zeros, and may be impractical if a large portion of the sample contains zeros, Fry et al. (2000). Zero expenditure signals a conscious decision by a consumer to not enter into the

---

[14] See Cooper & McLaren (1992), McLaren, Fry & Fry (1995), Fry, Fry & McLaren (1996) Fry, Fry & McLaren (2000), and Fry, Fry, McLaren & Smith (2001) for examples.

market for a particular commodity, and, consequently, needs to be recognised when conducting any demand analysis. Although there are many modelling techniques to address zeroes in the data, including Box-Cox transformations of the data and Tobit regressions, compositional data analysis techniques are econometrically less costly without a significant shortfall in benefits.[15] Furthermore, compositional techniques have been successfully applied to Australian household data, and, since the focus in the paper is on determining the validity of the technique for South African data, the technique is applied here.

The compositional analysis employed in this paper is an adaptation of Aitchison's (1986) zero (or trace) replacement. The modification, suggested by Fry et al. (2000), is meant to guarantee that the replacement of the zeros does not distort the ratios of the non-zero shares. This replacement technique supposes that a composition (in this case, a household) has M zero and $P - -M$non-zero components (budget shares). The zeros are replaced by:

$$(9) \qquad \tau_A = \delta(M + 1)/P^2,$$

while the nonzero shares are reduced by

$$(10) \qquad w_i \times \tau_S,$$

where

$$(11) \qquad \tau_S = \delta M(M + 1)/P^2$$

and δ is a chosen maximum rounding error. In order to apply this technique, sensible minimum and maximum values for the zero replacement τ_A are determined by the ratios: $0.01/$(maximum total expenditure) and $0.01/$(minimum total expenditure), respectively. Equation (9) is then solved for δ, allowing for the calculation of τ_S

---

[15] As can be seen in equation (8), the system is already non-linear. An additional non-linearity from the Box-Cox transformation would further complicate the estimation of the system.

in equation (11). The results reported in this paper are those obtained using the maximum zero replacement values.[16]

3.3. **Other Considerations.** As suggested earlier, the share ratios analysed in this paper do not conform to the lognormal distributional assumption, which is the basis for the stochastic lognormal distributional assumption. However, the primary reason for this conformity is the fact that only a limited number of commodities are analysed. For this analysis only four-commodity and six-commodity systems are considered. Extending the number of commodities beyond six leads to a potential to violate the distributional assumptions; in fact, one of the share ratios is consistently bimodal.[17] For nine and sixteen commodities, nearly half of the distributions are bimodal, even in very homogeneous samples. Although bimodality of total expenditure in a country such as South Africa is not surprising, there is no obvious reason to suspect that expenditure shares will have a bimodal distribution, or that the resulting share ratios will also have bimodal distributions. The potential effect of the bimodal distributions occurs in Section 4, however, the intuition is rather clear. A bimodal distribution implies two different populations, which may have very different behavioural properties. For that reason, estimates could be biased and not just imprecisely estimated, as is the case when variance assumptions are violated.

Beyond the potential for bias resulting from a distributional assumption violation, the 2000 South African Income and Expenditure Survey (SAIES) is known to have problems (see for example, Burger, van der Berg, and Nieftagodien, 2004). Either surveyed individuals are reporting incorrectly, the surveyors are recording incorrectly, or the data enterers are transcribing the data incorrectly. Whether or not the errors are random or, worse, are related to the survey itself, the result

---

[16] Fry *et al* (2000) show that the analysis is robust to the size of the zero replacement, and so I do not consider the differences in results that would obtain under different specifications of the replacement value.

[17] Although some of the share ratio distributions considered in the six-commodity systems appear to contain a second mode, and the effects of that will be discussed below, six-commodity systems were included in the analysis. Further research on dealing with the bimodality of the distributions is currently being undertaken.

of measurement, recording, or transcribing errors is biased estimates. The typical solution to a measurement error problem is locating an instrument to replace the incorrect variable. Importantly, however, the instruments must be appropriately related to the original variable of interest, and, more importantly, the instruments must not also be measured with error. If either of these conditions is not met, the instrumental variables estimate will not improve, and is likely to worsen, the measurement error bias. Although it is possible to create an instrument for total expenditure (the only right hand side variable in the model) by using demographic and dwelling variables that are unlikely to be measured with error, the demographic and dwelling variables are qualitative variables and cannot provide the appropriate level of continuity to provide useful budget share data. For that reason, instrumental variables were not used in this analysis. Rather, in order to eliminate some of the most egregious measurement error problems, individuals whose total expenditure was recorded as zero, when there household income was recorded as positive (especially if individual expenditure categories were recorded as positive) were eliminated from the data.

4. The Data

4.1. **The Data Source.** The data set used in this demand system analysis was obtained from the SAIES published by Statistics South Africa (StatsSA), for the year 2000. The SAIES is a comprehensive appraisal, as the name implies, of the income and expenditure patterns of households in South Africa. Using a master sample, based on the 1996 Population Census, 30 000 households were identified as sampling units, StatsSA (2000). The sampling methodology employed ensures that the households are chosen to represent the diverse demographic features of the South African population. The 2000 SAIES was conducted in October 2000, and included households throughout the country, from the thirteen historical metropolitan areas,[18] as well as other urban and rural areas StatsSA (2000).

---

[18] Cape Peninsula, Port Elizabeth/Uitenhage, East London, Kimberley, Bloemfontein, Free State Goldfields, Durban/Pinetown, Pietermaritzburg, Klerksdorp/Stilfontein/Orkney, Pretoria/Centurion/Akasia, Witwatersrand, Nelspruit/Witbank and Polokwane.

The main purpose of the SAIES is to determine the weights StatsSA uses in the compilation of its various consumer price indices (CPI and CPI-X, for example). However, apart from being a rich source of information concerning the income and expenditure capacity of South African households, as well as consumption patterns of households, the 2000 IES also contains data pertaining to demographics, development and a rough proxy for employment. Therefore, this data is a potentially fruitful source of analysis, despite many of the problems with the data, as reported by various authors.[19]

4.2. **Data Summary.** Before fitting the model to the data, a preliminary data analysis was conducted to ascertain the appropriateness and adequacy of the proposed techniques, as well as to enable a better understanding of any empirical results obtained. With this in mind, Table 1 provides the summary expenditure, income and budget share statistics for the two different groups of analysis and two population groups.

From Table 1, it is evident that the South African black population, living in single-person households, earns and spends less than their white South African counterparts. The data also shows that the, on average, poorer black households spend a larger proportion of their income on food and clothing, but a smaller proportion on housing and other goods. If the additional (other) expenditures are further disaggregated, we can see that white single-person households spend a larger proportion of their income on human capital (health and education) and public or private transportation than do black single-person households.

Another finding apparent in Table 1 is that the standard deviations of the data are extremely large, often very similar to or larger than their associated means. In other words, the data covers an exceptionally wide expenditure-price region, which further strengthens our case for using the MAIDS specification with its attendant improved regularity properties, Cooper & McLaren (1992) and Crawford, Laisney & Preston (2003). In the initial 4-commodity breakdown, there are very few

---

[19]See for example, Burger, van der Berg & Nieftagodien (2004).

zero expenditure shares for either population sub-group. However, if the catchall category is further disaggregated, a much larger proportion of zero shares are introduced, up to 31% of single-person black households did not spend (in 2000) on either education or health, for example.

Importantly, the empirical application also assumes that the underlying stochastic disturbance is multivariate lognormal, partly due to the fact that each individual share is likely to be approximately half-normal over the unit simplex, i.e., a larger proportion of small share values and a very small proportion of large share values are likely to be observed. In fact, Figures 1 and 2, for the four-commodity group and six-commodity group, respectively, tend to bear out this assumption. As can be seen in each of the figures, there are many more observed small share values than large share values, and in the case of food and clothing, in both cases, as well as human capital and transportation in the six-commodity case, the observed density follows a hyperbolic rate of decrease across the unit simplex.

Given the observations in Figures 1 and 2, the assumption of share ratio normality appeared to be plausible, and, therefore, a histogram for each of the ratio of the logged expenditure shares was calculated and graphed. The four-commodity group histograms are plotted in Figure 3, while the six-commodity group histograms are plotted in Figure 5; Figure 4 contains the observed density of the natural log of total expenditure in the sample. As can be seen in each of these plots, and can be verified by Jarque-Bera statistics[20], the distributions are not normal, tending to be skewed or suggestive of bimodalities. Although it is possible that the non-normality in the share ratios is offset by the non-normality of expenditures, so that the resulting residuals remained normal, a simple examination of the distributions cannot be conclusive on its own. However, given the differences in the distributions, we should not expect the stochastic disturbances in the underlying analysis to be normally distributed, and, in fact, as shown in the next section, they are not. The impact of violating the normality assumption is further discussed in the next section.

---

[20] These statistics for any of these distributions are available from the author upon request.

5. The Results

The MAIDS specification identified in equation (5) was used to explore the demand characteristics of different commodity groups in South Africa. The expenditure shares considered were those of food, housing, clothing, and other expenditures; additional analysis further disaggregated other expenditures into human capital (health and education), transportation (public and private) and a remainder category (see Table 1).

The estimation technique employed to estimate the system of demand equations was full information maximum likelihood (FIML), which was carried out using Eviews 5.0's SYSTEM object. As discussed earlier, one equation from the system (in this case it was the clothing equation, since it was the only one without zero shares) was used as the reference equation in order to render the system estimable. The parameter estimates and standard errors of the estimated parameters for the reference equation were then found by using the adding-up constraints.[21]

5.1. **The Four-Commodity System.** The results of the four-commodity system estimates for white and black single-person households are presented in Table 2. Given the differences in summary statistics between the two race subgroups, it is not surprising that there are differences in the estimated parameters. Regardless of the value of η that is chosen (although only the extreme values of 0 and 1 are chosen[22]), the estimated β parameters for food are larger for the black population, while the estimated β parameters for other goods and clothing are larger for the white population (budget share elasticities will be reported below).

According to Fry et al. (2000), sufficient conditions (when $\eta > 0$) for the regularity of the underlying indirect utility function are that α_i and β_i must be greater than or equal to zero, for all i (when $\eta > 0$). As is evident from the results in Table 2, these conditions are met across all the $\eta = 1$ equations for all African households,

[21] To check the sensitivity of our results to the choice of 'dropped' equation, the system was estimated with other equations as the denominator for all log-ratios, instead. The results, as expected, were not significantly affected by the choice of 'dropped' equation, although some of the estimated parameters were not exactly the same, a result common in non-linear estimation.

[22] Estimates for other values are available from the author upon request.

but not entirely for the white households. The irregularity in the white population could be due to the fact that the non-linear estimates are calculated with a limited number of observations. Another interpretation of the large negative for food share intercept term, although not completely reasonable, is that white households do not purchase food items, unless their total expenditure share is large enough.[23]

Although regularity conditions are not entirely invalidated by the estimates, a different problem exists regarding the interpretation of the results: the residuals are heteroskedasticy, and are, furthermore, as was expected (see Figure 6), not normally distributed.[24] In the normal case of heteroskedasticity, its impact is only upon the efficiency of the estimates. However, in this model, heteroskedasticity is combined with distributional bimodality. The result is a set of estimates that are certainly inefficiently estimated and could be inconsistently estimated as well.[25]

5.2. **The Six-Commodity System.** The parameter estimates from the six-commodity demand system are reported in Table 3. The results are very similar to those reported in Table 2. The estimated β parameters for food are larger for the African households, while the estimated β parameters for human capital, transport, and other goods are larger for the white households. Furthermore, the inclusion of additional share categories affects the estimates for single-person households in both the MAID and the AID System. For African households, the β parameter for food and clothing is larger, while the β estimates for housing and other goods are lower, with the addition reorganization of commodity aggregates. For single-person white households, the β estimates for other goods and clothing are always lower, while the

---

[23] Solving the equation, $w_f = 0 = \alpha_f + \beta_f \ln(y/989)$, results in $y = 28725$.

[24] The White Test reveals $nR_1^2 = 9.37$, $nR_2^2 = 11.0$, and $nR_3^2 = 19.52$ (where the subscripts denote the share ratio considered), while the Jarque-Bera statistic is 606.1,1328.4, and 3594.9, for the food to clothing share residual, the housing to clothing share residual, and the other goods to clothing share, respectively. All of these values result in rejection of the null hypotheses of no heteroskedasticity and normality, respectively, although the food to clothing share rejection of no heteroskedasticity is by a small margin. Similar tests were conducted for the case with $\eta = 1$, and the results were quantitatively the same, although the rejection margin was wider. Tests for the sample of white households resulted in similar findings, although the findings may have less meaning, given the irregularity of the MAID System.

[25] Further research will be undertaken to correct for these problems, however, it is believed that the solution requires a complete revision of the model, and not a simple weighted least squares parametric fix.

estimates for housing are always larger. In addition, the white household estimates point to irregularities in the system.

The observed estimated differences between the four-commodity and six-commodity estimates points to two issues. First, the utility function is not separable: the human capital, transportation, and remainder components in the six-commodity system do not represent the other component in the four-commodity system; the result can also be seen from the fact that the food, housing, and clothing estimates do not remain constant. Second, the distributional features of the data may have affected the results. In both the African and white populations, human capital and transportation share ratios with respect to clothing expenditure are bimodal (see Figure 5 for the African household data). Therefore, it is possible that the six-commodity estimates are biased. Even if the results are not biased, they are not consistent, because the residuals are not normally distributed and they are not homoskedastic.[26] The distributions of a selection of the residuals from the five-equation system are illustrated in Figure 7. As with the four-commodity group, it is uncertain whether the six-commodity results should be completely trusted, as they are presented.

5.3. **Budget Share Elasticities.** Not much discussion was presented regarding the interpretation of the results in Tables 2 and 3, partly because of the underlying estimation problems and partly because the comparative static exercises are not as obvious as a quick look at the parameter estimates imply. Given the assumption that heteroskedasticity only affects estimate efficiency, the primary interpretational issue in the model is related to the non-linearity of the system. In order to improve the intuitiveness of the comparisons, expenditure elasticities were calculated.[27] For

---

[26] The White Test reveals $nR_1^2 = 28.9$, $nR_2^2 = 45.7$, $nR_3^2 = 16.3$, $nR_4^2 = 59.8$, and $nR_5^2 = 144.8$ (where subscripts denote the share ratio), while the Jarque-Bera statistic is 74.0, 256.6, 225.4, 93.9, and 205.7 for the food to clothing share residual, the housing to clothing share residual, the human capital to clothing share, the transportation to clothing share, and the other goods to clothing share, respectively. All of these values result in rejection of the null hypotheses of no heteroskedasticity and normality, respectively. Similar tests were conducted for the case with $\eta = 1$, and the results were quantitatively the same. Similar analyses were conducted for the white households, with similar conclusions.

[27] See Appendix B for a derivation of the elasticities.

this analysis (see Appendix B), a commodity is a luxury if the expenditure elasticity exceeds unity and the commodity is a necessity if the elasticity is less than one. The elasticities are provided in Table 4.

For African households, the results are generally reasonable, if not slightly surprising. Regardless of the number of commodity groups, only necessity was clothing, while food, housing, other goods, transportation, and human capital investments were all luxuries. Given the fact that housing, food, and clothing are generally thought to be necessities, the results do not conform to economic theory. For the white households, the numbers were not consistent across estimates, another reason to be wary of the model, as it is applied to the data. However, in the four-commodity AID System, the elasticities do agree with economic theory, so that food, housing, and clothing are necessities, while other goods are luxuries. The four-commodity elasticities are rather similar regardless of whether the calculations were based on the AIDS or MAIDS models. The difficulty surrounding the white households can probably be attributed to the irregularities uncovered in the estimates, while the disagreement with economic theory exhibited by African households is likely to obtain from misspecification, especially related to the existence of bimodality.

6. Conclusion

Results from demand system estimates using data from the 2000 South African Income and Expenditure Survey were presented in this paper. The purpose of the analysis was to determine whether or not a model that has been reasonably successful in developed country contexts can be applied to South African data. The theoretical model used to develop the empirical model did not impose many restrictions, and was underpinned by the usual set of assumptions, especially regarding normality. More specifically proposed model provided a rather simplistic way of dealing with data that can only be observed in the unit simplex, and contains a large number of zeroes. Although there was a cost associated with the non-linear

estimation procedure, because it is time consuming, the benefit of the approach is that the budget share properties are necessarily met.

However, as was shown in the paper, even the simplest of assumptions can often fail in the case of South African data. In particular, the derived Jarque-Bera statistics failed to accept the null hypothesis of normality in all cases (although simple eye-balling of the data is convincing on its own), despite the fact that the sample sizes were very large, in excess of 4000 in the case of African households. The result of one feature of the non-normality was estimates that were inefficient, due to the presence of heteroskedasticity. Another result of the non-normality, which can be attributed to the bimodality, is the potential for inconsistent estimates.

The research presented here points to the need for research conducted with South African data to be analysed with techniques appropriate to the data. The research has shown that applying a model, which had been developed for the analysis of rich country data, may not lead to very useful results. The research has also shown that much more work must be done in this area to develop models, theoretical and empirical, which will more appropriately fit the needs of economic research in South Africa.

References

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman and Hall, Ltd.

Browning, M. & Chiappori, P. (1998), 'Efficient intra-household allocations: A general characterization and empirical tests', *Econometrica* **66**(6), 1241–1278.

Burger, R., van der Berg, S. & Nieftagodien, S. (2004), Consumption patterns of South Africa's black middle-class: Correcting for measurement errors. Paper delivered at the conference of the Centre for the Study of African Economies on Poverty Reduction, Growth, and Human Development in Africa; Oxford, UK.

Christensen, L., Jorgenson, D. & Lau, L. (1975), 'Transcendetal logarithmic utility functions', *American Economic Review* **65**, 367–385.

Cooper, R. & McLaren, K. (1992), 'An empirically oriented demand system with improved regularity properties', *Canadian Journal of Economics* **25**, 652–667.

Crawford, I., Laisney, F. & Preston, I. (2003), 'Estimation of household demand systems with theoretically compatible Engle curves and unit value specifications', *Journal of Econometrics* **114**, 221–241.

Deaton, A. & Muellbauer, J. (1980), 'An almost ideal demand system', *American Economic Review* **70**(3), 312–326.

Deaton, A. & Muellbauer, J. (1992), *Economic and Consumer Behavior*, Cambridge University Press, Cambridge.

Diewert, W. E. (1971), 'An application of the Shephard duality theorem: A generalized leontief production function', *Journal of Political Economy* **79**, 481–507.

Fry, J., Fry, T. & McLaren, K. (1996), 'The stochastic specification of demand share equations: Restricting budget shares to the unit simplex', *Journal of Econometrics* **73**, 377–385.

Fry, J., Fry, T. & McLaren, K. (2000), 'Compositional data analysis and zeros in micro data', *Applied Economics* **32**, 953–959.

Fry, J., Fry, T., McLaren, K. & Smith, T. (2001), 'Modelling zeroes in microdata', *Applied Economics* **33**, 383–392.

Greene, W. (2003), *Econometric Analysis*, 5 edn, Prentice Hall, Upper Saddle River, NJ.

Houthakker, H. (1960), 'Additive preferences', *Econometrica* **28**, 244–257.

Leser, C. (1963), 'Forms of Engle functions', *Econometrica* **31**, 674–703.

McLaren, K., Fry, J. & Fry, T. (1995), 'A simple nested test of the almost ideal demand system', *Empirical Economics* **20**, 149–161.

Moore, H. (1914), *Economic Cycles: Their Law and Causes*, Macmillan, New York, NY.

StatsSA (2000), Income and expenditure of households 2000. Statistical Release P0111.

Stone, J. (1954a), *The Measurement of Consumer Expenditure and Behavior in the UK: 1920-38*, Vol. 1, Cambridge University Press, Cambridge, UK.

Stone, J. (1954b), 'Linear expenditure systems and demand analysis: An application to the pattern of British demand', *Economics Journal* **64**, 511–527.

Theil, H. (1965), 'The information approach to demand analysis', *Econometrica* **33**, 67–87.

Thomas, R. (1987), *Applied Demand Analysis*, Longman, New York.

Working, H. (1943), 'Statistical laws of family expenditure', *Journal of the American Statistical Association* **38**, 43–56.

APPENDIX A. DEVELOPMENT OF SHARE EQUATIONS VIA ROY'S IDENTITY

From Roy's Identity, we know

$$(A.1) \qquad x_\ell = \frac{-\partial v/\partial p_\ell}{\partial v/\partial y}.$$

The expenditure share is given by

$$(A.2) \qquad w_\ell = \frac{p_\ell}{y} x_\ell.$$

Combining equation (A.1) with equation (A.2) yields

$$(A.3) \qquad w_\ell = \frac{p_\ell}{y} x_\ell = \frac{-\partial v/\partial p_\ell}{\partial v/\partial y} \cdot \frac{p_\ell}{y} = \frac{-p_\ell \cdot \left(\partial v/\partial p_\ell\right)}{y \cdot \left(\partial v/\partial y\right)}$$

Dividing the numerator and the denominator by indirect utility results in the expenditure share, which can be derived from an elasticity version of Roy's Identity.

$$(A.4) \qquad w_\ell = \frac{p_\ell}{y} x_\ell = \frac{-\frac{p_\ell}{v(p,y)} \cdot \left(\partial v/\partial p_\ell\right)}{\frac{y}{v(p,y)} \cdot \left(\partial v/\partial y\right)} = \frac{-\partial \ln v/\partial \ln p_\ell}{\partial \ln v/\partial \ln y}$$

APPENDIX B. DEVELOPMENT OF AIDS AND MAIDS EXPENDITURE ELASTICITIES

Rearranging equation (A.2) yields, $p_\ell x_\ell = w_\ell y$, i.e., expenditure on a good is equal to the expenditure share times total expenditure. The elasticity of the expenditure on a good is then the log derivative of that expenditure with respect to logged total expenditure:

$$(B.1) \qquad \xi_i \equiv \frac{\partial \ln (p_\ell x_\ell)}{\partial \ln y} = \frac{\partial \ln (w_\ell y)}{\partial \ln y} = \frac{\partial \ln w_\ell}{\partial \ln y} + \frac{\partial \ln y}{\partial \ln y} = 1 + \frac{\partial \ln w_\ell}{\partial \ln y}$$

In the MAIDS model,

$$(B.2) \qquad w_\ell = \frac{\alpha_\ell + \beta_\ell \ln\left(Y/K\right)}{1 + \eta \ln\left(Y/K\right)}$$

Therefore, $\ln w_\ell = \ln\left[\alpha_\ell + \beta_\ell \ln\left(Y/K\right)\right] - \ln\left[1 + \eta \ln\left(Y/K\right)\right]$, and, most importantly,

$$(B.3) \qquad \frac{\partial \ln w_\ell}{\partial \ln y} = \frac{\beta_\ell}{\alpha_\ell + \beta_\ell \ln\left[Y/K\right]} - \frac{\eta}{1 + \eta \ln\left[Y/K\right]}$$

Equation (B.3) is the share elasticity. Substituting equation (B.3) into equation (B.1) yields the expenditure elasticity. In the AID System, the last term in equation (B.3) is zero; otherwise, the calculations can be made with the same equations.

TABLE 1. Summary Expenditure, Income and Budget Shares

	Single Person African Households		
	$n = 4488$		
Expenditure Category	*Mean*	*Std. Dev.*	*% Zeroes*
Food	0.1322	0.1233	0.8
Housing	0.2303	0.2069	6.1
Clothing	0.2358	0.2709	0.0
Other	0.4107	0.2194	5.1
Total Expenditure	9634	26788	
Household Income	15089	19642	
Other Disaggregated			
Human Capital	0.0555	0.1183	31.0
Transport	0.0945	0.1263	17.7
Remainder	0.2517	0.2605	17.3
	Single Person White Households		
	$n = 360$		
Expenditure Category	*Mean*	*Std. Dev.*	*% Zeroes*
Food	0.0392	0.0363	0.3
Housing	0.4304	0.2128	0.8
Clothing	0.1024	0.0856	0.0
Other	0.4280	0.2236	0.8
Total Expenditure	37674	42066	
Household Income	58258	70239	
Other Disaggregated			
Human Capital	0.1146	0.1386	7.2
Transport	0.1519	0.1659	18.3
Remainder	0.1615	0.1308	2.2

Source: Author's calculations from 2000 SAIES.
Note: First four expenditure categories used for the 4−commodity estimates.
First three and last three categories used for the 6−commodity estimates.

TABLE 2. Non-linear Systems Estimates for African and White Single-person Households: MAIDS and AIDS over Four Commodities

	Single Person Black Households (n=4248)				
Expenditure Category	$\eta=1$			$\eta=0$	
	α	β		α	β
Food	0.0057	0.1643^a		0.0240^a	0.0234^a
	(0.019)	(0.006)		(0.007)	(0.002)
Housing	0.0014	0.2283^a		0.0233	0.0274^a
	(0.032)	(0.012)		(0.011)	(0.003)
Other	0.0009	0.5321^a		0.0128^b	0.0370^a
	(0.027)	(0.016)		(0.009)	(0.002)
Clothing^{d,e}	0.9920	0.0753^a		0.9399^a	-0.0877^a
	(0.046)	(0.021)		(0.046)	(0.004)
	Log Likelihood = 1035.1			Log Likelihood = 570.5	
	Single Person White Households (n=358)				
Expenditure Category	$\eta=1$			$\eta=0$	
	α	β		α	β
Food	-0.4238^a	0.1258^a		0.0759^a	-0.0067^a
	(0.088)	(0.017)		(0.116)	(0.002)
Housing	0.6427	0.3272		1.0146^a	-0.0858^a
	(1.272)	(0.201)		(0.544)	(0.010)
Other	0.6020	0.4345^b		-0.4331^a	0.1290^a
	(1.370)	(0.215)		(0.396)	(0.010)
Clothing^{d,e}	0.1791	0.1126		0.3426	-0.0366^a
	(1.871)	(0.294)		(0.683)	(0.015)
	Log Likelihood =314.7			Log Likelihood = 441.8	

Source: Eviews 5.0 SYSTEM object. Note: Standard errors in parenthesis.
^{a} – Significant at 1%, ^{b} – Significant at 5%, ^{c} – Significant at 10%.
^{d} – Parameter estimates calculated from the linear restriction imposed on the model (See equations 5 and 6). ^{e} – Standard errors calculated from the linear restriction imposed for the estimation of the parameter.

TABLE 3. Non-linear Systems Estimates for African and White Single-person Households: MAIDS and AIDS over Six Commodities

Expenditure Category	Single Person Black Households (n=4482)			Single Person Black Households (n=4482)	
	$\eta=1$			$\eta=0$	
	α	β		α	β
Food	0.0072	0.1963^a		0.1694^a	0.0004
	(0.133)	(0.026)		(0.014)	(0.003)
Housing	0.0018	0.3035^a		0.1881^a	0.0140^a
	(0.232)	(0.060)		(0.021)	(0.004)
Human Capital	0.0002	0.0358^a		0.0254^a	0.0011
	(0.000)	(0.000)		(0.004)	(0.001)
Transport	0.0003	0.0986^a		0.0752^a	0.0020
	(0.012)	(0.000)		(0.010)	(0.002)
Other	0.0005	0.2333^a		0.0142	0.0319^a
	(0.024)	(0.008)		(0.012)	(0.003)
Clothing^{d,e}	0.9900	0.1326^b		0.5278^a	-0.0494^a
	(0.269)	(0.066)		(0.030)	(0.006)
	Log Likelihood = 21787.1			Log Likelihood = 22111.4	
	Single Person White Households (n=360)			Single Person White Households (n=360)	
Expenditure Category	$\eta=1$			$\eta=0$	
	α	β		α	β
Food	0.1291	0.0240		0.0631^a	-0.0038^a
	(0.144)	(0.024)		(0.0203)	(0.003)
Housing	0.6146	0.5078^a		0.8418	-0.0480^a
	(0.548)	(0.090)		(0.087)	(0.013)
Human Capital	-0.4991	0.1547^a		-0.0611	0.0195^b
	(0.310)	(0.051)		(0.043)	(0.007)
Transport	-0.6950^a	0.2099^a		-0.0886^a	0.0270^a
	(0.173)	(0.034)		(0.022)	(0.004)
Other	-1.2347^a	0.3752^a		-0.1462^a	0.0465^a
	(0.333)	(0.058)		(0.048)	(0.008)
Clothing^{d,e}	2.6852^a	-0.2717^a		0.3911^a	-0.0411^a
	(0.747)	(0.126)		(0.112)	(0.018)
	Log Likelihood =1580.9			Log Likelihood = 1578.1	

Source: Eviews 5.0 SYSTEM object. Note: Standard errors in parenthesis.
^{a} – Significant at 1%, ^{b} – Significant at 5%, ^{c} – Significant at 10%.
^{d} – Parameter estimates calculated from the linear restriction imposed on the model (See equations 5 and 6). ^{e} – Standard errors calculated from the linear restriction imposed for the estimation of the parameter.

TABLE 4. Expenditure Elasticities for Four-Commodity and Six-Commodity: MAIDS and AIDS Models

Expenditure Category	African Households (n=4482)		White Households (n=350)	
Four Commodities	$\eta=1$	$\eta=0$	$\eta=1$	$\eta=0$
Food	1.0496	1.1771	3.2876	0.8752
Housing	1.0816	1.1832	0.987	0.8826
Other	1.1775	1.2045	0.9761	1.0356
Clothing	0.8716	0.8128	0.9652	0.8335
Six Commodities	$\eta=1$	$\eta=0$	$\eta=1$	$\eta=0$
Food	1.0491	1.0023	0.8653	0.925
Housing	1.0829	1.0547	0.9863	0.9299
Human Capital	1.0852	1.0358	0.7418	1.0071
Transportation	1.1188	1.0307	14.1907	0.0143
Other	1.1497	1.0282	-2.9308	2.7207
Clothing	0.9036	0.5985	0.5945	0.5852
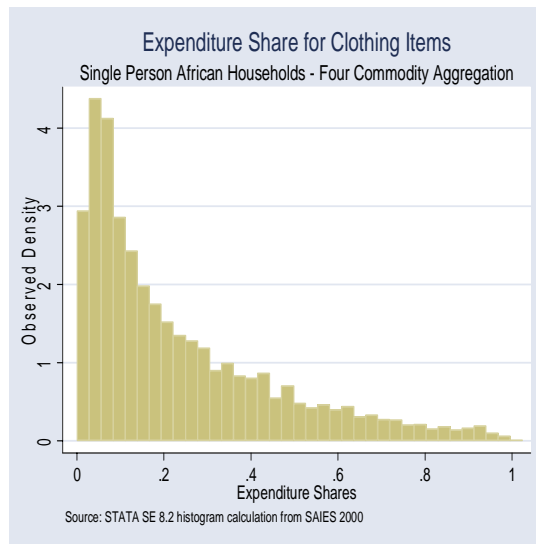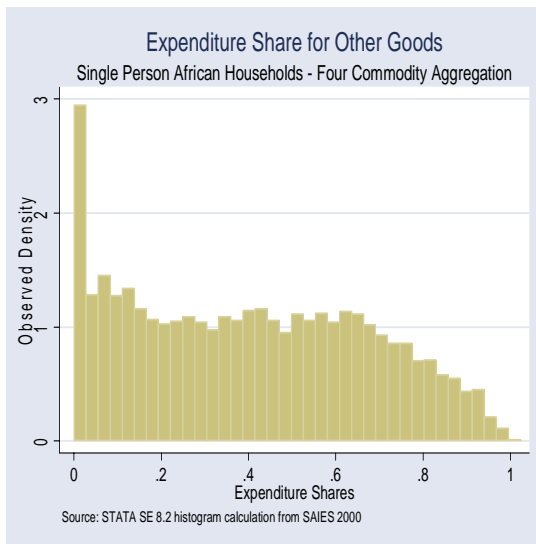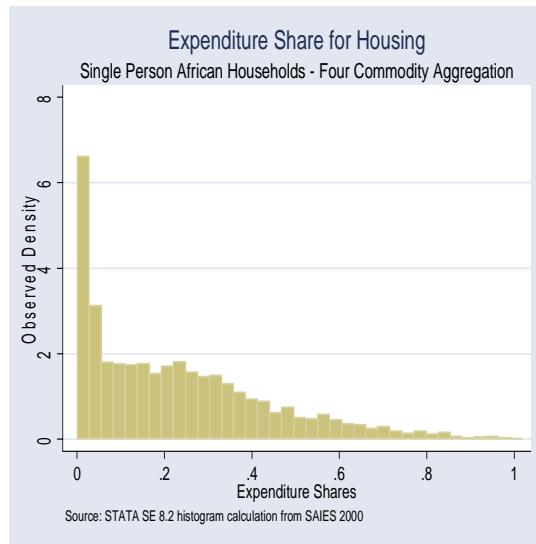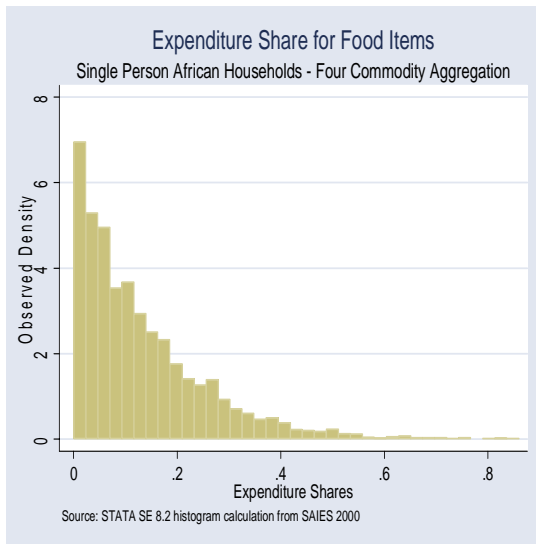
Figure 1. Single African Person Household Expenditure Shares for Four Commodity Groups
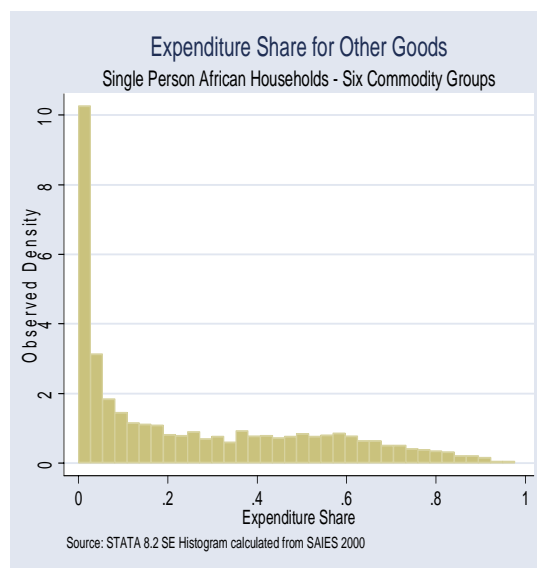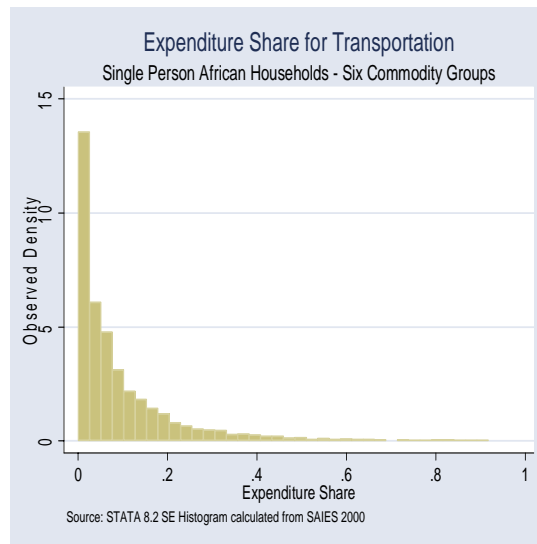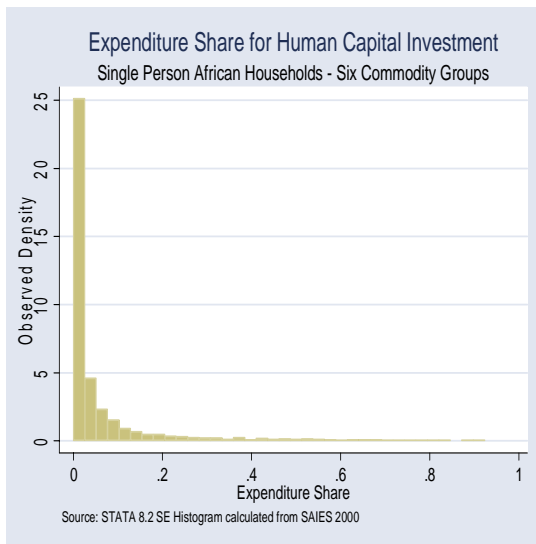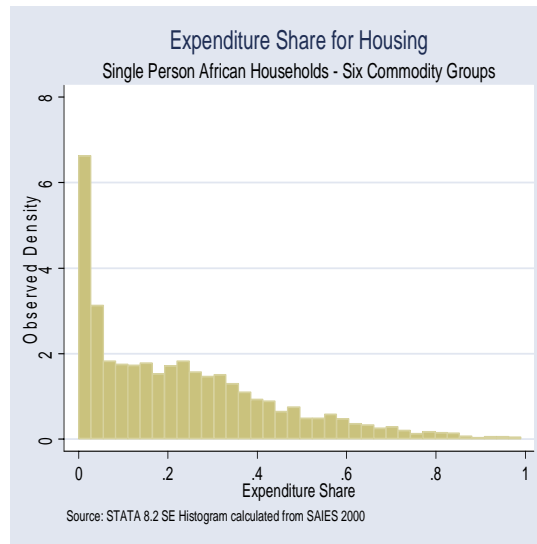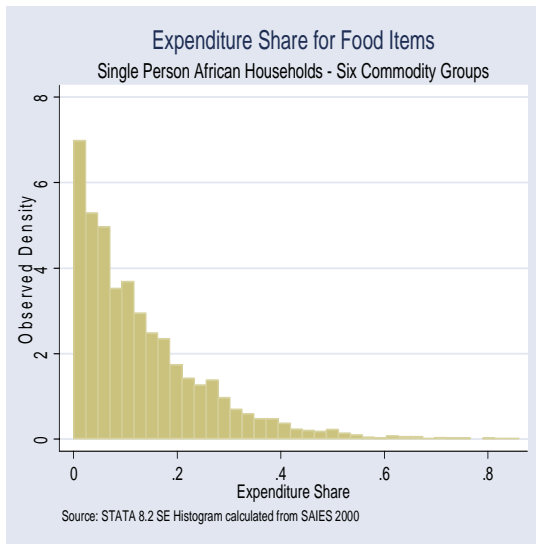
Figure 2. Single Person African Household Expenditure Shares for Six Commodity Groups
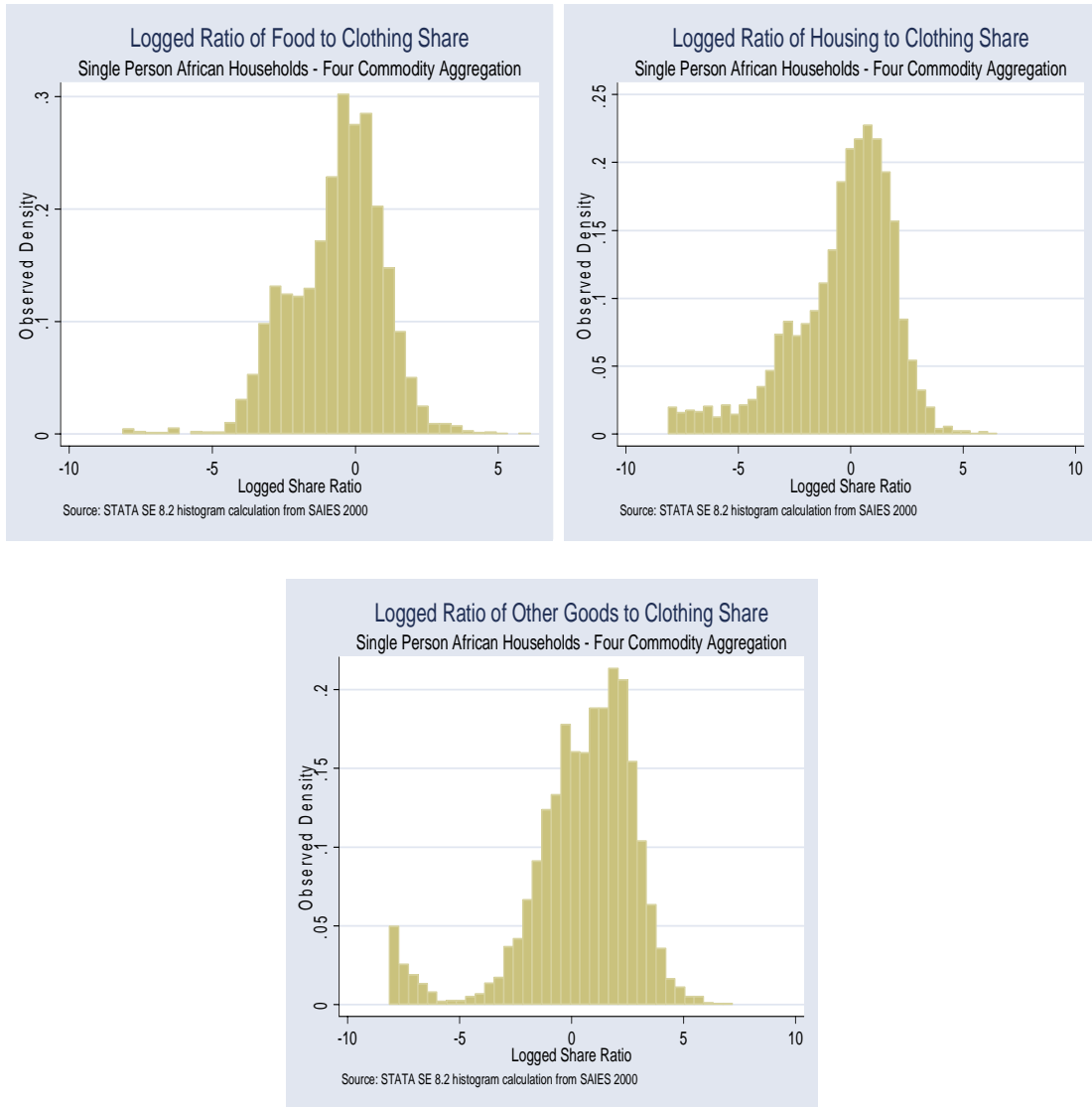
Figure 3. Four Commodity Group Logged Expenditure Share Ratios for Single Person African Households
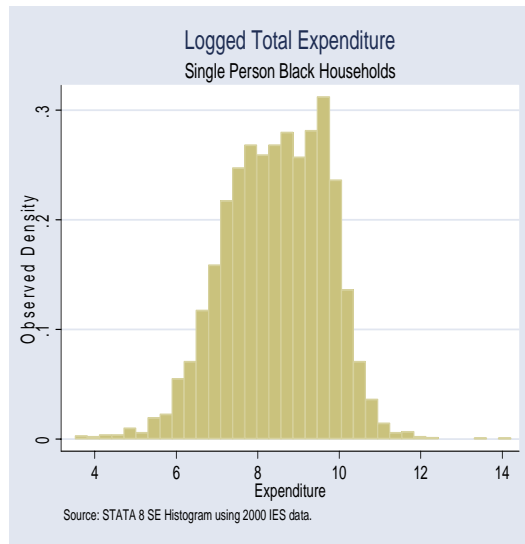
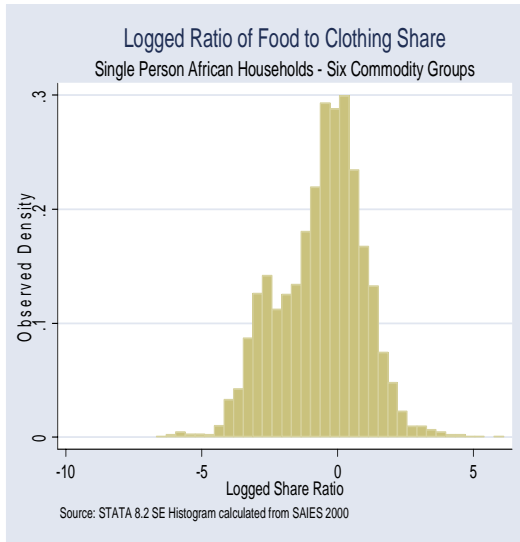**Figure 4. Logged Total Expenditure for Single-Person African Households**
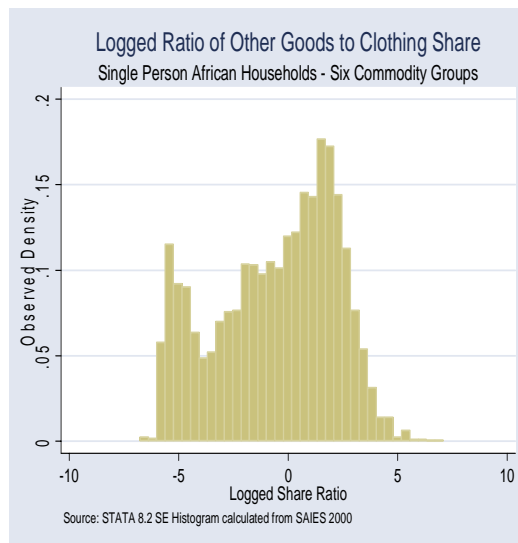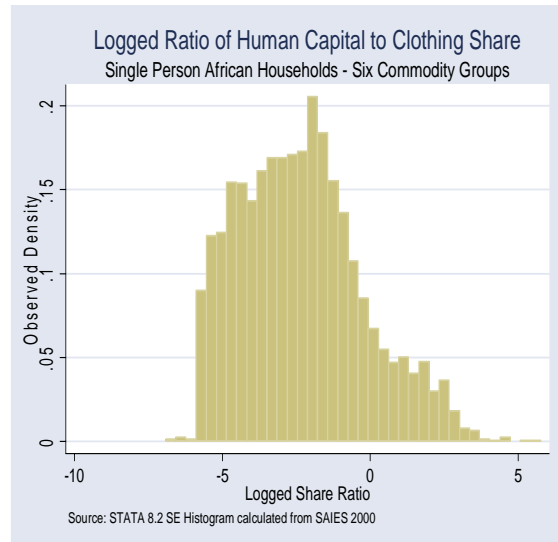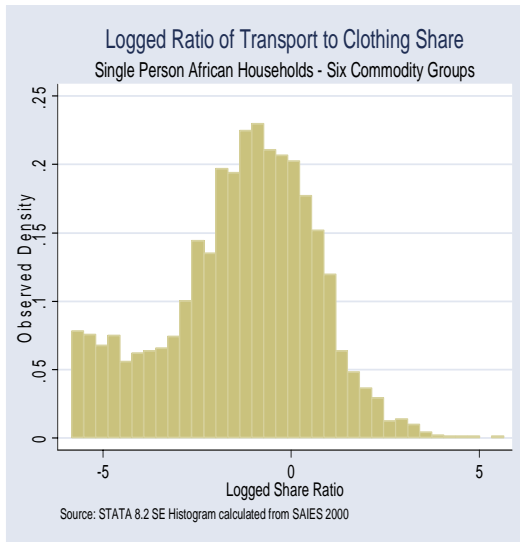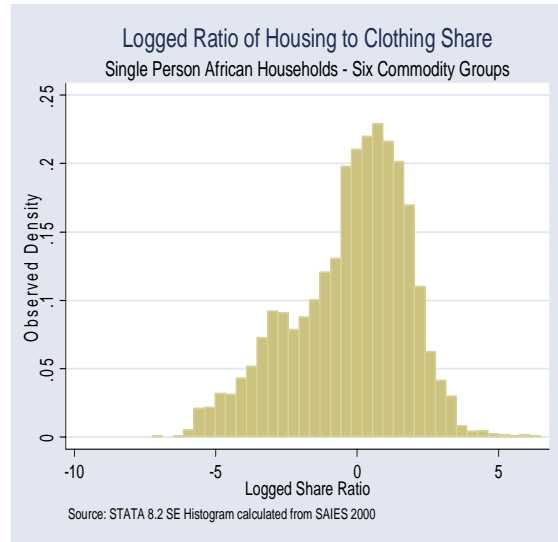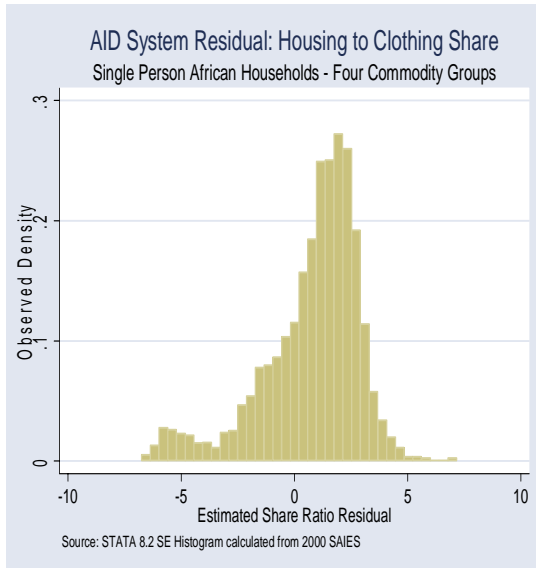
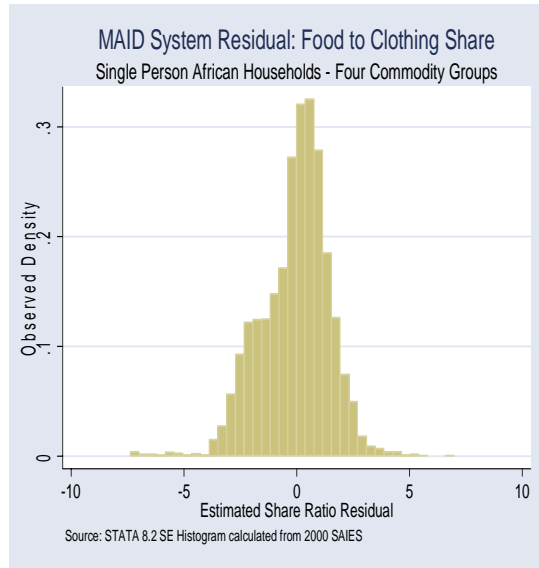**Figure 5. Six Commidity Group Logged Expenditure Share Ratios for Single Person Black Households**
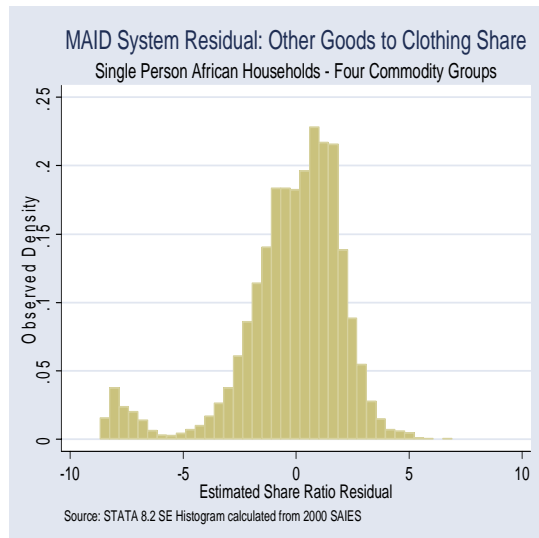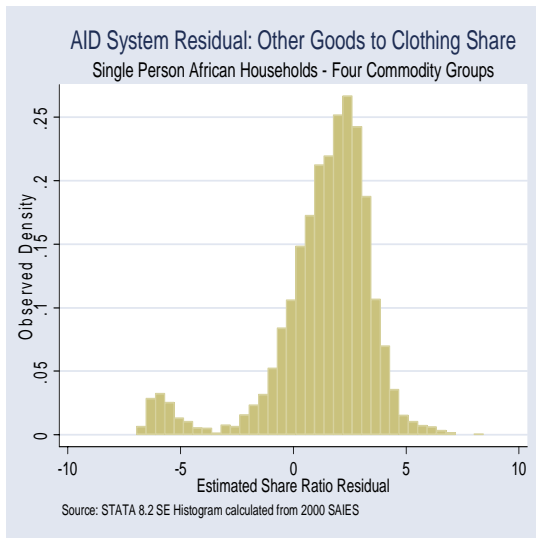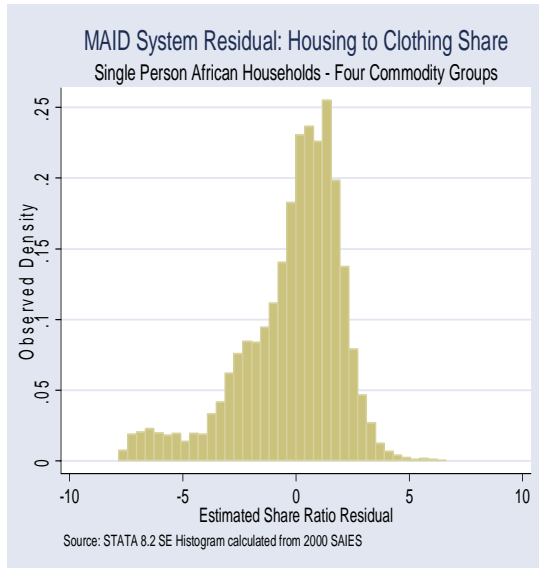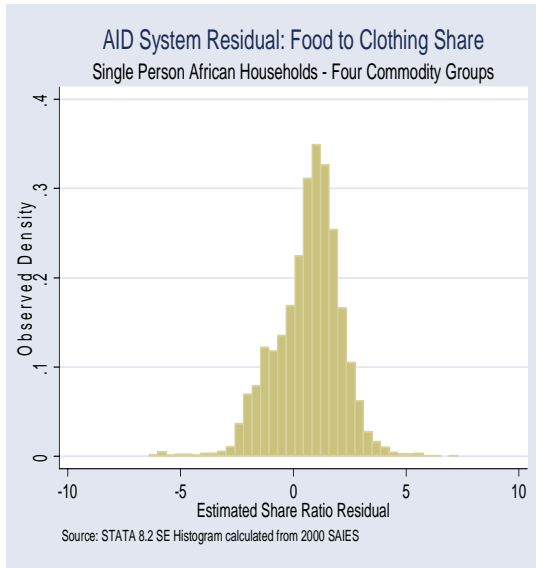
Figure 6. Non-normal Residuals from the Estimation of Equation (7) with Four Commodity Groups and $\eta = 0$ (on the left) and $\eta = 1$ (on the right)

Figure 7. Selected Non-normal Residuals from the Estimation of Equation (7) with Six Commodity Groups and $\eta = 0$ (on the left) and with $\eta = 1$ (on the right)