This is to certify that the
dissertation entitled

THE INFLUENCE OF DEMAND MODEL SELECTION ON
HOUSEHOLD WELFARE ESTIMATES: AN APPLICATION
TO SOUTH AFRICAN FOOD EXPENDITURES

presented by

LESIBA ELIAS BOPAPE

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Agricultural Economics

R. J. Myers

Major Professor's Signature

8/2/06

Date

THE INFLUENCE OF DEMAND MODEL SELECTION ON HOUSEHOLD WELFARE ESTIMATES: AN APPLICATION TO SOUTH AFRICAN FOOD EXPENDITURES

By

Lesiba Elias Bopape

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

2006

ABSTRACT

THE INFLUENCE OF DEMAND MODEL SELECTION ON HOUSEHOLD WELFARE ESTIMATES: AN APPLICATION TO SOUTH AFRICAN FOOD EXPENDITURES

By

Lesiba Elias Bopape

This study analyzes food expenditure patterns in South Africa, taking into account differences in demand behavior across rural and urban households, as well as across income groups. The study makes three main contributions from this study. First, it develops a Lagrange Multiplier (LM) test that can be used to determine whether the demand model should be specified with a quadratic or a linear expenditure term. The advantage of this test over the Wald test, which is based on the significance of the quadratic expenditure term in the demand model, is that it can be conducted without having to explicitly estimate the quadratic (in expenditure) demand model, which tends to be highly nonlinear. Second, the study examines the effects on household welfare of an indirect food tax reform, and evaluates the magnitude of the biases in the welfare estimates due to demand model is misspecification. The tax reform evaluated is the zero-rating of value-added tax (VAT) on meat products. Lastly, this study examines the differences in the consumption patterns between rural and urban households, and across households in different income groups. The study makes use of panel data on household food consumption in South Africa, collected as part of the KwaZulu-Natal Income Dynamics Study.

Results from both the LM and the Wald tests support the inclusion of the quadratic expenditure term. The implication of this finding is that popular functional forms such as the almost ideal demand system (AIDS), which have Engel curves that are linear in expenditure, would not give an accurate picture of demand behavior of the households considered in this study. Given these findings, this study estimates the quadratic almost ideal demand system (QUAIDS), which a generalization of AIDS that allows for a quadratic relationship between budget shares and expenditure. The QUAIDS model is used to estimate demand functions for seven food groups—grains, meat and fish, fruits and vegetables, dairy, oils and fats, sugar, and other foods. The endogeneity of expenditure in the demand model is explicitly tested and, where necessary, corrected for using the control function approach. The model is also adjusted to account for a large fraction of observed zero expenditures using a two-step procedure appropriate for equation system estimation.

Five of the seven food groups were found to be expenditure elastic, the exceptions being meat and fish and other foods. Demand behavior differs significantly between rural and urban households, as well as across income groups, implying that an accurate analysis of expenditure patterns in South Africa requires a disaggregated analysis that takes into account this heterogeneity in demand behavior. All households gain from the removal of VAT on meat, with welfare gains being larger for high-income households. On average, the AIDS expenditure elasticity estimates tend to be larger than the estimates based on QUAIDS. The AIDS model was also found to systematically overstate the welfare gains of the tax reform considered on this study, particularly for households with large expenditure levels.

To My Family
For Your Encouragement and Support

ACKNOWLEDGEMENTS

This dissertation is the product of joint efforts by many individuals, to whom I owe sincere gratitude. I am greatly indebted to the chair of my guidance committee, Professor Robert Myers, whose technical and professional advice most students can only hope for. The ingenious comments and suggestions from Professor Jeffrey Wooldridge greatly improved the quality of the methods used in this study. I also thank Professors John Staatz and Kellie Raper for their invaluable comments and suggestions during the various stages of this dissertation.

The technical support from Dr. Brian Poi of StataCorp greatly expedited the completion of this research. Professors Michael Carter of the University of Wisconsin, Julian May of the University of KwaZulu-Natal, and Ingrid Woolard of the University of Cape Town facilitated my access to the KIDS dataset. Dr. Jorge Agüero answered many of the questions I had on the technical details of the dataset.

This research was made possible by the financial support from the Department of Agricultural Economics at Michigan State University, Mandela Economics Scholars Program, the National Research Foundation of South Africa, and the Agricultural Research Council of South Africa.

I am greatly indebted to my friends and colleagues whose support during the four years I spent in Michigan greatly lightened the task of completing my studies. We trekked through the heavy Michigan snow with Gerald Nyambane for stimulating and intellectual discussions at the Peanut Barrel. My study partners, Lilian Kirimi and Mary Mathenge, still have to convince me on the workings of the Law of Iterated Expectations.

At last, I was able to convince my friend Elan Satriawan that I tackled endogeneity adequately in my model. I admire the ingenuity of Khathutshelo Sikhitha. Special mention also goes to the following individuals: Yanyan Liu, Anthony Chapoto, Wei Zhang, Ben Moholwa, Hikuepi Katjiuongua, Sindi Kirimi, Shaufique Sidique, Rui Benfica, Zhiying Xu, Vandana Yadav, Nomalanga Grootboom, and Annah Molosiwa.

I am deeply thankful to the love, support, and encouragement of my wife, Letta, and our two beautiful daughters, Napyadi and Pheladi. My family in South Africa made huge sacrifices during the long years of my graduate studies. My mother and late father have always believed in the power of education to transform lives. The weekend calls to my brothers Simon, Jonas, and Tshepo, and my sisters Maria and Monene, sustained my motivation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Aggregate per capita availability data suggest that South Africa is food secure in almost all basic foodstuffs. Furthermore, South Africa has the highest per capita income in Sub-Saharan Africa, and is categorized as a middle-income country with average per capita gross national income of US $3,650 in 2004 (World Bank, 2004). These facts suggest that hunger and food security should not be major policy issues in the country. However, these aggregate data mask a highly unequal distribution of income and a huge divide between relatively affluent urban areas and destitute conditions in many rural communities. The richest 20% of the population receives over 60% of the income while the poorest 20% receives less than 3% (World Development Report, 2002). At the household level, over 30% of the population is categorized as vulnerable to food insecurity and over 20% of the children are estimated to be stunted and vitamin A deficient (Human Science Research Council, 2004).

Policies designed to reduce income inequality, hunger, and malnutrition, have had mixed results. Major social, economic, and political reforms introduced since the demise of apartheid and the emergence of democratic government in 1994 have obviously redistributed wealth. But income inequality and household food insecurity remain. One of the problems is that little is known about how food expenditure patterns differ across different income groups, and across different geographic regions. Without a thorough understanding of the heterogeneity of food expenditure patterns, and how these patterns are changing over time, it will continue to be difficult to design policies that improve food security effectively over a broad range of heterogeneous low-income households. An accurate assessment of the distributional impacts of policies such as commodity tax reform requires accurate estimates of price and income effects, and how these differ across households in different socioeconomic groups.

This study seeks to improve knowledge and understanding of the heterogeneity in food expenditure patterns in South Africa. The study makes use of an unusually rich panel dataset on household food consumption, collected as part of the KwaZulu-Natal Income Dynamics Study (KIDS). The KIDS dataset contains detailed information on household socioeconomic and demographic characteristics, which permit heterogeneity effects to be analyzed. The dataset followed the same households over a ten-year period, with surveys in 1993, 1998, and 2004, to study changes in their incomes, expenditures, and poverty levels. Data on prices and expenditures on various food products consumed by households were also collected.

This study utilizes the KIDS data to estimate demand functions for seven food groups—grains, meat and fish, fruits and vegetables, dairy, oils and fats, sugar, and all other foods. Household locations, socioeconomic characteristics, and income levels are used to explain heterogeneity in food expenditure patterns.

1.2 Research Gap and Study Motivation

This study makes three main contributions. *First*, to the best of our knowledge this study is the only theoretically consistent panel data study of food consumption in South Africa done to date. Previous studies on food consumption in South Africa have either been limited to examining only one commodity (e.g., Taljaard, 2003; Nieuwoudt, 1998; Poonyth *et al.*, 2001) or have used highly aggregated composite commodity definitions, and typically ignored any impact of demographic factors on food demand (Bowmaker and Nieuwoudt, 1990; Liebenberg and Groenewald, 1997). The only theoretically consistent study of food demand in South Africa we are aware of that uses micro-level data incorporating household demographic characteristics is by Agbola (2003). However, Agbola uses cross-sectional data that was collected in 1993—one year prior to South Africa's first democratic government. Clearly, such data do not capture periods of important social and economic reforms that affect households' profiles and, hence, their food consumption patterns. As earlier mentioned, South Africa became a democracy in 1994 and as a result major policies were implemented with implications that the study by Agbola (2003) does not capture. The KIDS panel dataset allows for greater price variability across the sample, and also covers most of the period associated with major policy reforms in South Africa.

Unlike most previous food demand studies, this study explicitly tests for the endogeneity of expenditure in the budget share equations, and then controls for it. Among existing food demand studies, only LaFrance (1991) and Dhar *et al.* (2003) consider the problem of expenditure endogeneity. Expenditure endogeneity may arise whenever the household expenditure allocation process across food groups is correlated with other factors not captured by the explanatory variables used in demand estimation (i.e., bundled in the error term). In this case, least squares estimation of the demand model gives inconsistent parameter estimates. This study takes advantage of recent advances in econometric methods designed to overcome this problem, and to enhance demand estimation with micro level panel data.

We estimate a quadratic almost ideal demand system (QUAIDS) controlling for expenditure endogeneity, and explicitly accounting for the problem of observed zero-expenditures. Most of the previous food demand studies in South Africa use the almost ideal demand system (AIDS) model. The shortcoming of the AIDS model is that it assumes linear Engel curves and constant expenditure elasticity. Such assumptions have been shown to be restrictive, even in developing countries (examples include Meekashi and Ray (1999) and Abdulai (2004)).

Furthermore, because of high income inequality and large disparities in the economic conditions between rural and urban households in South Africa, pooling data across all households obscures important information on variability in demand behavior across households in different socioeconomic and demographic groups. To determine the impact of this household heterogeneity on demand, this study analyzes separately the food expenditure patterns of rural and urban households, as well as households in different income groups.

Second, this study builds on the work of Banks *et al.* (1997) to develop a test that can be used to determine whether the demand model should be specified with a quadratic (QUAIDS) or a linear (AIDS) expenditure variable. In particular, the implication of corollary 2 in Banks *et al.* (p.533) is that a utility-derived demand system that is rank 3 and exactly aggregable cannot have coefficients on both the linear and the quadratic expenditure terms that are independent of prices.[1] In other words, if such a demand model has a coefficient on the linear expenditure term that is independent of prices, then it *must* have a coefficient on the quadratic expenditure term that is price dependent. This study uses Bank *et al.*'s corollary 2 to develop a Lagrange Multiplier (LM) test that allows one to determine whether or not a QUAIDS specification is necessary. No other study was found to have explicitly conducted this test, certainly not with South African data. Hence, this study provides richer information on food consumption behavior in South Africa than has been obtained from existing studies.

Finally, the study examines the effects on household welfare of zero-rating the value-added tax (VAT) on meat products.[2] While most of the basic food commodities such as grains, milk, fruits and vegetables are zero-rated in South Africa, meat is not. Meat is taxed at the standard VAT rate of 14%. Whether or not meat should be zero-rated has been a subject of contention between the government and lobby groups (most notably the Congress of South African Trade Unions) since the introduction of VAT in 1991 (Watkinson and Makgetla, 2002). This study contributes to this issue by providing quantitative measures of the impacts of this tax reform on household welfare. We use the QUAIDS parameter estimates to calculate indirect utilities before and after the tax reform. These are then used to compute two money metric welfare measures of the tax effect, namely compensating variation and equivalent variation. To determine the sensitivity of these welfare measures to demand model selection, and the bias that results

[1] As will be made clearer below, the rank of a demand system has implications for aggregation and the non-linearity of Engel curves. Higher rank models are well suited to approximate non-linear Engel curves often found in empirical analyses. QUAIDS has a rank of 3.

[2] A commodity is zero-rated if it is taxable, but taxed at a rate of 0%. Zero-rating a commodity is different from exempting it; by law, exempted commodities cannot be taxed (i.e., they are not taxable).

when a restrictive functional form is used, we also estimate these welfare measures using parameters from the (nonlinear and linear) AIDS model. There are no studies we are aware of that compute these welfare measures for South Africa, certainly not with this dataset.

1.3 Objectives of the study

The broad objective of this study is to analyze the responsiveness of South African households to food price and income changes as well as other relevant socioeconomic factors, particularly focusing on the KwaZulu-Natal Province. This objective is accomplished by estimating a food demand model using appropriate econometric techniques.

The specific objectives are:

1. To estimate a household food demand model for South Africa. The model accounts for the effects of demographic and socioeconomic characteristics, and explicitly controls for expenditure endogeneity and observed zero expenditures.
2. To determine how food expenditure patterns differ across rural and urban households, as well as across income groups.
3. To estimate price and expenditure elasticities of demand for food using the model from objective (1), and to evaluate how these differ across rural and urban households as well as across households in different income groups.

4. To examine the effects on household welfare of zero-rating the value-added tax (VAT) on meat products

The results of this study will provide important insights into food policy formulation and implementation in South Africa. In particular, accounting for household heterogeneity in demand has implications on the likely effects of alternative policies on food consumption and food safety nets. These results should be particularly useful in implementation of the national food security strategy—the Integrated Food Security Strategy, established in 2002 by the South African cabinet—which emphasizes improvements in household-level nutrition and increases in the provision of food safety nets. These policies can be made more effective if they are based on behavioral parameters specific to particular demographic and socioeconomic groups.

1.4 Organization of the Dissertation

The rest of the dissertation is organized as follows. The next chapter presents a review of related literature. The relationship between utility maximization theory and demand functional forms is discussed, and a review of alternative approaches to modeling preferences is provided. Chapter three presents the empirical model and discusses various econometric tests to be implemented. Chapter four describes the survey and data sources, and then presents a descriptive analysis of expenditure patterns across time and household groups. Chapter five presents the empirical results, while Chapter six concludes with a summary and conclusions.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

2.1 Introduction

This chapter reviews the literature on the theory and empirical estimation of consumer demand. The first section discusses various approaches to restricting a large number of goods in the consumer's utility maximization problem to a smaller number, more manageable for empirical estimation. This is followed by section 2.3 which discusses demand system functional forms, and the assumptions about their underlying preferences. The various methods for incorporating demographic variables are discussed in section 2.4. This is followed by a review of the literature on censored demand modeling. The final section summarizes the key points in this chapter.

2.2 Commodity Grouping and Separability

In the standard utility maximization problem, a consumer makes budget allocation decisions on large numbers of goods with different relative prices. The solution to this problem gives the amount demanded of each good as a function of its price, prices of other goods, and the consumer's income. However, when the number of goods involved is too large, the consumer's allocation problem becomes complex for the empirical analyst. The problem of finding theoretically appealing approaches to reducing this large number of goods to a relatively small and more manageable number has attracted, and continues to attract, attention in the demand literature. The literature proposes two alternative approaches; one that groups commodities based on the behavior of their relative prices—the composite commodity theorem—; and another that makes assumptions about the consumers' preferences—separability and two-stage budgeting.

Originally proposed by Hicks (1936) and Leontief (1936), the *composite commodity theorem* asserts that, if a group of prices move in parallel, then the corresponding group of commodities can be treated as a single good. The price and quantity indices of the commodity groups are used to derive an expenditure function that satisfies the usual properties of expenditure functions (increasing in utility and prices, concave in prices, and linearly homogenous).

The usefulness of the composite commodity theorem in constructing commodity groupings for empirical analysis is limited (Deaton and Muellbauer, 1980a). One source of limitation is that relative prices fluctuate considerably in practice. Also, it would be difficult to justify some of the aggregates that are imposed. For example, a relatively volatile price of meat would prevent it being grouped with other foods whose prices are relatively stable.

In an attempt to circumvent some of these limitations, Lewbel (1996) develops a generalized composite commodity theorem, which is an extension of original Hicks-Leontief idea. The generalized composite commodity theorem relaxes the assumption of perfect correlation among group prices and instead allows for less than perfect co-movement among intra-group prices. It assumes that the distribution of an individual commodity's price is independent of the composite group price, and tests for the generalized composite commodity theorem are based on cointegration relationships between each of the good's prices and the price indices of groups to which they belong. Interesting applications of Lewbel's generalized composite commodity theorem are by

Davis (2003) and Reed *et al.* (2005). Davis extends Lewbel's bivariate Engel-Granger testing approach to a multivatiate framework, while Reed *et al.* applies the generalized composite commodity theorem to nonlinear demand systems.

In contrast to the composite commodity theorem which relies on an external factor (namely, the constancy of relative prices) to define commodity groups, *separability* defines commodity groups using consumer preferences themselves. If preferences are weakly separable, then commodities can be partitioned into groups so that preferences within groups can be described independently of consumption in other groups. This implies that a subutility function can be defined for each group, so that the values of each of these subutilities can be added to give total utility. The concept of a utility tree, proposed by Strotz (1957, 1959), allows consumers to break a decision into multiple steps. A closely related concept is two-stage budgeting, which hypothesizes that consumers first allocate total expenditure to broad groups of goods (the first stage) and then allocate each group expenditure to the individual commodities in that group (the second stage). Weak separability is both necessary and sufficient for the second stage of two-stage budgeting (Deaton and Muellbauer, 1980a).

Several studies attempt to empirically test the restrictions imposed by separability within flexible demand systems. These studies derive functional relationships that must hold between goods that belong to the same group and goods that belong to other groups, expressed in terms of price and expenditure elasticities. They then econometrically test for whether these relationships are supported by data. Included in these studies are those by Eales and Unnevehr (1988), Moschini (1992), Moschini *et al.* (1994), Nayga and Capps (1994), Edgerton (1997), and Carpentier and Guyomard (2001).

However, there are a number of problems associated with the tests mentioned above. One of the problems, as mentioned by Lewbel (1996), is that weak separability restrictions require that group prices depend on the parameters of the individual's utility function. Also, separability restrictions are difficult to test powerfully due to the multicollinearity of aggregate price data. In the context of this study, a more important problem with both the tests for separability and commodity groupings is that they have been developed within the context of time series data. Their implementability in the context of cross-sectional or panel data is limited. Given that this study uses panel data, we will follow the 'traditional' approach of maintaining the assumption of weak separability between foods and all other broad consumption goods. The food items that are closely related will then be grouped together into composite food commodities, where a group comprises items that are closely substitutable.

2.3 Modeling Preferences

The preference-based approach to modeling choice behavior treats the individual's tastes, summarized by a preference relation, as his or her primitive characteristic. The theory is developed by first imposing rationality axioms on the individual's preferences and then analyzing the consequences of these preferences on his or her choice behavior.

The preference-based approach to modeling choice behavior provides a useful framework for analyzing data on demand. Functional forms proposed for the econometric analysis of these data can be evaluated in terms of whether they are consistent with theory. Failure of data to conform to theoretical predictions may indicate excessive restrictiveness of the chosen functional form. It may also mean that people do not behave as theory suggests. When the chosen functional form gives meaningful results, then it becomes the basis for estimating such demand behavioral parameters as price and income elasticities.

The earliest empirical demand studies are characterized by extensive use of single equation methodology. At the center of these analyses has been the measurement of elasticities. The requirement that demand systems satisfy properties such as adding-up was ignored and perhaps unimportant because these early studies considered only a fraction of the total budget (Deaton and Muellbauer, 1980a). This made it tempting to choose explanatory variables pragmatically with the goal of getting better model fits.

The single equation approach to demand modeling changed with the introduction by Stone (1954) of the linear expenditure system (LES). The LES was among the first attempts to derive utility-based demand models. The derivation of LES imposes theoretical restrictions of adding-up, homogeneity, and symmetry. Among the implications of the LES are that goods cannot be inferior and that all goods must be substitutes, which obviously makes it too restrictive a functional form to model demand, except for cases where commodities are grouped into very broad categories so that it is reasonable not to expect inferiority or complementarity among them. Also, while the imposition of the theoretical restrictions helps generate degrees of freedom, it also forces the analyst to take theoretical restrictions as given (because these restrictions are embedded in the model). An interesting alternative is one that allows restrictions to be tested empirically.

The Rotterdam model proposed by Theil (1965) and estimated by Barten (1969) allows for restrictions to be tested statistically. In many ways, the approach followed in deriving the Rotterdam model is similar to Stone's, except the Rotterdam model is specified using variables in first-differences. Like the LES, the derivation of the Rotterdam model emphasized the use of theoretical restrictions to generate degrees of freedom. Among the limitations of the Rotterdam model are that it imposes constant price and expenditure elasticities and that it typically does not satisfy theoretical restrictions when applied to data (Deaton and Muellbauer, 1980a).

It became apparent that specifying functional forms using theoretical restrictions to generate degrees of freedom did not offer much promise. It was partly for this reason that research in the 1970s focused on developing flexible functional forms. This approach entailed approximating the direct utility function, the indirect utility function, or the cost function with some specific functional form that has enough parameters to be regarded as a reasonable approximation to whatever the true unknown function might be. Important contributions in this regard were the transcendental logarithmic (translog) model of Christensen, Jorgensen, and Lau (1975) and the almost ideal demand system (AIDS) model of Deaton and Muellbauer (1980b).

The indirect translog model (as originally proposed by Christensen *et al.* (1975)) is derived by applying Roy's identity to a function that approximates the unknown indirect utility function by a quadratic form in the logarithms of the price to expenditure ratios. Unfortunately, the demand functions derived from this indirect utility function are complicated and difficult to estimate. Its modified version (Jorgenson *et al.*, 1982), the direct translog model, makes the discomforting assumption that, for all goods, prices are determined by quantities rather than the other way round.

Deaton and Muellbauer's AIDS model marked an important breakthrough in the quest for flexible functional forms. In fact, no dramatic advances have been made since its introduction in 1980, although some refinements (discussed next) have been made. The particularly desirable properties of AIDS are that it satisfies the axioms of choice exactly and can be interpreted in terms of economic models of consumer behavior when applied to either aggregate or disaggregate (e.g., household) data. It also allows for consistent aggregation of individual demands to market demands.

Both AIDS and Jorgenson *et al.*'s translog models are members of the Price-Independent Generalized Logarithmic (PIGLOG) class of demand models (Muellbauer, 1976), which have budget shares that are linear functions of log total expenditure. Specification of Engel curves (i.e., relationships between a commodity's budget share and total expenditure) that are linear functions of log total expenditure are extensions of the earlier work by Working (1943) and Lesser (1963).

For many commodities, however, there is increasing evidence that Engel curve analysis based on this Working-Lesser form does not provide an accurate picture of behavior. Empirical Engel curve studies indicate that further terms in total expenditure are required for some, if not all, expenditure share equations (Lewbel, 1991; Blundell *et al.*, 1993). Also, Engel curves may vary with the labor market status and region (Browning and Meghir, 1991). Banks *et al.* (1997) generalize PIGLOG preferences to allow for nonlinearities in total expenditure. We discuss in general terms the differences in approach to modeling preferences using the PIGLOG class vis-à-vis other general classes.

Let w_i denote expenditure share of good i (i = 1, ..., K), x denote total expenditure, and $a(\mathbf{p})$ denote a price index used to deflate total expenditure, where $\mathbf{p}$ is a K-vector of prices. A general form that nest those derived from the PIGLOG class is

$$w_i = A_i(\mathbf{p}) + B_i(\mathbf{p})\ln x + C_i(\mathbf{p})g(x) \tag{2.1}$$

where $A_i(\mathbf{p})$, $B_i(\mathbf{p})$, $C_i(\mathbf{p})$, and $g(x)$ are differentiable functions. Equation (2.1) says that expenditure shares are linear in log total expenditure and another function of total expenditure, represented by $g(x)$. Thus, the $C_i(\mathbf{p})g(x)$ term allows for potential nonlinearity in demands. Engel curves of the PIGLOG class have $C_i(\mathbf{p})$ equal to zero, so that for this class of preferences, demands are modeled as linear functions of ln x. Lewbel (1991) defines the rank of a demand system as the dimension of the space spanned by its Engel curves. Based on this definition, the rank of equation system (2.1) equals the rank of the N × 3 matrix of Engel curve coefficients, having rows $[A_i(\mathbf{p}) : B_i(\mathbf{p}) : C_i(\mathbf{p})]$ for good i (Banks *et al.*, 1997). This matrix has three columns, so 3 is the maximum possible rank of equation system (2.1). Exactly aggregable demand systems are defined as demand systems that are linear in functions of x. Gorman (1981) proved that the maximum possible rank of any exactly aggregable demand system (with any number of terms) is 3. Thus, based on this theoretical result, there would be little or no gain in adding additional terms of the form $D_i(\mathbf{p})h(x)$ if exact aggregation is desired. In fact, Banks *et al.* show that all rank 3 exactly aggregable utility-derived demand systems of the form represented by equation (2.1) have $g(x) = (\ln x)^2$. So, given that rank 3 forces $g(x)$ to have this specific functional form, budget shares of form (2.1) are quadratic in ln x, and therefore, are quadratic in ln x itself.

Banks *et al.* (1997) also show that rank 3 exactly aggregable demand systems cannot have both $B_i(\mathbf{p})$ and $C_i(\mathbf{p})$ independent of prices. So, the AIDS model has the form of equation (2.1) with each B_i constant (that is, independent of prices) and every $C_i = 0$. To allow for potential nonlinearity in expenditure, it may be tempting to consider extending the AIDS model by simply adding a squared expenditure term with a constant, nonzero coefficient C_i. In fact, a number of studies do this (Blundell *et.al.*, 1993; Labeaga and Puig, 2002; Christensen, 2004; Browing and Collado, 2004). However, as Banks *et al.* show (Corollary 2, p. 533), B_i and C_i cannot both be constants for all commodities *i* in a rank 3 demand system. Based on these algebraic facts and restrictions from utility theory, Banks *et al.* derive an exactly aggregable rank 3 demand system —the Quadratic Almost Ideal Demand System (QUAIDS).

The QUAIDS model is of the form represented by equation (2.1), with the nonlinear expenditure term set to quadratic and its coefficient, $C_i(\mathbf{p})$, dependent on prices via the inverse of a Cobb-Douglas price aggregator. Hence, the QUAIDS model nests the AIDS model. Due to the restriction by AIDS that B_i is constant and $C_i = 0$, its expenditure elasticities are constant. In contrast, QUAIDS permits goods to be luxuries at some expenditure levels and to be necessities at others. In this study, we estimate the demand parameters and the price and income elasticities using the QUAIDS model, given its generality over AIDS and its other desirable properties to be explained in detail in the next chapter.

Recently, several demand studies have emerged that confirm the appropriateness of QUAIDS in modeling preferences. Examples using developed country data include Abdulai (2002) who applies QUAIDS to the food expenditure data from Switzerland, Moro and Sckokai (2000) who use Italian food expenditure data, Banks *et al.* (1997) and Blundell and Robin (1999) who both use expenditure data on broad consumption goods from the U.K., and Fisher *et al.* (2001) who apply QUAIDS to the U.S. aggregate consumption data. A number of studies in developing countries are also emerging that support QUAIDS. However, these studies are fewer compared to those from developed countries. Examples include Abdulai and Aubert (2004) using Tanzanian food expenditure data, Meenkashi and Ray (1999) using Indian food expenditure data, Gould and Villarreal (2006) using food expenditure data from urban China, and Molina and Gil (2005) using aggregate consumption data from Peru.

2.4 Demographic Variables in Demand

Demographic variables such as household size and age composition play an important role in determining household demand patterns. The treatment of demographic effects in the context of theoretically plausible demand systems dates back to Barten (1964). Since Barten's work, studies proliferated that were aimed at finding theoretically consistent techniques to incorporate demographic effects into demand analysis (for a review, see Pollak and Wales (1992)).

Ideally, there are two ways to incorporate demographic effects into a demand system; one is to use unpooled data and the other to use pooled data (Pollak and Wales, 1978). The first approach involves separating the entire dataset into sub-samples with identical demographic profiles and then estimating a demand system for each sub-sample separately. This approach allows all of the parameters of the demand system to depend on the demographic variables, so that there is no need to specify the form of the relationship between the parameters and the demographic variables. The major drawback of this approach, apart from its apparent inefficiency, is that it does not make it possible to draw inferences about households with one demographic profile from observations on the behavior of households with different profiles.

The second approach, which uses pooled data, involves three separate but interrelated steps (Pollak and Wales, 1980). The first step involves specifying a class of demand systems for every admissible demographic profile. The second involves specifying which parameters depend on the demographic variables and which do not, and the third involves specification of a functional form for each parameter which depends on the demographic variable.

Earlier attempts to incorporate demographic effects into complete demand systems using pooled data have led to the development of five general procedures: (i) demographic scaling of Barten (1964); (ii) Gorman's (1976) specification; (iii) the reverse-Gorman specification; (iv) the modified Prais-Houthakker procedure; and (iv) demographic translating of Pollak and Wales (1981). The procedures are general in the sense that they do not assume the original demand system has a particular functional form, but can be used in conjunction with any complete demand system.

Demographic translation replaces the original demand system, $w_i(\mathbf{p}, m)$, by $w_i(\mathbf{p}, m) = \delta_i + \overline{w}_i\left(\mathbf{p}, m - \sum_{j=1}^{K} p_j \delta_j\right)$, where the δ's are translation parameters. Hence, specifying demographic variables using translation can be viewed as allowing "subsistence" (typically the intercept) parameters of a demand system to depend on the demographic variables. Linear demographic translation, specifies demographic variables as intercept shifters

$$D_i(\mathbf{z}) = \sum_{s=1}^{S} \delta_{is} z_s \qquad (2.2)$$

where $\mathbf{z} = (z_1, \ldots, z_S)$ is a vector of demographic variables, and the δ's are parameters to be estimated. Linear demographic translation is the most common specification in empirical demand studies.

Demographic scaling involves applying scaling functions to prices and quantities. The scaling functions depend on demographic variables, and are interpreted as reflecting the number of 'equivalent adults" in the household (when the same scaling functions are the same for all goods), or as measuring the number of equivalent adults on a scale appropriate to each good (when the scaling functions differ from one good to another). This procedure leads to the interpretation of the household's preferences as depending not on the quantity of the raw commodities it consumes, but on the quantity per equivalent adult. The challenge when using demographic scaling is the criteria for choosing scaling functions or for choosing numerical scaling values. The criteria for specifying scaling functions have been based on such factors as nutritional and physiological needs, poverty measures, and expenditure behavior of households. While such criteria may be intuitively appealing, they are often inconsistent with theory. Scaling procedures that are made to be theoretically consistent typically lead to the imposition of implausible behavioral assumptions, such as zero substitution possibilities among goods (Deaton and Muellbauer, 1980a).

Gorman (1976) proposed a general form that incorporates demographic translating and scaling. Gorman's specification is obtained from the original demand system by first scaling and then translating. The "reverse Gorman" specification is very similar to Gorman's specification, except it is obtained first by demographic translating and then scaling. Given that this form proposed by Gorman and its "reverse" version nest demographic scaling, it inherits the same weaknesses associated with scaling.

Proposed in its original form by Prais and Houthakker (1955), the modified Prais-Houthakker procedure incorporates demographic variables into demand equations using a single income scale and a specific scale for each good. The Prais-Houthakker procedure replaces the original demand system by $w_i(\mathbf{p}, m) = s_i \overline{w}_i(\mathbf{p}, m / s_0)$, where the s_i's are "specific scales" for commodities which depend on the demographic variables, and s_0 is an "income scale" implicitly defined by the budget constraint $\sum_{i=1}^{K} p_i s_i \overline{w}_i(\mathbf{p}, m / s_0) = m$. However, Prais and Houthakker never reconciled their technique with an overall budget constraint (Pollak and Wales, 1981). The main limitation in applying the Prais-Houtkker procedure is that it does not yield a theoretically consistent demand system. Pollak and Wales (1981) show that this procedure yields theoretically plausible demand systems only under the very special case where the original demand system corresponds to an additive direct utility function. The limitation imposed by this additivity restriction is quite severe. The implication of the additivity restriction is that no good (or group of goods) occupies any special position in the utility function (Deaton and Muellbauer, 1980a). Since the function is additive, new groups can always be created by combining any others, such that no particular relationship exists between pairs of goods.

Lewbel (1985) extended Gorman's (1976) procedure by proposing a unified approach which combined the five procedures explained above. Lewbel's procedure modifies the expenditure function by first replacing each price by a function that depends on all prices and demographic variables and then subjecting the resulting expenditure function to a further transformation that depends on all prices and demographic variables. However, Lewbel's contribution was mainly theoretical and too general to apply empirically; hence it has rarely been used in empirical work.

A relatively recent study by Bollino *et al.* (2000) extends Gorman's (1976) procedure by following an approach similar to Lewbel's. Unlike Lewbel, Bollino *et al.* provide both a theoretical derivation of their technique and a procedure for empirical estimation. Unfortunately, the estimation procedure proposed in Bollino *et al.* is computationally complex, and it can accommodate the estimation of only a few consumption categories. In their paper, Bollino *et al.* applied their procedure to only three categories of goods.

An unambiguous ranking of these procedures is not possible (Pollak and Wales, 1981). One of the factors that make it difficult to rank these procedures is that not all of them are nested. Their assessment also depends on the functional form used to estimate the demand system. The theoretically more appealing technique of Bollino *et al.* (2000) restricts the number of consumption categories that can be analyzed. However, the number of goods estimated in empirical demand systems is large, so that its usefulness in practice can at best be very limited. In this study, we are estimating demand for seven food groups, which immediately rules out the use of Bollino *et al.*'s procedure. We estimate the QUAIDS model in its most flexible form, allowing for nonlinearity in the price index used to deflate total expenditure and allowing the coefficient of the quadratic expenditure term to depend on prices. Given that we are estimating QUAIDS in this highly nonlinear form, a preferred method to introduce demographic variables is one that will not create further nonlinearities. It is for this reason that we choose to incorporate demographic variables as intercept shifters through Pollak and Wales's (1981) linear demographic translation method.

2.5 Observed Zero Expenditures

The behavioral response of households to changes in their economic environment takes place on either the intensive or extensive margin (Meyerhoefer, 2002). Households respond along the intensive margin when they are consuming a non-zero amount of the good, so that a change in the independent variable (such as a commodity's price) leads them to marginally increase or decrease the amount they presently consume. The extensive margin refers to households that must make a decision whether or not to consume any amount of the good when its price (or some other exogenous factor) changes. These are households who are either not consuming the good initially, or those that respond to the exogenous change by completely exiting the market for that good. The response to changes in exogenous factors by households on the intensive margin entails a continuous change in the dependent variable, and can be easily modeled by traditional regression techniques. However, modeling scenarios with some households on the extensive margin and others on the intensive margin requires statistical analyses based on composite distributions. These are defined to contain a discrete probability mass on the boundary of the choice set, allowing for the positive probability of zero consumption, and a continuous density corresponding to positive consumption levels (Meyerhoefer, 2002).

The early empirical work in demand modeling estimated demand functions on aggregate time series data, or household level data with highly aggregated commodity groupings. Demand estimation with these aggregate data allow the use of standard econometric techniques that assume the dependent variables in the system of demand equations follow a joint normal distribution, and hence, do not allow for the positive probability of zero expenditure levels. When aggregate data are used, the number of observations with zero expenditure share values is typically very small, such that deleting these observations from the sample and carrying out estimations on only the positive observations consistently identifies the demand function (Meyerhoefer, 2002). Subsequent work on demand modeling made increasing use of micro data on highly disaggregate commodity groups. When micro data are used, it becomes increasingly likely to observe non-consumption of some commodities by a large number of households. This makes the strategy of deleting non-consuming households unattractive, particularly given the large number of degrees of freedom that is lost. Also, the exclusion of a large number of observations in this nonrandom manner may cause selection bias.

The first attempts to develop estimation techniques that explicitly capture consumer behavior on the extensive and intensive margins were done in a single equation context. In this context, several limited dependent variable models have been developed to deal with zero expenditure values generated by different underlying processes. One of the reasons for observed zero expenditures is that the market price for a given commodity exceeds the household's reservation price, leaving the household at a corner solution and censoring the expenditure distribution at the point of non-consumption. This reasoning has motivated the use of the Tobit model (Tobin, 1958) to estimated censored expenditure relationships. Under the Tobit formulation, the same variables are assumed to determine both the value of the continuous observations and the discrete switch to non-consumption at zero, making it only appropriate in cases where consumers are rationed out of the market by prices higher than they are willing to pay.

Other models have been developed that are appropriate for situations where zero expenditure values are a result of the infrequency of purchase, which occurs when the purchase of some commodities is not observed due to the short span of the survey period (Deaton and Irish, 1984; Blundell and Meghir, 1987). Popular among these is the "double hurdle model" of Craig (1971).

The study by Wales and Woodland (1983) was among the first attempts to derive econometric techniques to estimate a theoretically plausible demand system in the presence of zero expenditures. Wales and Woodland propose two alternative models to estimate censored systems of equations, based on assumptions about preferences. The first model assumes that preferences are randomly distributed in the population, so that each individual's direct marginal utility function for each good can be additively augmented with a normally distributed error term. These stochastic marginal utility functions are then substituted into the Kuhn-Tucker conditions to determine the set of goods with zero consumption and define the commodity demand functions. A multivariate normal density function for the vector of commodity demands is derived through a change of variables transformation, and used to assign a probability to each possible combination of consumption and non-consumption. The number of integrations to be performed on the density is equal to the number of non-consumption realizations. Unfortunately, the derivation of the maximum likelihood estimates in this case involves the evaluation of multiple integrals, which can be computationally infeasible for large equation systems.

The second model proposed by Wales and Woodland assumes that commodity demands are the result of individual nonrandom utility maximization subject to a budget constraint. This second model essentially extends Amemiya's (1974) Tobit estimator for a system of simultaneous equations to account for the budget constraint during estimation. An error term, assumed to follow a truncated multivariate normal distribution, is added to the demand share equations. As is the case with the first model, the truncated density is obtained by integrating non-consumed goods out of the joint normal density, and the likelihood function is constructed as the product of the individual truncated density functions.

The main difference between the two approaches lies in the assumptions each makes about the processes generating the zero consumption values. The first model assumes that zero consumption is determined by Kuhn-Tucker conditions, so that stochasticity enters the model through random preferences, while the second model incorporates stochasticity through additive disturbances on the share equations, so that the possibility of zero consumption occurs because disturbances follow a truncated joint normal density. The similarity between these models lies in the fact that both assume zero expenditures represent corner solutions where consumers are rationed out of the market by prices higher than they are willing to pay. The drawback in both models is that their empirical implementation is virtually infeasible for larger systems of equations, given the difficulty of performing multiple numerical integrations.

Building on Wales and Woodland's first model, Lee and Pitt (1986) develop a method for estimating censored demand systems that is dual to the Kuhn-Tucker approach. Lee and Pitt also assume that preferences are randomly distributed over the population but, unlike Wales and Woodland, they use the indirect utility function resulting from utility maximization without non-negativity constraints. Application of Roy's Identity to the indirect utility function defines what Lee and Pitt call unconstrained "latent notional demands", each of which is a function of market prices. They call the demands notional because they result from utility maximization with respect to the budget constraint only, allowing them to take on negative values, and latent in the sense that only nonnegative realizations are observable (Meyerhoefer, 2002). The notional demand functions can be related to observed demands by finding positive shadow prices, which are themselves functions of observable market prices, supporting the zero-valued demand levels. Households compare shadow prices to market prices to select a demand regime, so that if the shadow price of a good is less than its market price consumption is zero. The likelihood function is constructed as the product of the conditional density of disturbances for the consumed goods given the non-consumed disturbances with the density of disturbances for the non-consumed goods. The likelihood must be integrated over the domain of shadow price values. The drawback of Lee and Pitt's approach is the same as that of Wales and Woodland, namely that it can only be used where a small number of commodities is involved. Hence, despite the theoretical attractiveness of the Wales and Woodland and Lee and Pitt approaches, their computational infeasibility limit their usefulness in practice.

Arndt (1999) proposed a methodology that attempts to overcome the above limitation. Arndt follows Lee and Pitt's formulation of estimating equations by solving for the reservation prices and substituting them into the demand functions. But instead of specifying a likelihood function to carry out the estimation, Arndt proposes maximizing an entropy function subject to restrictions. These restrictions include prior distributions imposed on elasticity estimates, symmetry restrictions, constraints that the reservation prices are less than or equal to market prices, and the requirement that all outcome probabilities sum to one. Since entropy maximization problems do not involve the evaluation of numerical integrals, they can be solved using standard nonlinear optimization packages. However, whether the maximum entropy estimator meets the requirements of economic theory, particularly the curvature restrictions, has not been fully investigated (Meyerhoefer, 2002).

Motivated by the work of Heckman (1976), other alternative two-step procedures have been developed to reduce the computational burden of multiple integrals that bedevils one-step procedures. Heien and Wessells (1990) proposed a two-step estimation procedure for a system of demand equations with limited dependent variables. In the first step, a probit model is used to calculate the inverse Mills ratio (IMR) for each commodity. The IMR is then used as a selectivity regressor in each equation during the second stage. The system in the second stage is estimated with seemingly unrelated regression (SUR). Heien and Wessells's procedure has seen widespread use in empirical food demand studies (see Yen *et al.* (2002) for examples). However, it was later shown that Heien and Wessels's procedure is inconsistent due to the presence of a mathematical error in its derivation (Shonkwiler and Yen, 1999).

One feature to note about one-step estimation procedures, such as those due to Wales and Woodland (1983) and Lee and Pitt (1986), is that they are only appropriate for modeling corner solutions. This is because they assume that the same process that governs the positive observed demand also governs the consumption decision itself. But this is not the only explanation for the presence of zero expenditure levels. A natural alternative would be a procedure that captures other phenomena such as infrequency of purchases. Shonkwiler and Yen (1999) derive such as estimator as a multivariate generalization of Amemiya's (1985) type 2 Tobit model. Their model contains a separate binary censor used to predict the probability of consumption for each good in the system. This is then multiplied by the expectation of demand for the respective good conditional on positive consumption to generate the unconditional censored demand equations.

The Shonkwiler and Yen procedure is carried out in two steps. In the first step, single equation probit models are used to forecast the probability of consumption and construct the second stage demand system, which is subsequently estimated by either maximum likelihood or seemingly unrelated regression. This procedure is particularly suited to situations where large equation systems are involved, given the reduced computational burden associated with it. In this study, we estimate a demand system involving seven food groups. Given the suitability of Shonkwiler and Yen's procedure for large equation systems, and its consistency, it is our chosen approach to model the zero-expenditure problem. A detailed discussion of this procedure follows in the empirical model chapter.

2.6 Chapter Summary

This chapter reviewed literature on the relationship between utility maximization theory and demand functional forms, and discussed alternative approaches to modeling preferences. In empirical demand estimation, a large number of goods are involved, and to reduce this large number of goods into a manageable few, the assumption of weak separability is typically invoked. The developments in the literature on econometric tests for separability have focused mainly on time series applications. The AIDS model has been the most widely used functional form for empirically estimating price and expenditure elasticities. However, its assumptions that budget shares are linear in expenditure, and that expenditure elasticities are constant regardless of the point in the expenditure spectrum, are limiting. This study uses QUAIDS, which is a generalization of AIDS that allows for nonlinearity in expenditure and allows the goods to be luxuries at some expenditure levels and necessities at others. The effects of demographic variables are incorporated into the demand model using demographic translation. The methods that have been developed in the literature to model observed zero-expenditures are, for the most part, not suitable for large equation systems due to their computational complexity. Due to the large number of equations estimated in this study, we use a two-step procedure developed by Shonkwiler and Yen (1999).

CHAPTER THREE

EMPIRICAL MODEL

3.1 Introduction

This chapter discusses the specification and estimation of the empirical model. The general form of the model is discussed first, followed by the estimation form. Section 3.3 discusses estimation issues, paying particular attention to implications of the nonlinearity of the model, and deriving an LM test for nonlinearity. Econometric issues associated with QUAIDS estimation are the focus of section 3.4. In this section, the problems of expenditure endogeneity and non-consumption are discussed in more detail, and the strategies to modeling them are provided. The final section, section 3.5, is a summary of the chapter.

3.2 Empirical Model

Popular functional forms such as the almost ideal demand system (AIDS) of Deaton and Muellbauer (1980a) and the translog model of Jorgenson *et al.* (1982) have budget shares that are linear functions of log total expenditure. However, as discussed in the previous chapter, further terms in total expenditure may be required for some, if not all budget share equations. Banks *et al.* (1997) show that if some commodities require these extra terms, then parsimony, coupled with utility theory, restricts the nonlinear term to be quadratic in log income. Based on this restriction, they derive an extension of the AIDS model—the quadratic almost ideal demand system (QUAIDS)—which has log total expenditure as the leading term in budget share equations and higher order total expenditure terms.

3.2.1 Quadratic Almost Ideal Demand System: The General Form

The QUAIDS model assumes that household preferences belong to the following quadratic logarithmic family of expenditure functions:

$$\ln c(u,\mathbf{p}) \;=\; \ln a(\mathbf{p}) \;+\; \frac{ub(\mathbf{p})}{1-\lambda(\mathbf{p})b(\mathbf{p})u} \tag{3.1}$$

where u is utility, $\mathbf{p}$ is a vector of prices, $a(\mathbf{p})$ is a function that is homogenous of degree one in prices, $b(\mathbf{p})$ and $\lambda(\mathbf{p})$ are functions that are homogeneous of degree zero in prices. The corresponding indirect utility (V) function is:

$$\ln V \;=\; \left\{\left[\frac{\ln x - \ln a(\mathbf{p})}{b(\mathbf{p})}\right]^{-1} + \lambda(\mathbf{p})\right\}^{-1} \tag{3.2}$$

where x is total expenditure. The specific functional form for $\lambda(\mathbf{p})$ is:

$$\lambda(\mathbf{p}) \;=\sum_{i=1}^{K}\lambda_i \ln p_i, \qquad \text{where } \sum_{i=1}^{K}\lambda_i = 0 \tag{3.3}$$

and where $i = 1, \ldots, K$ denote the number of goods entering the demand model. Deaton and Muellbauer's AIDS model has an indirect utility function given by equation (3.2), but with $\lambda(\mathbf{p})$ set to zero. The specification of the functional forms for $a(\mathbf{p})$ and $b(\mathbf{p})$ in QUAIDS is similar to their specification in AIDS, in which they are made to be sufficiently flexible to represent any arbitrary set of first and second derivatives of the cost function.

Application of Shepard's lemma to the cost function (3.1) or Roy's identity to the indirect utility function (3.2) gives the QUAIDS model in budget shares form:

$$w_i = \alpha_i + \sum_{j=1}^{K} \gamma_{ij} \ln p_j + \beta_i \ln\left[\frac{x}{a(\mathbf{p})}\right] + \frac{\lambda_i}{b(\mathbf{p})}\left\{\ln\left[\frac{x}{a(\mathbf{p})}\right]\right\}^2 \tag{3.4}$$

where α, β, γ, and λ are parameters. As can be seen from the budget shares (3.4), the QUAIDS model specializes to AIDS when all of the λ's are zero across all equations. Hence, the AIDS model is nested within QUAIDS, and the AIDS specification can be tested based on the statistical significance of the λ's.

As with the original AIDS model, the theoretical restrictions of adding-up, homogeneity, and symmetry in the QUAIDS model are expressed in terms of its parameters. Adding-up requires $\sum_i w_i = 1$, and can be expressed in terms of model parameters as:

$$\sum_{i=1}^{K} \alpha_i = 1 \quad \sum_{i=1}^{K} \beta_i = 0 \quad \sum_{i=1}^{K} \lambda_i = 0 \quad \sum_{i=1}^{K} \gamma_{ij} = 0 \quad \forall j. \tag{3.5}$$

Since Marshallian demands are homogenous of degree zero in $(\mathbf{p}, x)$,

$$\sum_{j=1}^{K} \gamma_{ij} = 0 \quad \forall i. \tag{3.6}$$

Slutsky symmetry implies that:

$$\gamma_{ij} = \gamma_{ji} \quad \forall i, j. \tag{3.7}$$

The parameter α_i in the QUAIDS model can be interpreted as the share of an item in the budget of a subsistence household (i.e., the case of $u = 0$) at the base year prices (Meenkashi and Ray, 1999). The expression $\beta_i + 2(\lambda_i / b(P))[\ln(x / a(P))]$ measures the impact of a 1% increase in real expenditure on the budget share of commodity *i*. Unlike in the AIDS model where $\lambda_i = 0 \;\forall i$, this expression is capable of changing signs depending on the point in the expenditure spectrum. In other words, the QUAIDS model allows the possibility of normal goods becoming inferior or inferior goods becoming normal, as one moves along the expenditure spectrum of households. In contrast, expenditure elasticities are all constant in the AIDS model.

Formulas for the QUAIDS expenditure and price elasticities are derived by differentiating the budget share equations with respect to $\ln x$ and $\ln p_j$, respectively. Following Banks *et al.* (1997), we simplify the expressions for the elasticity formulas by using the intermediate results:

$$\mu_i \equiv \frac{\partial w_i}{\partial \ln x} = \beta_i + \frac{2\lambda_i}{b(\mathbf{p})}\left\{\ln\left[\frac{x}{a(\mathbf{p})}\right]\right\} \tag{3.8}$$

$$\mu_{ij} \equiv \frac{\partial w_i}{\partial \ln p_j} = \gamma_{ij} - \mu_i\left(\alpha_j + \sum_{l=1}^{K}\gamma_{jl}\ln p_l\right) - \frac{\lambda_i \beta_j}{b(\mathbf{p})}\left\{\ln\left[\frac{x}{a(\mathbf{p})}\right]\right\}^2. \tag{3.9}$$

In terms of the μ_i, the formula for expenditure elasticities can be written as:

$$e_i = 1 + \frac{\mu_i}{w_i}. \tag{3.10}$$

Using expression μ_{ij}, the formula for the Marshallian or uncompensated price elasticities can be written as:

$$e_{ij}^u = \frac{\mu_{ij}}{w_i} - \delta_{ij} \tag{3.11}$$

where δ_{ij} is the Kronecker delta taking the value $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. The Hicksian or compensated price elasticities are calculated by invoking the Slutsky equation:

$$e_{ij}^c = e_{ij}^u + w_j e_i \tag{3.12}$$

The QUAIDS model is used in this study to estimate price and expenditure elasticities using panel data on households. The next subsection focuses on the empirical estimation of the QUAIDS model.

3.2.2 Quadratic Almost Ideal Demand System: The Estimation Form

As before, denote commodities (and therefore, equations) by i, where $i = 1,...,K$, and let h = 1, …, N denote households, and t = 1, …, T index time periods. The empirical specification of the QUAIDS model is

$$w_{it}^{h} = \alpha_i + \sum_{j=1}^{K} \gamma_{ij} \ln p_{jt}^{h} + \beta_i \ln\left[\frac{x_t^h}{a(\mathbf{p}_t^h)}\right] + \frac{\lambda_i}{b(\mathbf{p}_t^h)}\left\{\ln\left[\frac{x_t^h}{a(\mathbf{p}_t^h)}\right]\right\}^2 + \sum_{s=1}^{S} \delta_{is} z_{st}^{h} + \varepsilon_{it}^{h} \tag{3.13}$$

where $\mathbf{z}_s = \left(z_{1t}^{h},...,z_{st}^{h}\right)$ is a set of demographic variables for household h at time t, $\ln a(\mathbf{p}_t^h)$ is the price index defined as

$$\ln a(\mathbf{p}_t^h) = \alpha_0 + \sum_{j=1}^{K} \alpha_j \ln p_{jt}^{h} + \frac{1}{2}\sum_{j=1}^{K}\sum_{l=1}^{K} \gamma_{jk} \ln p_{jt}^{h} \ln p_{lt}^{h} \tag{3.14}$$

and $b(\mathbf{p}_t^h)$ is the Cobb-Douglas price aggregator

$$b(\mathbf{p}_t^h) = \prod_{i=1}^{K} \left(p_{it}^{h}\right)^{\beta_i}. \tag{3.15}$$

The α's, γ's, and β's in the budget share equations (3.13) are restricted by theory to be the same as those in equations (3.14) and (3.15).

In estimating the QUAIDS model, total expenditure, x, is defined as expenditure on all food items consumed by the household. Price data is at the cluster level, which in most cases means at the village level for rural areas or at magisterially-defined districts for urban areas. So, households in different clusters face different prices, and this is the reason commodity prices are indexed with the h (household) superscript.

To control for varying preference structures and heterogeneity across households, demographic variables are incorporated in budget share equations through the linear demographic translation method of Pollak and Wales (1978). This method specifies observed household heterogeneity as a linear combination of socio-demographic variables observed in the data ($\mathbf{z}_{it}^{h}{}'\delta_i$). The socio-demographic variables considered here are household size, rural-urban dummy, race, and education of the household head. Dummy variables for the year of survey are included among the $\mathbf{z}$ variables to control for structural change in consumers' preferences and other aggregate time effects that may influence expenditure patterns (such as those related to the overall macroeconomic environment). The month of the survey is also included to control for the likely effects of seasonality on consumption behavior.

Given that there are existing food demand studies in South Africa based on the AIDS model, it is instructive to also estimate the AIDS model in this study, so that elasticity estimates can be compared with those obtained from QUAIDS. After all, once the unrestricted QUAIDS model has been estimated, the estimation of AIDS becomes a trivial task because it only requires restricting the coefficient on the quadratic expenditure term to zero. The empirical specification of AIDS is:

$$w_{it}^{h} = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln p_{jt}^{h} + \beta_i \ln\left[\frac{x_t^{h}}{a(\mathbf{p}_t^{h})}\right] + \sum_{s=1}^{S} \delta_{is} z_{st}^{h} + u_{it}^{h} \tag{3.16}$$

The β parameters of the AIDS model determine whether goods are luxuries or necessities (Deaton and Mueallbauer, 1980a). When $\beta_i > 0$, an increase in x leads to an increase in w_i so that good i is a luxury. Similarly, $\beta_i < 0$ for necessities. The γ_{ij} parameters measure the change in the ith budget share following a unit proportional change in p_j with $x/a(\mathbf{p})$ held constant. The formula for the AIDS expenditure elasticity is given by[3]:

$$e_i = \frac{\beta_i}{w_i} + 1 \ . \tag{3.17}$$

The Marshallian price elasticities are computed using the formula:

$$e_{ij}^{u} = \frac{\gamma_{ij}}{w_i} - \frac{\beta_i}{w_i}\left[w_j - \beta_j \ln\left(\frac{x}{a(\mathbf{p})}\right)\right] \tag{3.18}$$

where δ_{ij} is the Kronecker delta taking a value of one if $i = j$, and zero if $i \neq j$. The Hicksian price elasticities are obtained by invoking the Slutsky equation:

[3] The household superscript and time subscripts are not included here because the elasticities are calculated at time- and household-pooled sample means.

$$e_{ij}^{c} = e_{ij}^{u} + w_j e_i \ . \tag{3.19}$$

The elasticities in both the QUAIDS and AIDS models are estimated at sample means of prices, expenditures, and budget shares.

3.3 Estimation

For purposes of estimation, an error term, ε_{it}^{h}, is added to each of the commodity share equations. The errors $\boldsymbol{\varepsilon} \equiv [\ \varepsilon_{1t}^{h}, \varepsilon_{2t}^{h}, \ldots, \varepsilon_{Kt}^{h}\ ]$ are assumed to have a multivariate normal distribution with covariance matrix $\mathbf{\Sigma}$. However, due to the adding-up condition, direct estimation of the full equation system is not possible because $\mathbf{\Sigma}$ is singular. To get around this problem, one of the K demand equations is dropped from the system during estimation; the remaining (K-1) equations are estimated by maximum likelihood. The question of which among the K equations to drop is irrelevant because, as Barten (1969) shows, such a choice does not influence the demand parameter estimates. The full covariance matrix, together with the parameters of the K^{th} equation, are recovered by applying the delta method (Barten, 1969).

An interesting econometric feature of both the QUAIDS and AIDS models, is that they are both *conditionally* linear in the price aggregators $\ln a(\mathbf{p}_t^h)$ and $b(\mathbf{p}_t^h)$. This conditional linearity has been used in the AIDS model (in which the only source of nonlinearity is the $\ln a(\mathbf{p}_t^h)$ price aggregator) to simplify empirical estimation. In particular, most demand studies approximate the nonlinear price aggregator $\ln a(\mathbf{p}_t^h)$ by a linear index, which leads to a specification in which budget shares are linear in all parameters and therefore, can be estimated in a straightforward way. The most commonly used approximations for $\ln a(\mathbf{p}_t^h)$ are the Laspeyres index, Stone index, or modified (by Moschini (1995)) Stone index. Another reason for the linear approximation is that in practical applications, prices are relatively collinear, so that $\ln a(\mathbf{p}_t^h)$ is approximately proportional to any appropriately defined price index (Deaton and Muellbauer, 1980a). This latter reason is particularly relevant in demand studies that use time series data.

However, the imposition of a linear structure to variables whose true relationship is nonlinear can have undesirable consequences on the reliability of parameter estimates. Pashardes (1993) argues that the linearization of AIDS causes an omitted variables problem. Based on analytical expressions and empirical results, Pashardes shows that linearization of AIDS can understate own price elasticities and cross price elasticities of goods that are either luxuries or necessities, and overstate the cross price elasticities of the other goods. Buse (1994) also views linearization of AIDS as an omitted variable problem, and shows through Monte Carlo analyses that linearization may lead to inconsistency of the widely used seemingly unrelated regression (SUR) estimator. Both Pashardes and Buse's assessment of linearized AIDS are based on the original Stone price index.

Moschini (1995) shows that the Stone index fails to satisfy the commensurability property of index numbers; in other words, it is not invariant to changes in the units of measurement. Using the Laspeyres price index as a starting point, Moschini develops a price index—the modified Stone price index— which is invariant to units of measurement and which, he argues, approximates the nonlinear AIDS model well. We are not aware of studies that evaluate Moschini's modified Stone price index in a manner similar to those used by Buse and Pashardes to evaluate the original Stone index.

While approximating $\ln a(\mathbf{p}_t^h)$ by a linear price index solves the nonlinearity problem in the AIDS model, it does not solve the nonlinearity problem in the QUAIDS model due to the division by the price aggregator $b(\mathbf{p}_t^h)$ in the coefficient of the quadratic expenditure term. The temptation to include a linear (in parameters) quadratic term in expenditure may be natural. In fact, a number of demand studies force the coefficient of the quadratic expenditure term to be constant (included in these studies are Blundell *et al.* (1993), Christensen (2004), Labeaga and Puig (2002), and Browing and Collado (2004).). In the terminology of equation (2.1) of chapter 2, imposing a constant coefficient on the quadratic expenditure term is equivalent to assuming that $C_i(\mathbf{p})$ is independent of prices. However, as Banks *et al.* (1997) show, no rank 3 exactly aggregable utility-derived demand system exists that has both the coefficients on the linear and the quadratic expenditure terms independent of prices (Corollary 2 on p.533 of Banks *et al.* (1997)). Some of these studies (specifically Christensen (2004)) acknowledge the fact that forcing $b(\mathbf{p}_t^h)$ to be constant is to give away the integrability property of the demand system.[4]

In this study, we build on Banks *et al.*'s (1997) study (particularly corollary 2) and develop a formal test for the statistical significance of prices in the coefficient of the quadratic expenditure term. The logic behind this test is that if this coefficient does not depend on prices, then there is no need to include a quadratic expenditure term in the model once the linear term has been included, because the coefficient of the quadratic

---

[4] Integrability as used here means that for a given system of demand functions (which have a symmetric, negative semidefinite matrix), there should be a utility function from which these demand functions can be derived.

expenditure term must depend on prices. Higher order expenditure terms are also unnecessary because utility theory restricts the nonlinear expenditure term to be quadratic. So, a test of the statistical significance of the quadratic expenditure term is in effect, a specification test of the AIDS versus QUAIDS model. But because AIDS can be approximated linearly and QUAIDS cannot, a test for the statistical significance of prices can also be viewed as a test for nonlinearity of the demand model.

3.3.1 A Test for Nonlinearity of the Demand Model

To derive this test, it is necessary to relax the theoretical constraint in the QUAIDS model that the β_i parameters in the Cobb-Douglas price aggregator $b(\mathbf{p}_t^h)$ are the same as the coefficients on the linear expenditure terms (that is, the coefficient on $x_t^h / a(\mathbf{p}_t^h)$). This is because if we maintain the restriction that the β_i's in $b(\mathbf{p}_t^h)$ are the same as the β_i coefficients on the linear expenditure term, then the null hypothesis that the β_i's in $b(\mathbf{p}_t^h)$ are all zero will make the second term, $\ln(x_t^h / a(\mathbf{p}_t^h))$, in budget share equations to disappear. This will make the demand system to be a function only of the quadratic expenditure term, which is inappropriate. To avoid this problem, define a new price aggregator $b(\mathbf{p}_t^h) = \prod_{i=1}^{K} (p_{it}^h)^{\theta_i}$, where θ_i and β_i are allowed to differ from each other. For ease of exposition, we suppress the household (h) and time (t) subscripts, and absorb all the terms not involving the quadratic expenditure term into the vector $\mathbf{q}$ and their associated parameters (i.e., parameters not involving the θ_i's) into the vector $\boldsymbol{\varphi}$. With this new notation, the expenditure share equations (3.13) can now be expressed as:

$$w_i = g_i(\mathbf{q},\boldsymbol{\varphi}) + \lambda_i\left(\prod_{i=1}^{K} p_i^{\theta_i}\right)^{-1}\left\{\ln\left[\frac{x}{a(\mathbf{p})}\right]\right\}^2 + \varepsilon_i \tag{3.20}$$

We want to test the null hypothesis that the vector of coefficients $\boldsymbol{\theta}$ in $\left(\prod_{i=1}^{K} p_i^{\theta_i}\right)^{-1}$ is identically zero (i.e., H_0: $\boldsymbol{\theta} = \mathbf{0}$). The restricted model (with $\boldsymbol{\theta} = \mathbf{0}$) is easier to estimate than the unrestricted model, which makes the Lagrange Multiplier (LM) test an attractive approach.

Consider maximization of the log-likelihood subject to a set of constraints $c(\boldsymbol{\theta}) - r = 0$. Let κ be the Lagrange multiplier and define the Lagrangean function:

$$\Lambda = \ln L + \kappa\left(c(\boldsymbol{\theta}) - r\right) \tag{3.21}$$

where ln L is the log-likelihood function for commodity i given by:

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left[\frac{(w_i - E(w_i))^2}{\sigma^2}\right]. \tag{3.22}$$

The first derivative of the log-likelihood function with respect to θ_i is:

$$\frac{\partial \ln L}{\partial \theta_i} = \lambda_i\left(\prod_{i=1}^{K} p_i^{\theta_i}\right)^{-2}\left(\ln p_i \cdot p_i^{\theta_i} \cdot \prod_{j \neq i} p_j^{\theta_j}\right)\left\{\ln\left[\frac{x}{a(\mathbf{p})}\right]\right\}^2. \tag{3.23}$$

Evaluated at the null, H_0: $\boldsymbol{\theta} = \mathbf{0}$, the first derivative (3.23) becomes:

$$\frac{\partial \ln L}{\partial \theta_i} = \lambda_i \ln p_i \left\{ \ln \left[\frac{x}{a(\mathbf{p})} \right] \right\}^2 . \tag{3.24}$$

Based on equation (3.24), a test for statistical significance of prices in $b(\mathbf{p})$ reduces to adding price times expenditure-squared interaction terms ($\sum_{i=1}^{K} \ln p_i \cdot \{\ln[x/a(\mathbf{p})]\}^2$) to the demand model that is linear in expenditure (i.e., equation (3.13) with $\lambda_i = 0$) — the unrestricted model — and comparing it with the restricted model, which is just the QUAIDS expenditure share equations (3.13). To carry out this test, we first estimate the restricted model and obtain the residuals. These residuals are then regressed on all variables, including the price times expenditure-squared interaction terms. The R-squared, R_u^2, from this regression is used to compute the LM statistic, $LM = N \cdot R_u^2$. This *LM* statistic follows a Chi-squared distribution with degrees of freedom equal to the number of restrictions being tested. For testing purposes, the translog price aggregator, ln $a(\mathbf{p}_t^h)$, is approximated by the modified Stone price index suggested by Moschini (1995), $\ln a(\mathbf{p}_t^h) \equiv \sum_{i=1}^{K} \overline{w}_{i0} \ln(p_{it}^h)$, where $\overline{w}_{i0} = \frac{1}{N} \sum_{h=1}^{N} w_{it}^h$ is the mean budget share across households in the base period.

The LM test just discussed is useful for preliminary analysis of the data to determine whether the demand model should be specified with a quadratic (QUAIDS) or a linear (AIDS) expenditure variable. An obvious alternative would be to estimate the QUAIDS model and test for the statistical significance of the quadratic expenditure term.[5] However, the QUAIDS model is highly nonlinear and difficult to estimate. This LM test is a useful contribution because it allows one to test parametrically whether or not the quadratic expenditure is necessary, without having to estimate the highly nonlinear QUAIDS model.

3.3.2 Nonlinear Estimation

Given the speed and power of the nonlinear algorithms available today, maximum likelihood estimation of more flexible functional forms of the demand model is feasible. This can improve the precision with which income and price elasticites are measured. For this reason, we estimate the QUAIDS demand model (3.13)-(3.15) in its nonlinear form, allowing flexibility in the price aggregators $\ln a(\mathbf{p}_t^h)$ and $b(\mathbf{p}_t^h)$.

When modeling $\ln a(\mathbf{p}_t^h)$ in the flexible form (3.14), one of the problems is that it is virtually impossible to estimate α_0 empirically. Deaton and Muellbauer (1980b) suggest assigning a value to α_0 prior to estimation. In particular, they propose interpreting α_0 as the outlay required for a minimal standard of living when prices are unity (see Deaton and Muellbauer (1980b, p. 316). However, as Moschini *et al.* (1994) observe, the likelihood function is flat in α_0, so that the actual choice of α_0 does not matter for the approximation properties of the demand model. This implies that the computed elasticities are not affected by the choice of α_0. Moschini *et al.* choose a value of $\alpha_0 = 0$, which proved useful in their context given that it simplified the formulas for

[5] Also, one can use nonparametric methods to analyze the shape of the Engel curves (i.e., relationships between a commodity's budget share and total expenditure), as did Banks *et al.* (1997).

their separability tests. Our choice of α_0 in this study is based on the suggestion by Deaton and Muellbauer, primarily because choosing α_0 in this way has relevance to economic theory.

Apart from its inherent nonlinearity, the QUAIDS model has a very large number of parameters. To reduce the total number of parameters to be estimated, cross-equation restrictions are imposed during estimation. All the nonlinear AIDS and QUAIDS models are estimated by maximum likelihood using Stata, extending the programs written by Poi (2002) for estimating a four-equation demand system with no demographic variables to those that allow for a seven-equation system with demographic variables.

3.4 Econometric Issues

3.4.1 Attrition

A typical concern when using household panel data involves the extent of sample attrition and the degree to which attrition is nonrandom. While attrition is a common concern in any longitudinal study, it is particularly serious for studies conducted in developing countries, due to the generally poor communication infrastructures. Furthermore, the high levels of mobility and long distance migration associated with development are likely to complicate longitudinal survey work in developing countries. Partly offsetting these concerns, however, are the much lower refusal rates typical in developing countries, perhaps reflecting lower opportunity costs of time and possibly different cultural attitudes toward the interviewing process (Deaton, 1997). While a large literature exists in developed countries on the implications on nonrandom attrition, only a few studies have considered this topic in developing countries, perhaps reflecting the relative paucity of panel datasets in developing countries.These studies include those by Alderman *et al.* (2000), Maluccio (2000), and Thomas *et al.* (2001)).

In theory, three factors underlie the level of attrition in a survey: (1) the mobility of the target population, (2) the success with which those who move are followed and re-interviewed, and (3) the number of refusals. Thus, attrition is often closely linked to migration behavior (Maluccio, 2000). In the field, poor effort by enumerators and fieldworkers can also exacerbate attrition.

Attrition in panel surveys can be viewed as a specific type of nonresponse and, from a conceptual viewpoint, many of the insights regarding nonresponse in cross-sectional surveys carry over to panels. Fitzgerald *et al.* (1998) provide a statistical framework for the analysis of attrition bias. They distinguish between two types of sample selections; selection of variables observed in the data, and selection on variables that are unobserved. They develop tests for attrition using the two selection types. While neither of the two attrition/selection types necessarily imposes a bias on estimates, selection on observables is more amenable to statistical solutions. In particular, if one finds that there is attrition in the data, then one can determine whether or not there is selection on observables. Selection on observables basically means sample selection based on variables that are observed prior to attrition (e.g., in the first round of the survey). Even if there is selection on observables, this does not necessarily bias the estimates of interest. Thus, one needs to test for possible attrition bias in the estimates of interest as well.

More formally, assume that what is of interest is a conditional population density $f(y|q)$ where y is a scalar dependent variable and q is a scalar independent variable (an extension to make q a vector does not change to results of the discussion). The model takes the form

$$y = \pi_0 + \pi_1 q + \varepsilon, \qquad y_i \text{ observed if } A = 0 \tag{3.25}$$

where A is an attrition indicator equal to 1 if an observation is missing its value of y because of attrition, and equal to zero if an observation is not missing its value of y. Since (3.25) can be estimated only if $A = 0$ (that is, one can only determine $g(y|q, A = 0)$), one needs additional information or restrictions to infer f(.) from g(.). These can come from the probability of attrition, Prob($A = 0|y, q, z$), where z is an auxillary variable (or vector of variables) that is observable for all units but not included in x. This implies estimations of the form

$$A^{*} = \sigma_0 + \sigma_1 q + \sigma_2 z + \upsilon \tag{3.26}$$

$$\begin{aligned} A_i &= 1 \quad \text{if } A_i^{*} \geq 0 \\ &= 0 \quad \text{if } A_i^{*} < 0 \end{aligned} \tag{3.27}$$

Selection on unobservables occurs if z is independent of $\varepsilon|q$ but υ is not independent of $\varepsilon|q$. Selection on observables is the reverse: it occurs if z is not independent of $\varepsilon|q$ but υ is independent of $\varepsilon|q$. That is, selection on observables occurs if Prob($A = 0|y, q, z$) = Prob($A = 0|q, z$); selection on unobservables occurs if this equality fails to hold, so that the attrition function cannot be reduced from Prob(A = $0|y, q, z$).

Selection on unobservables is often presented as dependent on the estimation of the attrition index equation. Identification, however, usually relies on nonlinearities in the index equation or an exclusion restriction, i.e., some z that is not in q. It is difficult to rationalize most such exclusion restrictions because, for example, personal characteristics that affect attrition might also directly affect the outcome variable, i.e., they should be in q (Alderman *et al*., 2000). There may be some such identifying variables that are external to individuals and not under their control, such as characteristics of the interviewer in the various rounds. However, identifying restrictions are generally not available, which makes selection on unobservables an obstacle to accurate parameter estimation.

If there is selection on observables, the critical variable is z, a variable that affects attrition probability and that is also related to the density of y conditional on q. Two sufficient conditions for the absence of attrition bias due to attrition on observables are either (1) z does not affect A or (2) z is independent of y conditional on q. Attrition tests can be based on either of these two conditions. One test is simply to determine whether candidate variables for z significantly affect A. Another test is based on Becketti, Gould, Lillard, and Welch (BGLW) (1988). In the BGLW test, the value of y at the initial wave of the survey (y_0) is regressed on q and on A. The test for attrition is based on the significance of A in that equation.

The analysis of attrition in this study follows the approaches suggested by Becketti *et al.* (1988) and Fitzgerald *et al.* (1998). In particular, we test for whether or not attrition significantly affects estimated multivariate relations. Our analysis of attrition begins with a comparison of the means of a selected number of key household and community variables. Besides being informative, the comparison of means is also intuitively appealing, because the idea that attrition is likely to bias estimates is often made on the basis of such univariate comparisons (Alderman *et al.*, 2000). We then estimate probits for the probability of attrition in order to investigate what variables predict attrition and determine whether or not the probability of attrition can be explained significantly by observable variables. Finally, we test whether coefficient estimates differ for the two subsamples, one that attrits and one that is re-interviewed. The results from these attrition tests will help us in deciding how to deal with it.

3.4.2 Expenditure Endogeneity

Most empirical demand analyses do not cover all products and services that households purchase. Data limitations, finite computer memory, and the increased complexity and time required for estimating large models make it necessary to abstract from a completely specified demand system containing a different equation for each of the myriad goods available in the market (LaFrance, 1991). The practice is typically to assume that preferences are separable and estimate a set of conditional demands for the goods of interest as functions of prices and total expenditure on these goods (Pollak, 1969). However, such a practice raises questions regarding the possibility of simultaneity bias in the budget share equations. Total expenditure may be determined jointly with the expenditure shares of the individual commodities being analyzed, making it endogenous in the expenditure share equations. Also, expenditure endogeneity issues may arise whenever the household expenditure allocation process is correlated with other unobserved behavior not captured by the explanatory variables in the budget share equations. In this case, these unobserved effects would be bundled in the error term.

Estimation ignoring expenditure endogeneity may lead to inconsistent demand parameter estimates.

In cross-sectional demand studies, the common procedure to control for expenditure endogeneity is instrumental variables. With panel data, a number of possibilities to correct for unobserved heterogeneity are available, including linear transformations of the original model, such as through fixed effects and first differencing to remove the unobserved heterogeneity component of the error term. However, such transformations are difficult to implement with nonlinear models such as QUAIDS derived from consumer utility maximization theory. In this study, we follow Bundell and Robin (1999) and control for endogeneity using an extension of the limited information augmented regression technique suggested by Hausman (1978). This procedure is also known as the *control function approach.*

To illustrate how the augmented regression technique works, consider the regression of y_1, the dependent variable, on a set of exogenous explanatory variables, $\mathbf{z}$, and an endogenous explanatory variable, y_2, i.e., $y_1 = \mathbf{z}'\rho + \pi y_2$.[6] Also, suppose an instrumental variable, z_2, exists for y_2. Correction for the endogeneity of y_2 using the control function approach proceeds in two steps. The first step involves estimating a reduced form regression of the endogenous variable on a set of instrumental variables, where the set of instrumental variables include all the other exogenous explanatory variables (i.e., regress y_2 on $\mathbf{z}$ and z_2). The residuals, $\hat{v}$, from this first-stage regression are then included as an additional explanatory variable in the original y_1 equation. The OLS estimates of the parameters ρ and π in this augmented regression are identical to the Two-

[6] For illustration purposes, we consider the case of *one* endogenous variable and *one* instrumental variable. The case of multiple endogenous variables and multiple instruments can be handled in a straightforward way using the basic framework explained here.

Stage Least Squares (2SLS) estimator (Blundell and Robin, 1999). Moreover, testing for the significance of the coefficient on $\hat{v}$ is a test for the exogeneity of y_2. Following Banks *et al.* (1997), we use total household income and its square as instruments for expenditure (and expenditure squared).

3.4.3 Observed Zero Expenditures

In each equation of the QUAIDS model, the dependent variable w_{it}^h is observed with nonnegative values. In situations where micro data are used, it is very likely to observe non-consumption of some commodities due to purchase infrequency and corner solutions. If a nonnegligible proportion of the w_{it}^h values are identically zero, then the w_{it}^h variable becomes partly continuous with a positive probability mass at zero. OLS regression using the subsample for which $w_{it}^h > 0$ estimates the demand parameters inconsistently due to nonrandom sample selection problem, while OLS using all of the data will not consistently estimate the demand parameters due to the nonlinearity in the conditional mean of w_{it}^h (for a general discussion see Wooldridge (2002), pages 524-525).

In the case of a single-equation demand model, censoring in the dependent variable can be handled in a straightforward way by applying a maximum likelihood (ML) Tobit model. However, when estimation is for systems of equations, and censoring occurs in multiple equations, then direct estimation by ML becomes difficult because of the need to evaluate multiple integrals in the likelihood function. To a large extent, this problem of having to evaluate multiple integrals explains why many of the theoretical models discussed in chapter 2 have seen virtually no use in empirical demand studies.

The two-step procedure proposed by Heien and Wessells (1990) offered great promise as a solution to the computational infeasibility of these models, but as was shown by Shonkwiler and Yen (1999), it is inconsistent due to a mathematical error in its derivation. In this study, we apply the consistent two-step procedure developed by Shonkwiler and Yen, which corrects for the inconsistency associated with the Heien and Wessells procedure.

To introduce the Shonkwiler and Yen procedure, consider a structure in which censoring of each commodity i at time t is governed by a separate stochastic process $\mathbf{z}_{it}^{h}{}'\boldsymbol{\tau}_i + v_{it}^h$ such that

$$\begin{aligned} w_{it}^h &= w_{it}^h\left(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}\right) + \varepsilon_{it}^h & \qquad \text{if } \mathbf{z}_{it}^{h}{}'\boldsymbol{\tau}_i + v_{it}^h > 0 \\ &= 0 & \qquad \text{otherwise} \end{aligned} \qquad (3.28)$$

where w_{it}^h is observed expenditure share for hth household, $\boldsymbol{\psi}$ is a vector containing all parameters in a particular demand equation, $\mathbf{z}_{it}^h$ is a vector of exogenous variables, $\boldsymbol{\tau}_i$ is a conformable vector of parameters, and ε_{it}^h and v_{it}^h are random errors; $\mathbf{p}$ and m are interpreted as before. Assume that the vector of disturbances $\boldsymbol{\varepsilon}_t^h = \left[\varepsilon_{1t}^h, \varepsilon_{2t}^h, \ldots, \varepsilon_{Kt}^h\right]$ and $\mathbf{v}_t^h = \left[v_{1t}^h, v_{2t}^h, \ldots, v_{nt}^h\right]$ are normally distributed. Correlation is allowed only between $\mathbf{z}_{it}^{h}{}'\boldsymbol{\tau}_i + v_{it}^h$ and $w_{it}^h\left(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}\right)$ for each commodity and among $w_{it}^h\left(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}\right)$ and

$w_{jt}^h(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi})$, $i \neq j$.[7] Using equation (3.28) and the bivariate normality of $[\varepsilon_{it}^h, v_{it}^h]$, the mean of w_{it}^h conditional on a positive observation is

$$E\left(w_{it}^h \middle| v_{it}^h > -\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i\right) = w_{it}^h(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}) + \delta_i \frac{\phi(\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i)}{\Phi(\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i)} \tag{3.29}$$

where ϕ(.) and Φ(.) are the standard normal probability density and distribution functions, respectively. Based on the facts that $\text{Prob}(v_{it}^h > -\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i) = \Phi(\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i)$ and $E\left(w_{it}^h \middle| v_{it}^h < -\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i\right) = 0$, the unconditional mean of w_{it}^h is $E(w_{it}^h) = \Phi(\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i) w_{it}^h(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}) + \delta_i \phi(\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i)$. Based on $E(w_{it}^h)$, the system of share equations can be written as

$$w_{it}^h = \Phi(\mathbf{z}_{it}'\boldsymbol{\tau}_i) w_{it}^h(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}) + \delta_i \phi(\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i) + \xi_{it}^h \tag{3.30}$$

where $\xi_{it}^h = w_{it}^h - E(w_{it}^h \mid \mathbf{p}_t^h, m_t^h, \mathbf{z}_{it}^h)$. System (3.30) can be estimated in two steps: (i) first, obtain the maximum-likelihood probit estimates $\hat{\boldsymbol{\tau}}_i$ of $\boldsymbol{\tau}_i$ using the binary outcomes $w_{it}^h = 0$ and $w_{it}^h > 0$, and then (ii) calculate $\phi(\mathbf{z}_{it}^h \hat{\boldsymbol{\tau}}_i)$ and $\Phi(\mathbf{z}_{it}^h \hat{\boldsymbol{\tau}}_i)$ for all *i* and estimate $\boldsymbol{\psi}$, δ_1, δ_2, ..., δ_n in the augmented system

---

[7] That is, the binary censoring mechanism $\mathbf{z}_{it}^{h\prime}\boldsymbol{\tau}_i + v_{it}^h$ is independent of $\mathbf{z}_{jt}^{h\prime}\boldsymbol{\tau}_j + v_{jt}^h$ and the level equations $r^{\perp}$ for i $\neq j$.

$$w_{it}^h = \Phi(\mathbf{z}_{it}^{h}{}'\hat{\boldsymbol{\tau}}_i) w_{it}^h(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}) + \delta_i \phi(\mathbf{z}_{it}^{h}{}'\hat{\tau}_i) + \xi_{it}^h \quad (3.31)$$

by ML or SUR. Elasticities can be calculated as in equations (3.8) – (3.11), except the intermediate derivatives μ_i, μ_{ij} are now replaced by $\bar{\mu}_i, \bar{\mu}_{ij}$

$$\bar{\mu}_i = \partial E(w_i)/\partial \ln m = \Phi(\mathbf{z}_i{}'\hat{\boldsymbol{\tau}}_i)\partial w_i / \partial \ln m \quad (3.32)$$

$$\bar{\mu}_{ij} = \partial E(w_i)/\partial \ln p_j = \Phi(\mathbf{z}_i{}'\hat{\boldsymbol{\tau}}_i)\partial w_i / \partial \ln p_j \quad (3.33)$$

which are obtained by taking first derivatives of equation (3.30). Shonkwiler and Yen's procedure has seen use in some recent food demand studies (Dong *et al.*, 2004; Yen *et al.*, 2004; Aguero and Gould, 2003).

3.5 Chapter Summary

The chapter discussed how the empirical model will be estimated. The demand parameters and price and expenditure elasticities will be estimated using the QUAIDS functional form. For comparison with previous demand studies in South Africa, the demand parameters will also be estimated using the AIDS functional form. The statistical significance of prices in the coefficient of the quadratic expenditure term will be tested using the LM test. Demographic translation will be used to incorporate demographic variables into the demand model. The endogeneity of expenditure will be corrected for using the augmented regression technique. The two-step procedure suggested by Shonkwiler and Yen (1999) will be used as a corrective procedure for the zero-expenditure problem.

CHAPTER FOUR

DATA SOURCES AND DESCRIPTION

4.1 Introduction

This chapter discusses the data sources used in this study, and presents summary statistics of the variables entering the demand model. Section 4.2 describes the data and methods used to collect them. Section 4.3 provides background information on KwaZulu-Natal Province, focusing on the economic and social reforms that occurred between the three panel waves. Section 4.4 provides a descriptive analysis of the expenditure patterns of households, and presents summary statistics of their socioeconomic and demographic characteristics. The extent of the zero-expenditure problem is discussed in section 4.5, and a summary if the chapter provided in the final section.

4.2 Surveys and Data Description

Data to be used in this study comes from the KwaZulu-Natal Income Dynamics Study (KIDS). KIDS is a panel dataset comprising three surveys: the 1993 Project for Statistics on Living Standards and Development (PSLSD) survey, and the 1998 and 2004 surveys which conducted further interviews on households from the 1993 PSLSD survey who resided in KwaZulu-Natal Province.

The PSLSD is a nation-wide survey undertaken in the last half of 1993 by a consortium of South African groups and universities under the leadership of the South African Labor and Development Research Unit at the University of Cape Town, with technical expertise from the World Bank. The main instrument was a comprehensive household survey collecting a broad array of information on the socio-economic condition of households. The topics covered included household demographics, education, food and non-food expenditures, remittances, employment and income, agricultural activities, health and anthropometry.

In addition to the household questionnaire, a community questionnaire was administered in each cluster to collect information common to households, such as the availability of schools and health facilities. The community questionnaire also collected data on prices for a detailed list of food products commonly purchased by households. These prices were collected from at least two sources in or near each community. The first source was a formal retail store such as a supermarket, and the second source was a less formal business such as a "corner café" (corner café's are commonly referred to as *spaza* shops in South Africa).[8] In this study, we calculate the price for each food product as the average of prices from the two sources.

The 1993 sample was selected using a two-stage, self-weighting design. In the first stage, clusters were chosen proportional to size from census enumerator districts. The census enumerator districts were based on the 1991 population census. In the second stage, all households in each chosen cluster were enumerated and then a random sample selected. Nationwide, a total of 358 clusters were surveyed, and information on 8848 households was collected.

In 1998, households surveyed by the PSLSD in KwaZulu-Natal Province were reinterviewed in the KIDS survey. The 1998 KIDS survey was undertaken by a

[8] A *spaza* shop is a small store typically owned by an individual or a household, selling frequently purchased consumer items such as food, body care products, and alcoholic beverages. These forms of business are found in black townships and villages, and are operated from owners' homes or in such places as busy roads or intersections and taxi or bus stations.

consortium including the University of Natal, the University of Wisconsin-Madison, and the International Food Policy Research Institute.[9]

The 1993 KwaZulu-Natal portion of the PSLSD sample was representative at the province level (conditional on the accuracy of the 1991 census and other information used in the sampling frame) and contained households of all races. In the 1998 KIDS survey however, it was decided not to reinterview white and coloured households. The reason was that the proportion of white and coloured households in the sample was small (7% were whites and 3% coloureds), which precluded comparative ethnic analyses. Another reason given was that white and coloured households were located in a small number of clusters (due to the general lack of spatial integration of races in South Africa), which appear to be non-representative at the ethnic group level.

The 1998 household questionnaire largely followed the 1993 version, except that the former was appended to include aspects of asset ownership and economic shocks (as was driven by the focus of the study, on "income dynamics"). The 1998 target sample (i.e., the sample that would have resulted in the absence of attrition) was 1,354 households, of which 1,171 (86 percent) were successfully re-interviewed—success here defined as having re-interviewed at least one member from the 1993 household. The 1998 survey tracked and interviewed households that had moved or migrated to newer locations. If a member split from the 1993 household (for example, to establish his/her own household), this member was tracked for reinterviewing if he/she was regarded as "core" in 1993. A core member is defined as one who had influence in household

[9] There were no specific reasons for choosing to focus only on KwaZulu-Natal (and not on other provinces); "[t]he choice of KwaZulu-Natal was in part the result of practical considerations including a confluence of research interests, resources, and the feasibility of locating the households interviewed in 1993" (1998 KIDS Codebook)

decision-making in 1993, such as older and/or working children of the household head. Of the 1,171 households that were reinterviewed, 1,058 were the original 1993 households. The interviews were undertaken during a three-month period, from March to June.

The third wave of the KIDS survey was undertaken in 2004. The survey was undertaken by a consortium including the Universities of waZulu-Natal and Wisconsin-Madison, the International Food Policy Research Institute (IFPRI), the London School of Hygiene and Tropical Medicine (LSHTM), and the Norwegian Institute of Urban and Regional Studies (NIBR).

The structure of the 2004 questionnaire was very similar to 1998, except new modules were introduced to collect information on children's literacy rates and household deaths. Similar to the 1998 survey, the 2004 survey tracked and interviewed core members of the 1998 households. As a result, a total of 1428 households were interviewed, 727 of which were interviewed in both the 1993 and 1998 surveys. Most of the interviews were completed during the months of April and July. We restrict our analysis to the households who were interviewed in all the three surveys, so that our analysis is based on a total of 2,181 (3×727) data points. This has the advantage of ensuring greater variability in prices and income faced by each household, and hence, enhances the identification of demand behavioral parameters.

The next section provides a brief overview of the socioeconomic conditions of the KwaZulu-Natal Province, as well as the economic and social reforms that occurred between the three KIDS waves.

4.3 Background Information: KwaZulu-Natal Province

KwaZulu-Natal, the most populous province in South Africa, is home to approximately 20% of South Africa's population of 44 million and was formed in 1994 by combining the former Zulu homeland with the old Natal province.[10] Although not the poorest province in South Africa, it arguably has the highest incidence of deprivation in terms of access to services and perceived well-being (Carter and May, 2001; Leibbrandt and Woolard, 1999). The province's urbanization rate of 42 percent is relatively high compared to poorer provinces such as the Eastern Cape (37 percent) and Limpopo Province (11 percent). KwaZulu-Natal is also home to most of South Africa's ethnically Indian people who constitute 14% of the province's population. Black Africans constitute about 76% of the province's population, with people of European descent (largely British) and coloureds constituting 7% and 3%, respectively.[11]

The economic, social, and racial stratification of KwaZulu-Natal mirrors that of the country as a whole: the province includes a wealthy metropolitan area, Durban, poor townships surrounding it, and a poor and largely rural former homeland, KwaZulu. Poverty and inequality in the province are similar to those at the national level (Woolard *et al.*, 2002), so that reasonable generalizations can be made about the rest of the country based on findings from this study.

The period covered by the KIDS data, coincides with the switch from minority to democratic governance in South Africa. Since its inception in 1994, the democratic government's orientation toward addressing the problems of poverty and inequality

---

[10] During the racially segregationist apartheid government, homelands were created in South Africa as reservation areas for black Africans. Infrastructural developments are usually low, and poverty rates high in the homelands compared to the rest of the country.

[11] The population of South Africa has been historically categorized into five racial groups: blacks, whites, Indians, coloureds, and Asians. Coloureds are defined as people of "mixed" race.

underwent marked shifts. Aggressive programs were introduced to improve access to shelter, sanitation, and education to the previously segregated black communities. In 1996, the government introduced a program known as Growth, Employment and Redistribution (GEAR) which focused on macroeconomic stabilization and structural adjustments. Under GEAR, many of South Africa's sectors were reformed, agriculture being among the most aggressively reformed. For instance, the Agricultural Marketing Act of 1996 called for closure of state supported commodity marketing boards and the termination of all forms of subsidies to agriculture. Other major reforms were carried out in the industrial, labor, and finance sectors.

Empirical evidence points to the deepening of inequality in post-apartheid South Africa (Adato *et al.*, 2004). In KwaZulu-Natal, the Gini coefficient grew from 0.38 in 1993 to 0.42 in 1998. The poverty headcount in KwaZulu-Natal during the same period increased from 27% to 42% (Carter and May, 2001).

In summary, the KIDS survey covers a period of significant social and economic changes in the South Africa. Hence, the KIDS dataset provides a unique opportunity to learn about the household behavior in a period of rapidly changing economic and social environment.

4.4 Descriptive Statistics

The questionnaires from each of the three survey waves contain two sections on household expenditures. The first section collected expenditure information on broad consumption goods (such as food, housing and clothing), and the second collected information on food expenditure and food prices.

Table 4.1 Average Budget Shares of Broad Consumption Goods

Consumption Good	Year 1993	1998	2004	Budget Share Change 1993-2004	% Change 1993-2004
Food	0.52	0.40	0.37	-0.15	-29
Housing	0.11	0.18	0.16	0.05	45
Clothing	0.04	0.03	0.03	-0.01	25
Transportation	0.06	0.06	0.07	0.01	17
Health	0.01	0.01	0.01	0	0
Education	0.02	0.03	0.04	0.02	100
Utilities (energy, water, etc.)	0.10	0.10	0.10	0	0
Personal Items	0.06	0.08	0.09	0.03	50
Insurance	0.02	0.02	0.03	0.01	50
Remittances	0.02	0.01	0.02	0	0
Other expenditure	0.04	0.06	0.07	0.03	75

Table 4.1 presents a list of the broad consumption goods and their expenditure shares for each of the panel years. The budget share for food has consistently declined across the panel years. This may be the result of a variety of factors, including the possibility of structural change. We include year dummies to account for the possibility of structural change. These dummies will, of course, capture the effects of other aggregate time effects, not just structural change. The budget share on food is consistently the highest category during each panel year. The decline in the share of food in the households' budgets may be reflection of Engel's law, which holds that when a family's income increases, the proportion of money spent on food decreases.

The average household (nominal) income for 1993 was 1444.08 rands (R), and it grew to R2757.66 and R3797.71 in 1998 and 2004, respectively.[12] This represents nominal income growth of over 150% during the 1993-2004 sample period. Expenditure shares on education, housing and transportation experienced slight increases during the sample period, while the share on clothing decreased.

The food expenditure information was collected by asking households if they had eaten a particular food item in the month preceding the survey, and for the items eaten, the amount spent on each item was requested. To maintain a reasonable number of parameters, these food items are grouped into seven commodities: grains; meat and fish; dairy products; fruits and vegetables; oils, butter, and fats; sugar and sugar products; and other foods. In grouping these food items, an attempt was made to place goods that are close substitutes in the same group whenever possible, in accordance with the commodity grouping theories discussed in chapter 2. Table 4.2 lists the food groups and the individual items in each group.[13] The budget shares of each of these food groups and their average prices at each of the panel years are listed in Table 4.3.

Throughout the three panel years, grains constituted the largest share of households' total food expenditure. Maize, wheat flour, and bread, all in the grains food group, are staples in South Africa.

---

[12] Rand (R) is the South African currency; the exchange rate between South African Rand (SAR) and United States Dollar (USD) averaged 6.50 SAR = 1 USD between March 2005 and March 2006 (Statistics South Africa).

[13] Alcohol products are not included among the list of food items in Table 4.2 because no data were collected on alcohol consumption. Thus, we are effectively assuming that the direct utility function is weakly separable between alcohol and other foods. Although this assumption may be reasonable in some cases, its violation may lead to econometric problems of expenditure endogeneity. However, we explicitly test for expenditure endogeneity in the next chapter, which may address this problem, at least in part.

Table 4.2 The Composition of Composite Food Commodity Groups

Food Group	Food Items
Grains	maize grain (samp), maize flour, rice, white bread, brown bread, wheat flour, breakfast cereal
Meat and fish	mutton, beef, pork, chicken, fresh fish, tinned fish
Fruits and vegetables	dried peas, lentils, beans, potatoes, madumbes, sweet potatoes, pumpkin, squash, carrots, cabbage, tomatoes, bananas, apples, citrus fruit
Dairy products	fresh milk, sour milk, yoghurt, milk powder, cheese
Oils, butter, and fats	cooking oil, margarine, butter
Sugar and sugar products	Sugar, soft drinks
Other	baby formula, salt

Table 4.3 Average Budget Shares and Prices of the Food Groups

	Average Budget Shares			Budget Share-Weighted Price Indices[1]		
Commodity	**1993**	**1998**	**2004**	**1993**	**1998**	**2004**
Grains	0.35	0.32	0.31	0.38	1.09	7.18
Meat and fish	0.21	0.25	0.22	1.06	2.57	14.00
Fruits and vegetables	0.17	0.17	0.19	0.34	1.52	4.99
Dairy products	0.07	0.05	0.08	1.63	1.48	17.09
Oils, butter, and fats	0.06	0.05	0.05	1.09	2.31	4.80
Sugar and sugar products	0.05	0.05	0.04	0.40	1.41	5.91
Other	0.09	0.11	0.11	0.55	2.14	15.12
Total food expenditure	*R 715.35*	*R 693.25*	*R 1013.27*			

[1]. With the exception of fresh milk, sour milk, cooking oil, and soft drinks, the prices for all the food items are Rands per kilogram; prices for the former items are Rands per litre.

Meats and vegetables are typically consumed jointly with cooked maize meal, which partly explain their relatively large share in households' food budgets.

The share of grains in the households' total food expenditure decreased slightly between 1993 and 2004, which provides some evidence of Bennet's law, which holds that households switch from less to more expensive calorie consumption as their incomes rise. The budget shares of meat products, a more expensive source of calories, rose slightly during the same period. Total expenditure on food (last row in Table 4.3) decreased between 1993 and 1998, but increased sharply between 1998 and 2004. This is partly due to the unusually large increase in food prices between 2000 and 2002. In fact, prices for many of major food products rose by over 100%. Concerns by consumer groups and the government about possible price manipulation and unfair use of market power even led to the establishment in 2002 of a commission of enquiry into food pricing in South Africa (FPMC, 2003).

Income-groups and rural-urban differences

As has been mentioned in chapter one, income distribution is South Africa is highly uneven. Pooling incomes of all households (i.e., ignoring this income inequality) may obscure important information on differences in the behavioral patterns of households. An understanding of the differences in the food expenditure patterns of high and low income households also has important implications for such policies as food pricing reforms and safety nets. For these reasons, in this study, we divide households into three income groups (low, middle, and high), and analyze the consumption behavior of households in each group. Households are ranked from lowest to highest based on their average (averaged over the three panel years) real incomes and divided into three groups of equal sizes; the national consumer price index is used to convert incomes from nominal to real. The average budget shares of the three income groups are reported in Table 4.4.

Table 4.4 Income-groups and Rural-urban differences: Budget shares and prices

	Average Budget Shares					Mean Prices	
	Income Groups			Regions			
Commodity	*Low*	*Middle*	*High*	*Rural*	*Urban*	*Rural*	*Urban*
Grains	0.37	0.35	0.26	0.37	0.24	4.15	4.68
Meat and fish	0.21	0.22	0.26	0.21	0.27	7.05	7.20
Fruits and vegetables	0.18	0.18	0.16	0.18	0.17	7.39	7.86
Dairy products	0.05	0.06	0.08	0.05	0.09	6.98	6.69
Oils, butter, and fats	0.05	0.05	0.05	0.05	0.05	8.18	7.87
Sugar and sugar products	0.06	0.05	0.03	0.05	0.04	8.87	8.74
Other	0.07	0.08	0.14	0.08	0.14	9.38	9.59
Mean household expenditure	584.97	746.51	1089.90	725.94	967.34		
Mean household income	774.34	1696.70	6064.22	1981.55	4553.35		

[1] With the exception of fresh milk, sour milk, cooking oil, and soft drinks, the prices for all the food items are Rands per kilogram; prices for the former items are Rands per litre.

It is clear from Table 4.4 that expenditure patterns differ across households in different income groups. Low-income households spend about 42% more of their total food budgets on grains than high-income households. This is to be expected, given the role of grains as a relatively inexpensive source of calories. On the other hand, high-income households spent more on meats and dairy products, relatively expensive sources of calories. The differences in incomes between high and low income households are also very high; the average income of high-income households is over seven times larger than that of low-income households. Table 4.4 also reports the average budget shares of rural and urban households, as well as rural-urban prices. On average, urban incomes are more than two times rural incomes, and urban households have larger (over 30%) budgets for food than rural households. This explains, in part, the low share of grains and the high share of meat and dairy products in the food budgets of urban households. Part of the reason urban households spend more on food than rural households is that prices of major foods (namely, grains, meat/fish, and fruits/vegetables which constitute about 70% of the food budget) are more expensive in urban than rural areas.

The socioeconomic and demographic characteristics of households play an important role in determining their demand patterns. The demand model will be estimated incorporating these variables to control for varying preference structures and heterogeneity across households. The summary statistics of the socioeconomic and demographic variables considered are reported in Table 4.5.

Table 4.5 Summary statistics of household composition variables

Variable		Income Groups			Rural/Urban	
	Pooled	**Low**	**Middle**	**High**	**Rural**	**Urban**
Household size	6.8	6.4	7.4	6.6	7.3	5.7
Education of head	5	3.1	3.9	7.8	3.8	7.3
Age of household head	54.8	53.7	57.0	53.4	55.6	52.7
Proportion male headed	0.6	0.5	0.6	0.7	0.6	0.6
Proportion rural	0.7	0.9	0.7	0.4	-	-
Proportion black (race)	0.8	1.0	0.9	0.6	1.0	0.6

On average, urban households are of smaller sizes and headed by younger males with high levels of education. Most of these characteristics are shared by high-income households, except the latter have larger family sizes. The rural and low-income groups comprise mainly black households. Overall (i.e., in the pooled sample), black households constitute 85% of the total sample. These statistics are in line with those for the entire KwaZulu-Natal Province; the proportion of blacks in the province is 82 %, and 57 % of the province's population resides in rural areas.

4.5 Zero-Expenditure

When modeling demand using micro data, it is typical to observe a significant number of households that purchase zero quantity of some of items (or commodities) during the survey period. A sample with a large number of non-purchasing households poses a number of econometric challenges. One of the challenges arises when price data are not available and unit values (calculated as purchase value divided by quantity) are used in place of "prices". In that case, the problem becomes one of missing price and quantity data. In cases where price data are available, a large percentage of zero consumptions requires estimation techniques that jointly capture the behavior of both consuming and non-consuming households. Table 4.6 reports the percentages of households with zero consumption in our sample. Clearly, the problem of zero consumption is not severe in the food groups considered here. This is not surprising, given that each of these food groups is an aggregate of several food items.

Table 4.6 Proportions of households with zero consumption for various food groups

	Proportions of Zero Purchase						
Groups	**Grains**	**Meats/fish**	**Fruits/Veg.**	**Dairy**	**Fats/Oils**	**Sugar**	**Other**
Rural (*n* = *1446*)	0.00	0.02	0.00	0.18	0.03	0.01	0.01
Urban (*n* = *735*)	0.00	0.01	0.00	0.05	0.01	0.01	0.00
Low (*n* = *729*)	0.00	0.03	0.00	0.24	0.04	0.02	0.02
Middle (*n* = *723*)	0.00	0.02	0.00	0.12	0.03	0.01	0.00
High (*n* = *729*)	0.00	0.01	0.00	0.05	0.00	0.00	0.00
All (pooled) sample *(n = 2181)*	*0.00*	*0.01*	*0.00*	*0.14*	*0.02*	*0.01*	*0.01*

Typically, the seriousness of the zero-expenditure problem increases with the level of disaggregation in the food groups. The dairy food group is the only one with a significant number of non-consuming households, particularly in the rural and lower-income samples. Appropriate econometric techniques (explained in chapter three) will be applied to the estimation of the budget share equation for dairy.

4.6 Chapter Summary

This chapter discussed the sources of the data to be used in estimating the demand model. Descriptive statistics of the variables entering the model are presented. Food is found to constitute the largest share of the households' total expenditure. Within the food commodity, the budget share of grains is the largest, although this share decreases at higher income levels. Expenditure patterns differ among rural and urban households, especially in the grains, meat and fish, and dairy commodities, with rural households consuming more grains and less of the latter. Observed zero-expenditures are found to be significant only in the dairy food group.

CHAPTER FIVE

EMPIRICAL RESULTS

5.1 Introduction

This chapter presents results from estimating the demand model. The first section tests for the presence of nonrandom attrition. Section 5.3 presents the results of the LM tests for quadratic expenditure specification and expenditure endogeneity. The demand estimation results are discussed in section 5.4, focusing mainly on the expenditure and price elasticity estimates. Heterogeneity in the food expenditure patterns of rural versus urban households, as well as across income groups, are analyzed in section 5.5. The demand estimation results with correction for zero expenditures are presented in section 5.6. Section 5.7 presents estimates of the welfare effects of an indirect tax reform based on different model specifications, and estimates the biases that result when due model misspecification. Section 5.8 summarizes and compares elasticity estimates from this study with those from existing demand studies in South Africa. The last section, section 5.11, summarizes the key findings from this chapter.

5.2 Attrition

This section analyzes and tests for attrition startig with a simple comparison of means of selected key variables for two subsamples, one for households that were re-interviewed in all the three panel rounds and the other for households that subsequently drop from the sample. Although not a formal test for attrition, this univariate comparison of means will give an idea of the extent to which the unconditional means of variables differ among attritors and nonattritors. In order to determine the household and community characteristics that explain which households attrit and which do not in a conditional mean sense, a probit model for the probability of attrition will be estimated. The results of the multivariate probit estimates and the univariate means-comparison tests will be compared in terms of the significance of the household and community characteristics in explaining attrition, and the direction of these effects on attrition. To determine whether or not attrition affects consistent estimation of the coefficients of interest, we estimate a reduced form for the household food expenditure function, and test for whether there are significant differences between the coefficient estimates for households that attrit and those that stay. The base sample for these analyses includes black and Indian households that were interviewed in 1993, since these were the two racial groups targeted for re-interview in the 1998 and 2004 surveys. The total number of households in the 1993 sample is 1241, and 727 of these were re-interviewed in both the 1998 and 2004 surveys.

5.2.1 Difference-of-means tests for major outcomes and control variables

The means compared here are of household and community characteristics as well as other control variables measured in the 1993 survey. The variables selected for this purpose are either those that are most relevant to this study, in the sense that they are related to household food expenditure allocation decisions, or those that are expected to influence attrition. A list of these variables is presented in Table 5.1. The first column of Table 5.1 reports the means and standard deviations for households that are re-interviewed and the second column for those that attrit. The third column lists the differences in the means of the nonattritor and attritor samples, with the results of the t-tests for equality of these means reported in parentheses.

Table 5.1 Differences-of-Means Tests between the Attritors and Nonattritors in KIDS 1993

	Nonattritors (B)		Attritors (A)		Difference (B - A)	
	Means	(S.D.)	Means	(S.D.)	In Means[2]	(t-test)[1]
Household Characteristics						
Household size	6.956	(3.526)	5.066	(2.945)	1.889***	(10.25)
Race (1 if black)	0.855	(0.351)	0.794	(0.405)	0.062***	(2.79)
Gender of head (1 if male)	0.673	(0.469)	0.687	(0.464)	-0.014	(-0.53)
Education of head	3.645	(3.664)	4.196	(4.101)	-0.551**	(-2.43)
Age of head	52.35	(13.881)	49.25	(15.199)	3.098***	(3.66)
Per capita monthly food expenditure	119.98	(72.338))	139.58	(90.517)	-19.598***	(-4.07)
Per capita total monthly income	274.81	(837.88)	334.33	(468.13)	-59.52	(-1.59)
Community Characteristics						
Rural (1 if rural)	0.663	(0.473)	0.632	(0.021)	0.031	(1.11)
Mean per capita food expenditure	126.19	(35.81)	130.78	(41.10)	-4.59**	(-2.04)
Mean per capita income	264.37	(281.66)	349.19	(419.24)	-84.72**	(-3.99)
Number of schools	2.756	(2.131)	2.276	(1.766)	0.480***	(4.32)
Health facility (1 if available)	0.554	(0.497)	0.535	(0.022)	0.019	(0.67)
Sample size	727		514			

[1.] two-sample t-tests with unequal variances. [2.] The asterisks *** indicate significance at 1 percent level; ** and * indicate significance at 5 % and 10 %, respectively.

Based on the signs of the differences in means, it can be seen that on average households that were most likely to be re-interviewed are black, from rural areas, with larger sizes, and an older head. Households headed by males, with lower levels of education, and residing in communities with low mean per capita food expenditure and income had a low likelihood of being re-interviewed.

These results confirm a priori expectations. For instance, in the case where attrition was due to migration, one would expect large-size households to be less mobile, given the larger costs of relocating a family with many members. Higher income facilitates mobility, and this effect is likely to be strengthened by factors such as education and gender, because households headed by an educated male have higher incomes on average.[14] The number of schools in a community and the presence of a health facility increased the likelihood of re-interview, partly due to the role they play as primary sources of information for the survey teams. The means for household size, education and age of household head, race, total monthly food expenditure, community mean food expenditure and income differ significantly (at the 5 percent level) between the sample of households that are re-interviewed and those that attrit.

The problem with the foregoing univariate analysis is that it compares the unconditional means of variables, and does not take account of the relationships among variables. However, because interest in this study is in the relationships among variables, it is more appropriate to examine the presence of non-random attrition in a multivariate framework. The next section focuses on the question of whether or not attrition can be predicted by the observable household and community variables, and on whether or not the effects of each of these variables are attenuated when other socioeconomic variables are conditioned-on in a regression framework.

---

[14] However, since the 1998 and 2004 KIDS surveys tracked households that moved, other factors may offset this effect of income on migration; for instance, wealthier households are likely to be well known in their communities, so that their departure from the community is likely not to be without trace.

5.2.2 A probit model for probability of attrition

In this subsection, we estimate a probit model for the probability of attrition in order to determine whether attrition occurred in a nonrandom manner, and to determine what variables predict attrition. The dependent variable in the probit model is a binary variable of whether or not a household is re-interviewed in all the three surveys (=1 if the household is re-interviewed; =0 if not).[15] The explanatory variables considered include those that were in the univariate difference-of-means tests, as well as additional household and community variables that are expected to affect re-contact and re-interview. These additional variables are: a binary variable for whether or not a household owned its place of residence in 1993, the month of the survey, a binary variable for whether or not any member of the household has been a victim of crime in the 12 months prior to the survey, and dummy variables for supervisor codes (there was total of eight supervisors in the 1993 survey). The dummy variables for supervisor codes control for the quality of the fieldwork, since the ability to successfully track a respondent in subsequent periods depends on the accuracy and detail of information on the base period questionnaires.

The first column of Table 5.2 reports coefficient estimates of the probit model. Household size, race, household income, and community income are significantly associated with attrition. The dummy variables for supervisor codes are not statistically significant individually, but they are jointly significant as a group. Education and age of the head, household food expenditure, community-mean food expenditures, and the number of schools in the community are not significant in the probit regression.

[15] Following Fitzgerald *et al.* (1998), we do not estimate a dynamic model of survey-by-survey (i.e., between-surveys) attrition. Hence, the estimates in this case can be viewed as coming from a model of cumulative attrition, as opposed to survey-by-survey attrition.

Table 5.2 A Selection Probit Model for analyzing Attrition between KIDS Panel Waves

	Selection Probit (Dep. Var. = 1 if attrited)[1]	
Household Characteristics		
Household size	-0.0795***	(0.0177)
Race (1 if black)	5.8547***	(0.6755)
Gender of head (1 if male)	0.0789	(0.0864)
Education of head	-0.0074	(0.0137)
Age of head	0.0011	(0.0031)
Food expenditure	-0.0004	(0.0006)
Monthly income	0.1162**	(0.0569)
Property ownership (1 if own)	-0.3461***	(0.1183)
Community Characteristics		
Rural (1 if rural)	0.2469*	(0.1287)
Mean food expenditure	0.0001	(0.0019)
Mean income	0.0004**	(0.0002)
Number of schools	-0.0435	(0.0233)
Health facility(1 if available)	-0.1810*	(0.1100)
Month of survey	0.0515	(0.0516)
Crime (1 if occurred)	0.1461	(0.1051)
Supervisor codes jointly significant?	*Yes*	
Constant	-6.2356	
Chi^2 test: overall relation [$p > Chi^2$]	164.08	[0.0000]

[1] Standard errors in parentheses [2] *** indicates significance at 1 percent level; ** and * indicate significance at 5 % and 10 %, respectively.

These variables are also not jointly significant (p=0.4297). Households who owned the dwellings they lived in 1993 were more likely to be re-interviewed, and those who had been victims of crime 12 months prior to the 1993 survey were less likely to be re-interviewed. The Chi^2 statistic for the overall significance of the relation is significant, which implies that attrition between 1993 and 2004 was not purely random. The issue of whether or not attrition affects the consistent estimation of the parameters of interest in this study is dealt with in the next section.

5.2.3 Difference-of-coefficient tests: attritors versus nonattritors

The question addressed in this section is whether or not the coefficients for the outcome variables that are relevant to this study are affected by attrition. In other words, the task is to find out whether or not the coefficients of interest in this study can be estimated consistently using only the sample of households that were re-interviewed. To accomplish this task, we examine whether or not the households who subsequently leave the sample differ in their initial behavioral relationships from those who stay. Following Becketti *et al.* (1988) and Alderman *et al.* (2000), we regress an outcome variable at the initial wave of the survey on predetermined variables and test whether the coefficients of the predetermined variables differ for those respondents who are subsequently lost due to attrition versus those who are re-interviewed. The outcome variables chosen for this analysis are the budget shares of the food groups that enter the demand model. We estimate reduced forms for the budget share equations, and test for whether there are significant statistical differences in the estimated coefficients of nonattitor and attritor households.[16]

To carry out the above difference-of-coefficients test, we divide the 1993 sample into two groups, group 1 comprising households who are interviewed in all the three panel surveys, and group 2 comprising households who subsequently leave the sample. Two indicator variables are then created, corresponding to each of the two groups. The first indicator variable, G_1, takes on the value 1 if the household belongs to the first group and 0 otherwise, and the second indicator variable, G_2, takes on the value 1 if a household belongs to the second group and 0 otherwise. The budget share equations are then

[16] In the next sections, we estimate two demand models separately, one using all households and another using only households in the panel, to further determine the impact of attrition on coefficient estimates.

regressed on a set of household and community variables and their interactions with the G_2 variable. The test for equality of coefficients involves testing whether the coefficient on each non-interacted variable is equal to the coefficient of the corresponding variable interacted with G_2.

To fix ideas, let $\mathbf{z}^h$ represent the set of household and community variables included as explanatory variables in the reduced-form food expenditure function, and let $\mathbf{z}_2^h$ represent a full set of interaction terms involving G_2 and the $\mathbf{z}^h$ variables. The following reduced form regressions are estimated

$$w_i^h = \alpha + \mathbf{z}^{h\prime}\boldsymbol{\beta} + \mathbf{z}_2^{h\prime}\boldsymbol{\gamma} + v$$

The test for equality of coefficients between attritors and nonattritors is simply an F-test of the null hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$. The same household characteristics that enter the selection probit model are used in the reduced form for household food expenditure, except the property ownership variable is left out given that it is not expected to directly affect food consumption decisions. Also, the aggregate food price index is replaced by the individual commodity prices. The results of the F tests are reported in Table 5.3 (to conserve space, the full set results are not reported). We report two F statistics at the bottom of Table 5.3. The first F statistic is for the joint equality of all coefficient pairs including the constant term, and the other for joint equality of all coefficients excluding the constant term. None of these F statistics is statistically significantly different from zero, implying that we cannot reject the null hypothesis of equality of coefficients for attritors and nonattritors. In other words, it can be concluded that the coefficients on the variables in these budget share equations can be estimated consistently using the sample of only nonattriting households.

Table 5.3 Testing Impact of Attrition on the Coefficients of the Budget Share Equations for the Individual Food Groups

	F tests for equality of coefficients						
	Grains	**Meats/Fish**	**Fruits/Veg**	**Dairy**	**Oils/fats**	**Sugar**	**Other**
Joint effect of attrition on all estimates including the constant	0.75 (0.6816)	0.67 (0.7494)	1.31 (0.2172)	0.84 (0.5867)	0.98 (0.4611)	1.02 (0.4238)	1.31 (0.2172)
Joint effect of attrition on all coefficients but not the constant	0.43 (0.9209)	0.68 (0.7249)	1.07 (0.3790)	0.93 (0.5008)	1.08 (0.3764)	1.12 (0.3422)	1.34 (0.2092)

p-values (for prob > F) in parentheses

Hence, based on these F-test tests, it can be concluded that, overall, attrition does not have a significant impact on the slope coefficients of the budget share equations.[17]

5.3 Nonlinearity

5.3.1 LM test results: OLS and SUR estimations

Our test for nonlinearity, developed in chapter three, builds on the work by Banks *et al.* (1997). In particular, the implication of corollary 2 in Banks *et al.* (p.533) is that a utility-derived demand system that is rank 3 and exactly aggregable cannot have both the coefficients on the linear and the quadratic expenditure terms that are independent of prices. In other words, if such a demand model has a coefficient on the linear expenditure term that is independent of prices, then it *must* have a coefficient on the quadratic expenditure term that is price dependent. Hence, our test for nonlinearity involves testing

[17] Another way to check for the impact of attrition on coefficient estimates is to estimate the model of interest without attrition correction and with attrition correction (using a procedure such as inverse probability weighting), and then examine if there are large differences in the coefficient estimates. However, this is computationally burdensome when using nonlinear models that require difficult user-defined programs.

for the statistical significance of prices in the coefficient on the quadratic expenditure term. The top part of the table reports results of the LM tests for nonlinearity in each of the budget share equations. The first column of Table 5.4 reports the results of these tests from pooled OLS estimation.

Table 5.4 Tests for Nonlinearity of the Demand System based on Statistical Significance of Prices of the Coefficient on the Quadratic Expenditure Term

Nonlinearity tests in the individual budget share equations

	Significance of Prices: LM Statistics (Heteroskedasticity-robust)	
Commodity	**OLS (p-value)**	**IV-2SLS (p-value)**
Grains	22.03 (0.0025)	24.64 (0.0009)
Meat and fish	5.23 (0.6314)	19.63 (0.0064)
Fruits and vegetables	23.99 (0.0011)	21.59 (0.0030)
Dairy	23.77 (0.0012)	7.63 (0.3660)
Oils, butter, and other fats	12.43 (0.0872)	6.98 (0.4310)
Sugar	19.63 (0.0064)	11.34 (0.1244)
Other	75.46 (0.0000)	41.87 (0.0000)

System-wide test for nonlinearity (all budget share equations)

	SUR	**3SLS**
Chi-square (p-value)	158.63 (0.0000)	87.82 (0.0000)

Based on these results, the null hypothesis that the coefficient on the quadratic expenditure term is independent of prices is rejected (at the 10% significance level) in all individual budget share equations, except that for meat and fish. The implication of these results is that the individual budget share equation for meat and fish only require a linear expenditure term, so that the inclusion of the quadratic expenditure term is unnecessary in this case.

To allow for cross equation correlations, we also estimate the equations jointly in a systems framework and test the null hypothesis that the coefficients on the expenditure terms across all equations do not depend on prices. Results of the Chi^2 test computed from the SUR estimation of budget share equations are reported at the bottom part (first column) of Table 5.5. This test provides strong evidence against the null of price-independence of the coefficient on the quadratic expenditure term.

5.3.2 LM test results: IV-2SLS and 3SLS estimations

As discussed in chapter three, total household food expenditure is likely to be correlated with the error term, and this may affect the results of the LM tests. In this subsection, we explicitly test for the endogeneity of expenditure to determine whether or not the LM test needs to be adjusted. Total household income is used as an identifying instrumental variable (IV) for total household food expenditure.[18] For income to be a good IV for expenditure, it must meet two conditions: the *relevance* condition, which requires that income be sufficiently correlated with expenditure (the endogenous variable), and the *exogeneity* condition, which requires that income must not be correlated with the error term in the demand model. The former condition is testable, and the latter cannot be tested (see Wooldridge, pages 118-122). A test for the relevance condition involves determining whether income is partially correlated with expenditure.

[18] For brevity, the term *expenditure* will be used to refer to total household expenditure, and *income* will be used to refer to total household income.

This test involves determining whether log total household income ($\ln m$) is statistically significant in the reduced form regression for log expenditure ($\ln x$).

The middle column of Table A1 in the appendix to this chapter reports parameter estimates of the reduced form regression for $\ln x$. Based on simple t-tests, the coefficient on $(\ln m)^2$ is significantly different from zero, while the coefficient on $\ln m$ is not, mainly due to the collinearity between $\ln m$ and $(\ln m)^2$. A formal test for the relevance condition involves testing for the joint significance of the coefficients on both instrumental variables in the reduced form. The results of the F tests for the joint significance of $\ln m$ and $(\ln m)^2$ are presented in the bottom row of Table 5.4. These tests provide evidence of a strong partial correlation between $\ln m$ and $(\ln m)^2$ and $\ln x$. Thus, based on the results of this test, it can be concluded that income and income squared are relevant instruments for expenditure, and hence, the former will be used as instrumental variables for the latter in the analyses that follow (the exogeneity assumption is, of course, maintained).

Table 5.4 also reports the results of the tests for expenditure exogeneity. The procedure for carrying out endogeneity tests is as discussed in chapter three; it involves augmenting the budget share equation for each food group with residuals from the reduced forms for expenditure ($\hat{v}$), then testing for their statistical significance. Table A2 at the end of this chapter presents estimates of the reduced forms for each budget share equation. The results of the tests for the significance of $\hat{v}$ (and hence, for the test of the null hypothesis that expenditure is exogeneous) are reported in the top portion of Table 5.5. As can be seen from these test results, there is limited statistical evidence against the null hypothesis of expenditure exogeneity. The hypothesis is rejected (at the 10% significance level) only in the budget shares equations for grains, meat and fish, and dairy. The strong statistical evidence supporting expenditure exogeneity in the grains budget share is somewhat counterintuitive. One would have expected that, because grains comprise food items that are major staples in South African diets, its relation with total food expenditure would be somewhat 'fixed.'

The problem with testing for endogeneity in the individual budget share equations is that it ignores correlations among the equations. To allow for these correlations, we estimate the budget share equations as a system using the seemingly unrelated regressions (SUR) and then test the null hypothesis that $\hat{v}$ is statistically significant across all the equations. Results of the χ^2 tests for the statistical significance of $\hat{v}$ are reported in middle part of Table 5.5 (to conserve space, not all the coefficient estimates of the SUR estimation are reported). The first χ^2 statistic is computed from the unrestricted SUR estimation, while the second one is calculated from the theory-restricted SUR estimation; the demand theory restrictions imposed are homogeneity and symmetry. Both the restricted and unrestricted SUR-based tests reject the null hypothesis that the coefficients on $\hat{v}$ are jointly zero across all equations, providing evidence that expenditure is endogenous in the system.

Since we are using two identifying instruments ($\ln m$ and $(\ln m)^2$) for $\ln x$, we have one overidentifying restriction. Our test for overidentifying restriction follows Wooldridge (2002), pp. 122-124. This test (also known as the Sargan test) involves estimating each budget share equation by IV-2SLS (using $\ln m$ and $(\ln m)^2$ as instruments) and obtaining residuals, regressing these residuals on all exogenous variables and obtaining the R-squared statistic (call this R_u^2), and then computing $\chi^2(1)$ as the product of the number of observations times R_u^2. The results of these tests are also presented in the top portion (second column) of Table 5.5.

Table 5.5 Results of the Test for the Endogeneity of Expenditure

Tests for endogeneity of expenditure in the individual budget share equations		
Commodity	**Endogeneity tests t stat (p-value)**	**Overidentification tests χ^2 (p-value)**
Grains	3.59 (0.000)	0.01 (0.958)
Meat and fish	-4.11 (0.000)	8.40 (0.004)
Fruits and vegetables	1.37 (0.171)	0.27 (0.600)
Dairy	-1.85 (0.064))	4.58 (0.032)
Oils, butter, and other fats	-0.26 (0.796)	2.17 (0.141)
Sugar and sugar products	1.13 (0.258)	10.03 (0.001)
Other	0.22 (0.825)	11.12 (0.000)
Tests for endogeneity of expenditure across all budget share equations in the system		
SUR (Unrestricted)	26.73 (0.0002)	
SUR (Restricted)[1]	17.64 (0.0072)	
Test for the relevance of income and income-squared as IVs for expenditure		
F stat. (p-value)	116.90 (0.0000)	

[1] The demand theory restrictions imposed during estimation are symmetry and homogeneity (additivity is satisfied automatically by the data)

As with the findings from expenditure exogeneity tests, statistical evidence on the validity of the instruments based on the overidentification tests is mixed across budget share equations. Statistical evidence in support of the exogeneity of ln m and (ln m)^2 is strong in the budget share equations for grains, fruits and vegetables, and oils and fats. There is no clear relationship between findings from exogeneity tests and overidentification restrictions (i.e., the budget share equations that pass the overidendification tests are not the same as those that pass (or do not pass) expenditure exogeneity tests). The finding that not all the budget share equations pass the overidentification tests indicates that one of the instruments, ln m or (ln m)^2, may not be completely exogenous in these equations.

In summary, the exogeneity tests indicate that expenditure is endogenous in the budget share equations for grains, meat and fish, and dairy. However, the null hypothesis that expenditure is exogenous across all budget share equations is rejected. Given these findings, it is necessary to determine whether or not the LM tests results stay the same when expenditure endogeneity is adjusted, particularly for the grains, meat and fish, and dairy commodities.

The results in the second column of Table 5.4 are computed from budget share equations estimated using instrumental variables two-stage least squares (IV-2SLS), with income used as an IV for expenditure. As can be seen from these results, after adjusting for expenditure endogeneity, the null hypothesis that the budget share equation for meat and fish is linear expenditure in expenditure is rejected. We consider these results (based on IV-2SLS) to be more reliable in the case of meat and fish, given the finding from the endogeneity tests which indicated that expenditure is endogenous in the budget share equation for meat and fish. Contrary to the LM test results based on the equation-by-equation OLS estimations, the null hypotheses that expenditure is linear in the budget share equations for dairy and oils and fats are not rejected. The null hypothesis of expenditure exogeneity was not rejected in the budget share equation for oils and fats, which implies that if the individual budget share equations are considered, the OLS results may be more reliable.

To account for cross equation correlations, we also estimated the equations as a system, and tested the null hypothesis that the coefficient on the quadratic expenditure term does not depend on prices. The results of this test are reported in the bottom part (second column) of Table 5.4. The χ^2 statistic for this test is computed from three-stage least squares (3SLS) estimation with the expenditure instrumented by income. Similar to the results of the SUR test, this test provides strong evidence against the null of price-independence of the coefficient on the quadratic expenditure term.

In summary, the null hypothesis that expenditure enters the demand model linearly is rejected in six of the seven budget share equations based on the endogeneity-unadjusted OLS estimations, and in four budget share equations when expenditure endogeneity is corrected for using IV-2SLS. When cross-equation correlations are allowed for using SUR and 3SLS, we find strong evidence in favor of the QUAIDS model specification.

Given these results of the nonlinearity tests, and the finding in the previous subsection that expenditure is endogenous in some budget share equations and that it is endogenous in the equation system, we will proceed by estimating the demand models that include the quadratic expenditure term (that is, the QUAIDS model) with expenditure endogeneity endogeneity adjusted for. Results of the endogeneity-unadjusted models will also be presented, so as to determine the extent to which ignoring endogeneity biases the parameter estimates.

5.4 Demand Model Results

This section reports estimation results of the demand models. All of the demand models are estimated using pooled maximum likelihood (ML), with convergence occurring at 0.000001, the default tolerance level for MLE in Stata. Theoretical restrictions of adding-up, homogeneity, and symmetry are maintained during estimation. The budget equation for the 'other' food group was deleted during estimation to avoid singularity of the variance-covariance matrix when all seven equations are included. Parameters of the deleted equation are recovered through the adding-up restrictions.

Due to the large number of parameters estimated, not all estimation results will be presented. Presentation will concentrate mainly on the elasticity estimates, and how these change when different model specifications are considered.

This section starts by conducting model specification tests. The LM tests for QUAIDS versus AIDS model specifications based on the statistical significance of prices in the coefficient on the quadratic expenditure term (Table 5.5) supported the QUAIDS specification. In this section, we conduct the likelihood ratio (LR) tests for AIDS versus QUAIDS specifications. The LR testing approach differs slightly from the LM-based tests of the previous section in that the LR tests are based on the observation that the only difference between the AIDS and QUAIDS models is that AIDS contains only the linear expenditure term, while QUAIDS contains both the linear and quadratic expenditure terms. In other words, QUAIDS nests AIDS, so that once QUAIDS has been estimated, a test for whether or not AIDS is the appropriate model specification involves simply checking for the statistical significance of the quadratic expenditure term. The coefficient on the quadratic expenditure term is $\lambda_i / b(\mathbf{p}) = \lambda_i / \prod_{i=1}^{K} p_{it}^{\beta_i}$, so that testing for its significance involves simply testing for the statistical significance of lambda (λ_i). The main difference between this test and the LM tests is that this test requires the estimation of the QUAIDS (unrestricted) model, as opposed to the LM test for which only the estimation of the restricted model is needed. Also, the approaches followed in deriving these tests are different in that the tests based on LM statistic tests for the statistical significance of prices in $b(\mathbf{p})$, while the LR test checks for the significance of λ in the coefficient on the quadratic expenditure term.

The tests for significance of λ_i are conducted in the endogeneity-uncorrected QUAIDS model. The reason for using endogeneity-uncorrected model, as opposed to the model with endogeneity corrected for, is to avoid problems of inferential invalidity caused by generated regressors (Wooldridge, 2002; pp. 115-118). This problem arises here because the use of the control function procedure involves including the residuals from the reduced form regressions as regressors in the demand model. But, because these residuals are generated using the same data used for demand estimation, their inclusion as regressors raises questions in terms of the asymptotic validity of the estimated standard errors and test statistics on other regressors (parameters can still be estimated consistently).

The first column of Table 5.6 reports results of the tests of the null hypothesis that lambda (λ_i) is zero in the budget share equation of each of the food groups. This hypothesis is rejected in the budget share equations of four of the seven food groups, thus providing some evidence in support of the QUAIDS model specification. So, in addition to the budget share equation for meat and fish, which was also found to require only the linear expenditure term by the LM tests (Table 5.5), the LR tests indicate further that the budget share equations for fruits and vegetables and oils and fats only require the linear expenditure term. The findings in the case of oils and fats are the same as those from IV-2SLS. As explained above, the approaches followed in constructing the two tests differ, so that their leading to different conclusions regarding which budget share equation is linear in expenditure and which is not is not necessarily unexpected. However, what we can conclude is that both provide evidence against the AIDS model as a system specification. A test of the null hypothesis that λ is not different from zero in all the budget share equations is strongly rejected (Chi^2 = 72.56, p=0.0000).

Table 5.6 Tests Endogeneity of Expenditure, and for the Statistical Significance of Lambda in the demographically-extended QUAIDS Model

Tests based on the budget share equations for each commodity				
		Expenditure Endogeneity Tests		
Commodity	**Significance of Lambda χ^2 tests (p-val.)**	**QUAIDS t-tests (p-val.)**	**Nonlin. AIDS t-tests (p-val.)**	**LA/AIDS t-tests (p-val.)**
Grains	12.42 (0.0004)	-0.05 (0.964)	-0.11 (0.911)	0.87 (0.385)
Meat and Fish	0.01 (0.9501)	-1.53 (0.126)	-1.75 (0.081)	-2.60 (0.009)
Fruits and Vegetables	0.73 (0.3929)	1.17 (0.241)	1.06 (0.288)	1.81 (0.071)
Dairy	12.37 (0.0004)	0.31 (0.753)	-0.02 (0.986)	-0.67 (0.502)
Oils, butter, other fats	1.97 (0.1600)	-0.99 (0.321)	-1.12 (0.261)	-0.68 (0.498)
Sugar	11.99 (0.0005)	-0.24 (0.807)	-0.83 (0.405)	-0.09 (0.929)
Other	50.94 (0.0000)	1.07 (0.284)	1.93 (0.053)	1.12 (0.263)
System-wide (all equations) test				
χ^2 (p-value)	72.56 (0.0000)	5.03 (0.5398)	8.22 (0.2226)	10.22 (0.1158)

The next step in the analysis in this section is to test for the exogeneity of total household food expenditure in the QUAIDS model. As discussed in chapter three, expenditure endogeneity is adjusted for using the control function approach, which involves augmenting each of the budget share equations in the demand system with the residuals from the reduced form regression for expenditure. Once the reduced form residuals have been augmented to the budget share equations, testing for exogeneity becomes straightforward, as it involves simply testing for the statistical significance of the coefficient on the residuals. We follow Blundell and Robin (1999) and include only the residuals from the reduced form for $\ln x$, using income and income squared as instruments for $\ln x$.

Results of the Chi^2 tests for expenditure exogeneity in the QUAIDS-estimated budget share equations of the individual food groups are reported in the second column of Table 5.6. As is clear from these results, statistical evidence against expenditure endogeneity in the individual budget share equations is very weak, much weaker than was found with OLS-based tests. The results of the test of the null hypothesis that expenditure is exogenous across all budget share equations in the demand model are presented in the last row of Table 5.6. This test gives a Chi^2 statistic of 5.03 ($p=0.5398$), implying that in the case of system estimation of the budget share equations, it may not be necessary to control for the expenditure endogeneity. Thus, based on these results, if one were to test the null hypothesis of expenditure exogeneity in the individual budget share equations using tests based on OLS estimations, there would be a higher likelihood of rejecting the null hypothesis than if one used tests based on the nonlinear QUAIDS estimation. The third and fourth columns of Table 5.6 also report results of the tests for expenditure endogeneity based on the AIDS and LA/AIDS models. These will be discussed in detail in the latter sections. For now, it suffices to mention that exogeneity tests in the nonlinear AIDS model lead to the same conclusion as tests based on the QUAIDS model—that expenditure is not endogeneous across all the budget share equations. Tests based on the LA/AIDS model provide some evidence, although weaker ($\chi^2 = 10.22$, $p = 0.1158$), of expenditure endogeneity.

To summarize the results in this section, we found that both the LM and LR specification tests support QUAIDS, as opposed to the AIDS, model. Tests for expenditure exogeneity based on the QUAIDS estimation led to failure to reject the null hypothesis of expenditure exogeneity. Hence, based on these test results, our preferred estimates are the results of the QUAIDS model without adjustment for expenditure endogeneity. However, given the findings from the OLS-based tests that expenditure is endogenous, we also report the results of the endogeneity-adjusted QUAIDS model.

Maximum likelihood parameter estimates of the QUAIDS model are presented in Table A3 at the end of this chapter. The third column reports parameter estimates of the endogeneity-unadjusted QUAIDS model, while the fifth column reports those for estimated from QUAIDS controlling for expenditure endogeneity. Only 9 of the 28 price effects are significantly different from zero at the 10% significance level, suggesting that there is not much quantity response to movements in relative prices, possibly due to the level of aggregation in the commodity groups. Most (34 out of 49) of the coefficient estimates on the demographic variables are statistically different from zero. Households with large sizes consume more grains and dairy products while their small-sized counterparts consume more meat and fish and fruits and vegetables. These results are as expected, given that grains provide a relatively cheap source of calories compared to such foods as meat and fish, and given that large households are likely to have more children who consume milk and other dairy products. Formal tests of the effects of household demographic characteristics and time effects on expenditure patterns are presented in Table 5.7. As is clear from these results, the null hypothesis of no demographic effects is strongly rejected. The null hypothesis that preferences have remained stable over time (i.e., that there was no structural change) is also rejected, as shown by highly significant year dummies. The month of the survey is also significant across the budget share equations, indicating the importance of seasonality in food purchase and consumption patterns.

Table 5.7 Results of the Wald Tests for Demographic effects, Structural Change, and Seasonality

	χ^2	**Degrees of freedom**	***p*-value**
Demographic effects	1244.97	24	0.0000
Structural change (aggregate time effects)	125.08	12	0.0000
Month/seasonality effects	45.24	6	0.0000

The interpretation of price and income effects is best discussed in terms of elasticities. Estimates of expenditure elasticities based on both the endogeneity-adjusted and endogeneity-unadjusted QUAIDS models are reported in Table 5.8. We focus in this section on the elasticity estimates from the endogeneity-unadjusted models. All expenditure elasticity estimates are statistically significant at less than 1% level. The estimates are all positive, as would be expected for broadly defined commodities like the ones considered here. Hence, all of the seven food groups can be classified as normal, which implies that their demand increases as total household expenditure increases. Meat and fish are luxuries, with expenditure elasticities in excess of unity.

Table 5.8 Expenditure Elasticities estimated from QUAIDS Models with and without Endogeneity Adjustments

Commodity	**Expenditure Elasticity**	
		Endog. Adjusted
Grains	0.3881	0.3881
	(0.1254)	(0.1254)
Meat, fish	1.1363	1.1363
	(0.1326)	(0.1326)
Fruits, vegetables	0.9739	0.9739
	(0.1095)	(0.1095)
Dairy	0.6509	0.6509
	(0.1518)	(0.1518)
Oils, butter, fats	0.5572	0.5572
	(0.0989)	(0.0989)
Sugar	0.2924	0.2924
	(0.0886)	(0.0886)
Other	2.9761	2.9761
	(0.1818)	(0.1818)

Standard errors in parentheses

Table 5.9 presents estimates of the own price elasticities. Estimates of all own and cross price elasticities are reported in Table A4 at the end of this chapter. As can be seen from Table 5.9, all estimates of the own price elasticities are highly significant (at 1% level). These estimates are negative, as expected. Apart from dairy, all food groups are either (approximately) unitary elastic or price elastic based on the estimated Marshallian price elasticities. This indicates the degree of responsiveness that households have to changes in the prices of these foods. However, when only the substitution effects are considered, grains and meat and fish become less price elastic, as shown by the inelastic compensated own-price elasticities.

Table 5.9 Own-price elasticities estimated from QUAIDS models with and without endogeneity adjustments

Marshallian/uncompensated own-price elasticities

		Endog. Adjusted
Grains	-1.0296	-1.0458
	(0.0742)	(0.0856)
Meat, fish	-1.1077	-1.1058
	(0.1586)	(0.1943)
Fruits, vegetables	-0.9803	-0.9739
	(0.0957)	(0.0967)
Dairy	-0.8143	-0.8238
	(0.0788)	(0.0977)
Oils, butter, fats	-1.0604	-1.0679
	(0.0745)	(0.0755)
Sugar	-1.0463	-1.0477
	(0.0749)	(0.0842)
Other	-4.6426	-4.9186
	(0.6605)	(0.7423)

Hicksian/compensated own-price elasticities

Grains	-0.9032	-0.9330
	(0.0408)	(0.0456)
Meat, fish	-0.8455	-0.8312
	(0.0306)	(0.0342)
Fruits, vegetables	-0.8087	-0.8082
	(0.0193)	(0.0217)
Dairy	-0.7715	-0.7842
	(0.0100)	(0.0118)
Oils, butter, fats	-1.0307	-1.0334)
	(0.0053)	(0.0064)
Sugar	-1.0320	-1.0333
	(0.0043)	(0.0053)
Other	-4.3462	-4.6264
	(0.0181)	(0.0210)

Standard errors in parentheses

5.4.1. The effects of controlling for expenditure endogeneity

Given the findings from the OLS-based tests for expenditure endogeneity, we also report results of the QUAIDS model with adjustment for expenditure endogeneity. The fifth column of Table A3 at end of this chapter reports the parameter estimates of the endogeneity-corrected QUAIDS model. The estimates of most interest in comparing the two QUAIDS specifications (one with, and the other without endogeneity adjustment) are the expenditure elasticities. Expenditure elasticities estimated from endogeneity-adjusted QUAIDS model are reported in the second column of Table 5.8. These estimates remain virtually unchanged across all equations after controlling for expenditure endogeneity. Their signs are the same, and their magnitudes and standard errors are very similar. The own-price elasticity estimates computed from the endogeneity-adjusted QUAIDS model are reported in the second column of Tables 5.9. As can be seen, controlling for expenditure endogeneity does not change the estimates of the price elasticities significantly.

Given the finding that expenditure is not endogenous in the QUAIDS model, our preferred estimates are those from QUAIDS without endogeneity adjustment.

5.4.2. The effects of home production

To determine the possible effects of home production on expenditure patterns, the elasticities were also estimated excluding the households who reported engaging in home production. The elasticity estimates are reported in Table A5 (Part I) at the end of the chapter. We compare these elasticity estimates with those estimated from QUAIDS with demographic variables. Apart for the meat and fish food group, the elasticity estimates are similar to those obtained when all households are used. The exclusion of households who report home production changes the classification of meat and fish from luxuries to necessities. However, these changes in magnitudes are also the result of the elimination of a large number of observations from the main dataset. About 31% of the observations have nonzero own production. The elimination of this large number of households is expected to change results, so that these changes cannot all be attributed solely to own production.

5.4.3. The effects of attrition

In addition to the attrition tests of section 5.2, we investigate the possible effect of attrition on demand parameters by estimating the elasticities using data on all households, including those who attrited. The elasticities are estimated using the QUAIDS model with demographics. The results are reported in Table A5 (Part II) at the end of the chapter. Compared these estimates with the panel data-based estimates, it can be seen that, apart from dairy, the expenditure elasticities estimated using the two datasets are very similar. The own-price elasticities, particularly the Hicksian elasticities, are also very similar. These findings are consistent with those from the attrition tests in section 5.2.

5.4.4. The effects of excluding the quadratic expenditure term

This subsection examines the impact of excluding the quadratic expenditure term in the demand model. As explained in chapter three, the exclusion of the quadratic expenditure term from the QUAIDS model results in the AIDS model, so that the AIDS is nested within QUAIDS. If no assumptions are made about the translog price aggregator, ln a(**p**) (equation (3.14), chapter 3), then the AIDS model is nonlinear in parameters, which makes it hard to estimate empirically. However, once the QUAIDS model has been estimated, estimation of the AIDS model is straightforward, because this requires simply restricting the coefficient on the quadratic expenditure term to zero. In this subsection, we estimate the nonlinear AIDS model, and compare elasticities computed from it with those computed from the QUAIDS model. Table A6 at the end of this chapter presents the maximum likelihood parameter estimates of the AIDS model. Estimates of the AIDS expenditure and own-price elasticities are reported in Table 5.10 and Table 5.11, respectively.

Table 5.10 Expenditure Elasticities estimated from LA/AIDS and AIDS models with and without endogeneity adjustments

	LA/AIDS		**Nonlinear AIDS**	
		Endog. Adjusted		**Endog. Adjusted**
Grains	0.8237	0.7937	0.8288	0.8325
	(0.0138)	(0.0372)	(0.0137)	(0.0346)
Meat, fish	1.1482	1.2671	1.1434	1.2153
	(0.0185)	(0.0494)	(0.0183)	(0.0452)
Fruits, vegetables	0.8783	0.7993	0.8827	0.8408
	(0.0176)	(0.0471)	(0.0174)	(0.0432)
Dairy	1.1799	1.2438	1.1704	1.1720
	(0.0378)	(0.1024)	(0.0374)	(0.0940)
Oils, butter, fats	0.6846	0.7333	0.6899	0.7627
	(0.0270)	(0.0771)	(0.0268)	(0.0704)
Sugar	0.5800	0.5860	0.5874	0.6378
	(0.0249)	(0.0727)	(0.0246)	(0.0657)
Other	1.7038	1.5950	1.6900	1.5208
	(0.0383)	(0.1049)	(0.0379)	(0.0955)

Standard errors in parentheses

Estimated expenditure elasticities are all significant at the 1% level. Given the finding in the previous sections that the QUAIDS elasticity estimates are also significant at the 1% level, the choice between QUAIDS and AIDS expenditure elasticity estimates cannot be based on their statistical significance. On average, the AIDS expenditure elasticity estimates are larger than the QUAIDS estimates, which is consistent with the QUAIDS model not allowing for adequate curvature in the expenditure response of households. On average, the expenditure elasticity estimates differ by about 53% between the two models, with most of the AIDS estimates being larger. The largest differences occur in the elasticity estimates for grains, dairy, and sugar food groups, with expenditure elasticity estimates from the AIDS model being about 113%, 80%, and 101%, respectively, larger than those based on the QUAIDS model. Exclusion of the quadratic expenditure term has a larger effect on the magnitudes of expenditure elasticities than price elasticities. This is to be expected, given that the main difference between the AIDS and QUAIDS models lies in the specification of the expenditure term.

These are also the three food groups whose budget share equations were found to require the quadratic expenditure term based on both the LM and LR tests. This partly explains why exclusion of the quadratic expenditure term has the largest effect on elasticity estimates of these food groups. Estimates of the own price elasticities are reported in Table 5.11. The differences in the magnitudes of the estimated own price elasticities are substantial also in the grains and sugar.

The expenditure and own-price elasticities do not change significantly when correction is made for endogeneity. This is to be expected because the null hypothesis of expenditure exogeneity was rejected in the AIDS model (Table 5.6).

Table 5.11 Own-price elasticities estimated from LA/AIDS and AIDS models with and without endogeneity adjustments

Marshallian/uncompensated own-price elasticities

	LA/AIDS		Nonlinear AIDS	
		Endog. Adjusted		**Endog. Adjusted**
Grains	-0.8828	-0.8451	-0.9279	-0.9274
	(0.0384)	(0.0593)	(0.0356)	(0.0395)
Meat, fish	-1.0822	-1.0364	-1.0996	-1.1129
	(0.0461)	(0.0532)	(0.0472)	(0.0476)
Fruits, vegetables	-0.8694	-0.8169	-0.9015	-0.8935
	(0.0348)	(0.0486)	(0.0339)	(0.0346)
Dairy	-0.8545	-0.8441	-0.8797	-0.8775
	(0.0784)	(0.0810)	(0.0774)	(0.0778)
Oils, butter, fats	-1.0293	-1.0522	-1.0668	-1.0845
	(0.0747)	(0.0828)	(0.0733)	(0.0761)
Sugar	-0.8986	-0.8973	-0.9455	-0.9312
	(0.0613)	(0.0614)	(0.0622)	(0.0641)
Other	-0.8830	-0.9873	-1.2065	-1.2240
	(0.1094)	(0.1387)	(0.1029)	(0.1037)

Hicksian/compensated own-price elasticities

	LA/AIDS		Nonlinear AIDS	
		Endog. Adjusted		**Endog. Adjusted**
Grains	-0.6146	-0.5866	-0.6580	-0.6563
	(0.0368)	(0.0500)	(0.0350)	(0.0357)
Meat, fish	-0.8172	-0.7440	-0.8357	-0.8325
	(0.0466)	(0.0600)	(0.0471)	(0.0470)
Fruits, vegetables	-0.7146	-0.6761	-0.7459	-0.7454
	(0.0340)	(0.0429)	(0.0338)	(0.0338)
Dairy	-0.7769	-0.7623	-0.8027	-0.8003
	(0.0786)	(0.0828)	(0.0773)	(0.0774)
Oils, butter, fats	-0.9928	-1.0131	-1.0300	-1.0439
	(0.0745)	(0.0809)	(0.0731)	(0.0751)
Sugar	-0.8704	-0.8688	-0.9169	-0.9001
	(0.0613)	(0.0615)	(0.0623)	(0.0650)
Other	-0.7133	-0.8284	-1.0382	-1.0726
	(0.1108)	(0.1459)	(0.1031)	(0.1051)

Standard errors in parentheses

5.4.5. The effect of imposing linearity

Nonlinearity in the AIDS model is caused by the translog price index, ln $a(\mathbf{p})$ (equation (3.14), chapter 3). It is common in applied demand analysis to linearize the AIDS model by replacing this translog price index with a share-weighted price index. Moschini (1995) develops an index which enhances the approximation abilities of the linear AIDS model compared to the commonly used indices (such as Stone's, Laspeyres, and Paasche indices). The replacement of the translog index with Moschini's or any of these linear indices leads to a model that is linear in parameters, which greatly simplifies empirical estimation. In fact, the only existing theory-based food demand study in South Africa that considers an exhaustive list of food commodity groups similar to the one considered in this study uses the linearized AIDS model (LA/AIDS model). This subsection examines the effect on the price and expenditure elasticities, of imposing linearity on the AIDS model. To accomplish this, an LA/AIDS model is estimated with the translog index replaced by Moschini's price index, and then the LA/AIDS elasticity estimates are compared with those calculated from the nonlinear AIDS and QUAIDS models. The estimates of the LA/AIDS expenditure and own-price elasticities are reported in the first two columns of Table 5.11.

The expenditure elasticities estimates and standard errors of the nonlinear AIDS and LA/AIDS are very similar; on average, the elasticity estimates differ by less than 1% and the standard errors by less than 2%. However, the differences in the magnitudes of the elasticity estimates from the two models are larger (about 5%) when expenditure endogeneity is corrected for. A possible explanation is the finding of some evidence of expenditure endogeneity in the LA/AIDS model. The differences in the magnitudes of own price elasticity estimates from the two models are also small. On average, the own price elasticity estimates from LA/AIDS tend to be larger than those from the nonlinear AIDS model.

There is an important difference between the elasticity estimates from the endogeneity-corrected LA/AIDS and AIDS models: the estimates of the Marshallian and Hicksian own-price elasticities calculated from the LA/AIDS model are larger in absolute value than those computed from the nonlinear AIDS model for all food commodities. Hence, at least for the current sample, the linearization of the AIDS model leads to a downward bias in the estimates of own-price elasticities.

Similar to the findings in previous section when nonlinear AIDS was compared with QUAIDS, the differences in the estimated expenditure elasticities between the LA/AIDS and QUAIDS models are largest for grains (112%) and sugar (108%). The reason for these large differences is that, in addition to excluding the quadratic expenditure term, the LA/AIDS model further imposes linearity on the price index, ln a(**p**). As will be shown later, welfare measures that require estimates of expenditure elasticities will differ substantially, depending on which model (QUAIDS or AIDS) is used.

5.5 Rural-Urban and Income-Groups Differences

This section examines whether or not there are any significant disparities in food expenditure patterns between rural and urban households, as well as among households in different income groups. We estimate demand models separately for rural and urban households, as well as for households in each of the three income groups. To create the income groups, the CPI-deflated income of each household is averaged over the three panel years, and then households are ranked from lowest to highest based on their averaged incomes. Households are then divided into three income groups, with households in each income group comprising about one-third of the total sample.

Although the inclusion of the quadratic expenditure term may be appropriate for the pooled data (i.e., pooled across all households), it may not be necessary when specific household groups are considered. For this reason, we start the analyses in this section by testing for whether or not the quadratic expenditure term is significant in the QUAIDS model estimated for each of the sub-samples (i.e., in each of the rural, urban, and income groups samples). As before, these tests are conducted in the endogeneity-unadjusted models. Results of these tests, based on the statistical significance of lambda, are reported in Table 5.12.

Similar to the findings with pooled data, evidence in support of the quadratic expenditure specification in the budget share equations for the individual commodities is mixed. However, when all the budget share equations are considered jointly, the evidence in favor of QUAIDS is robust across all the five sub-samples. An interesting observation about these test results is that there is a clear rural-urban difference, with the statistical evidence in support of QUAIDS stronger for the urban than for the rural sample. The statistical evidence in favor of QUAIDS tends to also be weaker in the individual income groups data than in the pooled data. A possible explanation is that because the role of the quadratic expenditure term is to capture the curvature of the expenditure responses of households, grouping households with similar incomes together homogenizes the sample, thereby playing the role that is supposed to be played by the quadratic expenditure term. Meenkashi and Ray (1999) found a similar result in India, where statistical evidence in support of QUAIDS was weak when data on the individual Indian states were considered, but strong in the data pooled across states.

Table 5.12 Tests for Quadratic Expenditure Specification based on the Statistical Significance of Lambda in the QUAIDS model

Statistical Significance of Lambda: χ^2 tests[1]

Commodity	Rural	Urban	Low	Middle	High
Grains	26.28	4.23	1.10	9.08	0.25
	(0.0000)	(0.0397)	(0.2951)	(0.0026)	(0.6165)
Meat, fish	11.30	32.75	1.34	1.73	6.78
	(0.0008)	(0.0000)	(0.2473)	(0.1886)	(0.0092)
Fruits, vegetables	0.97	1.11	0.01	0.01	2.17
	(0.3245)	(0.2923)	(0.9861)	(0.9775)	(0.1404)
Dairy	2.70	8.88	2.10	0.29	15.06
	(0.1004)	(0.0029)	(0.1470)	(0.5924)	(0.0001)
Oils, butter, fats	0.19	3.89	2.02	2.76	0.15
	(0.6661)	(0.0486)	(0.1556)	(0.0964)	(0.7018)
Sugar	12.52	3.11	6.53	8.84	4.08
	(0.0004)	(0.0777)	(0.0106)	(0.0029)	(0.0433)
Other	17.37	25.35	5.53	15.70	7.78
	(0.0000)	(0.0000)	(0.0187)	(0.0002)	0.0053
All equations	*49.10*	*64.93*	*14.20*	*28.04*	*32.82*
	(0.0000)	*(0.0000)*	*(0.0275)*	*(0.0001)*	*(0.0000)*

[1] p-values in parentheses

Another interesting observation about these test results is that the statistical evidence in support of the inclusion of the quadratic expenditure term strengthens with income. The main result derived from the tests in Table 5.12 is that it is appropriate to include the quadratic expenditure term when estimating the demand model for each of the five sub-samples, and hence, that all of the analyses in this section are based on QUAIDS.

Since it is possible that the results of the tests for expenditure endogeneity may change depending on the specific sub-sample being considered, the next task in our analysis is to F-tests for the endogeneity of expenditure in each of the five sub-samples.

Table 5.13 Tests for Expenditure Endogeneity in the Rural-Urban and Income Groups Samples

Relevance of Instruments: F-tests

	Rural	Urban	Low	Middle	High
Expenditure	59.94	77.68	14.40	14.81	16.71
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Expenditure Endogeneity: χ^2 tests based on QUAIDS estimation[1]

Commodity	Rural	Urban	Low	Middle	High
Grains	0.06	30.86	1.36	0.27	1.07
	(0.8121)	(0.0000)	(0.2443)	(0.6012)	(0.3012)
Meat, fish	3.57	0.45	0.28	1.87	0.58
	(0.0590)	(0.5010)	(0.5990)	(0.1711)	(0.4473)
Fruits, vegetables	1.59	4.10	5.12	0.31	0.38
	(0. 2075)	(0. 0428)	(0.0237)	(0.5749)	(0.5355)
Dairy	0.01	14.65	1.14	0.01	0.20
	(0.9417)	(0.0001)	(0.2857)	(0.9988)	(0.6557)
Oils, butter, fats	3.19	3.19	0.61	0.38	0.01
	(0.0740)	(0.0740)	(0.4364)	(0.5366)	(0.9873)
Sugar	2.09	7.58	2.52	0.66	1.62
	(0.1479)	(0.0059)	(0.1121)	(0.4158)	(0.2028)
Other	3.09	0.60	0.39	7.51	0.39
	(0.0786)	(0.4372)	(0.5302)	(0.0061)	(0.5328)
All equations	*9.89*	*51.06*	*10.73*	*8.95*	*3.28*
	(0.1295)	*(0.0000)*	*(0.0972)*	*(0.1767)*	*(0.7559)*

Overidentification: χ^2 tests

Grains	0.10	3.43	0.19	2.57	1.11
	(0.7423)	(0.0640)	(0.6611)	(0.1089)	(0.2919)
Meat, fish	0.05	14.72	1.41	1.30	1.71
	(0.05)	(0.0001)	(0.2343)	(0.2540)	(0.1912)
Fruits, vegetables	0.95	11.23	0.37	32.46	0.43
	(0.3290)	(0.0008)	(0.5439)	(0.0000)	(0.5106)
Dairy	2.75	1.83	2.46	3.33	1.18
	(0.0971)	(0.1759)	(0.1166)	(0.0682)	(0.2771)
Oils, butter, fats	0.12	2.28	0.16	3.99	0.04
	(0.7299)	(0.1311)	(0.6895)	(0.0457)	(0.8482)
Sugar	5.58	9.57	4.65	6.65	0.04
	(0.0182)	(0.0020)	(0.0310)	(0.0099)	(0.8410)
Other	4.39	4.08	2.40	1.12	7.37
	(0.0384)	(0.0433)	(0.1215)	(0.2904)	(0.0066)

p-values in parentheses

The results of the tests for the relevance of income and income squared as instrumental variables for expenditure are reported in the top part of Table 5.13. Consistent with the previous findings based on the pooled sample, total household income explains a significant portion of the variation in total household food expenditure, so that assuming the endogeneity condition is satisfied, income can be used as a valid instrument for expenditure.

Since the QUAIDS model has already been estimated, testing for expenditure endogeneity simply requires augmenting the residuals from the reduced form regressions for expenditure to each of the budget share equations in the demand model, and testing for their statistical significance. The middle part of Table 5.13 reports results of these χ^2 tests, first in the individual budget share equations, and then across all budget share equations in the QUAIDS equation system.

Overall, statistical evidence against the null hypothesis of expenditure exogeneity is very weak in the sub-samples. It is only in the urban sub-sample that statistical evidence strongly supports the alternative of expenditure endogeneity. The null of expenditure exogeneity is also rejected on the low income sub-sample, albeit weakly so ($p = 0.0972$).

Similar to the findings with pooled data, results of the overidentification tests, (reported in the bottom portion of Table 5.13) are mixed in the individual budget share equations. There appears to be a problem with exogeneity of the instruments in the urban sub-sample, which, unfortunately, happens to be the sub-sample that requires correction for expenditure endogeneity. There does not appear to be problems with the exogeneity of instruments in the budget share equations estimated using the low income sub-sample.

In accordance with the preceding model specification and endogeneity test results, the analyses that follow will be based on the QUAIDS model, with correction for expenditure endogeneity in urban and low income sub-samples. The maximum likelihood parameter estimates of the QUAIDS model for rural and urban samples are reported in Table A8 at the end of this chapter. Given that the finding of expenditure endogeneity in the urban sample, we report parameter estimates for both endogeneity-corrected and uncorrected models for the urban samples in Table A8. From the parameter estimates, it can be seen that in both rural and urban areas, larger-sized households consume more grains and less meat and fish and fruits and vegetables.

The estimates of expenditure elasticities are presented in Table 5.14. For some food groups, the difference in the estimated expenditure elasticities between rural and urban samples is quite substantial. For urban households, a 1% increase in total food expenditure leads to only 0.02% increase in the quantity consumed of meat and fish. It is very different with the rural households, where the same 1% expenditure increase leads to about 1.73% increase in consumption of meat and fish. This is one of the reasons why it is necessary to examine expenditure patterns of rural households separately from urban households. The finding that expenditure on grains is less responsive to increases in total household expenditure in the rural areas than urban areas is counterintuitive. The reason for the unresponsiveness of grain purchases to expenditure increases among rural households could be due to their engagement in home production (90% of households who reported nonzero home production reside in the rural areas). Endogeneity correction does not significantly impact the expenditure elasticity estimates of urban households.

Estimates of all price and expenditure elasticities are reported in Table A9 at the end of this chapter.

Table 5.14 Estimated Expenditure Elasticities: Rural-Urban and Income Group Differences

	Rural	**Urban**	**Low**	**Middle**	**High**
		Endog.Adjust	**Endog.Adjust**		
Grains	0.0662	0.9338	0.2122	0.1294	0.8626
	(0.1524)	(0.1937)	(0.6568)	(0.2370)	(0.2566)
Meat, fish	1.7359	0.1164	1.8501	1.5624	0.3554
	(0.1650)	(0.1863)	(0.6344)	(0.2632)	(0.2754)
Fruits, vegetables	1.0300	1.0548	0.7247	0.8229	1.2257
	(0.1285)	(0.1743)	(0.5033)	(0.2139)	(0.2135)
Dairy	0.9512	0.7034	0.3841	1.0100	0.0249
	(0.1991)	(0.2277)	(0.5209)	(0.3445)	(0.2626)
Oils, butter, fats	0.8042	0.3323	0.4615	0.4310	0.7425
	(0.1194)	(0.1822)	(0.3627)	(0.1429)	(0.2365)
Sugar	0.2423	0.1524	0.0332	0.1945	0.7116
	(0.1051)	(0.1664)	(0.3796)	(0.1425)	(0.1448)
Other	2.6091	3.1482	3.4620	3.2054	2.6957
	(0.2251)	(0.2942)	(0.8175)	(0.3584)	(0.3541)

Standard errors in parentheses

We focus now on the income-group differences. The maximum likelihood estimates of the QUAIDS model for each income group are presented in Table A10 at the end of this chapter. As in the rural-urban samples, households with larger sizes consume more grains and less meat and fish and fruits and vegetables in all the income groups. Estimated expenditure elasticities for each of the income groups are also presented in Table 5.14.

Grains are a necessity across income groups, while meat and fish are luxuries in the low and middle income groups. High income households spend more of their total expenditure increases on fruits and vegetables than middle and low income households.

Again, these differences in demand behavior reaffirm the need for disaggregated analysis of expenditure patterns by income groups in a country with high income disparities like South Africa. Nevertheless, the estimated expenditure elasticities are as one would expect a priori. Given the budget share of oils and fats is about the same among all income groups (approximately 6%), the finding that low income households respond to total expenditure increases by consuming more oils and fats is of policy interest, given the likely negative effects if the consumption levels of these products exceed levels recommended for healthy diets. Correction for expenditure endogeneity in the low income sub-sample has the largest impact on grains and dairy. Interestingly, expenditure endogeneity was not found to be a problem in these commodities.

Table 5.15 reports estimates of the Marshallian and Hicksian price elasticities. Own-price elasticities are all negative as expected. Rural households are less responsive to increases in prices of grains and meat and fish, than rural households. This behavioral pattern is also shared by high income households in comparison to low and middle income households. The response to the price increase of meat and fish food group makes the differences in the demand behavioral patterns among the rural-urban and income group more vivid. High income households respond to a 1% increase in the price of meat and fish by decreasing consumption by 0.59%. This is very different than low income households, for whom the same 1% price increase leads to a decrease of about 2.48% in the consumption of meat and fish. This further underscores the need to undertake disaggregated analysis of demand behavior in a South Africa.

Table 5.15 Estimated Marshallian and Hicksian Own Price Elasticities

Marshallian Own Price Elasticities

Commodity	Rural	Urban	Low	Middle	High
		<u>Endog.Adjust</u>	<u>Endog.Adjust</u>		
Grains	-1.0083	-0.7474	-0.8889	-1.1924	-0.8770
	(0.1077)	(0.0998)	(0.1868)	(0.1987)	(0.0668)
Meat, fish	-1.9868	-0.5248	-2.4796	-1.6434	-0.5898
	(0.3020)	(0.1889)	(1.4364)	(0.4451)	(0.1550)
Fruits, vegetables	-1.0614	-0.9598	-0.9468	-0.8690	-1.2072
	(0.1287)	(0.1568)	(0.2343)	(0.1374)	(0.2764)
Dairy	-0.9247	-0.7245	-0.9815	-0.7310	-1.1427
	(0.0993)	(0.1383)	(0.3596)	(0.1742)	(0.3081)
Oils, butter, fats	-1.1463	-0.8409	-0.9035	-1.2419	-0.9912
	(0.0918)	(0.1440)	(0.1712)	(0.0966)	(0.1720)
Sugar	-1.0731	-1.1547	-1.1357	-1.0992	-0.9203
	(0.0905)	(0.1502)	(0.4463)	(0.1241)	(0.1075)
Other	-2.8749	-5.1792	-7.2895	-5.1092	-3.9698
	(0.5953)	(1.1354)	(3.6996)	(1.3480)	(1.2688)

Hicksian Own Price Elasticities

Grains	-0.9840	-0.5191	-0.8111	-1.1471	-0.6515
	(0.0559)	(0.0474)	(0.2406)	(0.0828)	(0.0671)
Meat, fish	-1.6203	-0.4934	-2.0963	-1.2967	-0.4964
	(0.0348)	(0.0502)	(0.1314)	(0.0584)	(0.0724)
Fruits, vegetables	-0.8751	-0.7835	-0.8119	-0.7226	-1.0054
	(0.0232)	(0.0291)	(0.0937)	(0.0381)	(0.0351)
Dairy	-0.8719	-0.6640	-0.9617	-0.6669	-1.1407
	(0.0110)	(0.0196)	(0.0268)	(0.0219)	(0.0215)
Oils, butter, fats	-1.1040	-0.8227	-0.8781	-1.2191	-0.9524
	(0.0063)	(0.0100)	(0.0200)	(0.0076)	(0.0123)
Sugar	-1.0597	-1.1492	-1.1337	-1.0890	-0.8952
	(0.0058)	(0.0061)	(0.0221)	(0.0075)	(0.0051)
Other	-2.6709	-4.7331	-7.0282	-4.8478	-3.5879
	(0.0176)	(0.0417)	(0.0617)	(0.0292)	(0.0502)

Standard errors in parentheses

5.6 The Problem of Observed Zero Expenditures

For reasons such as purchase infrequency and corner solution outcomes from agents' optimization problems, it is common to observe zero budget shares for some commodities in household expenditure data. Estimation of a demand system with a large percentage of zero expenditure shares may lead to inconsistent parameter estimates unless econometric techniques appropriate for such data structures are used.

As indicated in chapter four, the problem of zero expenditure shares is severe for the dairy commodity (see Table 4.6). Apart from the dairy commodity, the percentage of observations with zero expenditure shares is very low for the other six commodities (4% at most). Hence, it would not be appropriate to model the budget shares for these commodities using econometric techniques meant for data structures with large number of zeros in the dependent variable. In the pooled sample, about 14% of the dairy budget shares are zeros. Non-purchase of dairy products is higher among rural households (18%) and among households in the lower income brackets (24% for low-income households and 12% for middle-income households). Hence, adjustment for zero expenditure shares is made in the estimation of the dairy budget share equations for these household groups. In this study, the adjustment for zero expenditure shares follows the two-step procedure proposed by Shonkwiler and Yen (1999).

The Shonkwiler and Yen procedure is carried out in two steps. In the first step, a single equation probit model is estimated to compute the probability and the cumulative density values. The dependent variable in the probit models is a binary variable taking a value of one if positive purchase occurs and zero otherwise. The independent variables are the exogenous variables income and income-squared, household demographics that enter the demand model (household size, urbanization, and education of household head), year dummies and survey month variable. Four additional variables not included in the demand model are included in the probit models, namely household size-squared, gender and age of household head, and a dummy variable for household ownership of a refrigerator. In the second step, equation (3.30) is re-specified with the following changes made: (1) the budget shares of the dairy commodity are multiplied by the cumulative density values ($\Phi(.)$), and (2) the probability density values ($\phi(.)$) are included as an additional regressor in the budget share equation for the dairy commodity.[19] The discussion here focuses on the statistical significance of δ—the coefficient on $\hat{\phi}$ — and on the impact on the elasticity estimates for the dairy commodity of controlling for the zero expenditure problems. As before, the tests for the statistical significance of δ are conducted in the endogeneity-unadjusted demand model. The results of these tests are reported in the last row of Table 5.16. As can be seen from these results, the statistical significance of δ is robust across the sample groups, indicating that the additional information provided by the probability density values explains a significant part of the variation in the budget share of the dairy commodity.

Table 5.16 also reports estimates of the expenditure and own-price elasticities. For the rural and middle income sub-samples, these elasticities are computed from the censored QUAIDS model with no adjustment made for expenditure endogeneity, while for the low income and pooled samples, they are estimated from the QUAIDS model with adjustment for both expenditure endogeneity and censoring. Comparing these elasticity estimates with those computed from QUAIDS without correcting for censoring (Table 5.10), it can

[19] For ease of reference, we reproduce eq. (3.30) here: $w_{it}^h = \Phi(\mathbf{z}_{it}'\boldsymbol{\tau}_i)w_{it}^h(\mathbf{p}_t^h, m_t^h; \boldsymbol{\psi}) + \delta_i\phi(\mathbf{z}_{it}^h{}'\boldsymbol{\tau}_i) + \xi_{it}^h$

be seen that the correction for the zero-expenditure problem has the effect of decreasing the magnitudes of the expenditure elasticity estimates for the dairy commodity.

Table 5.16 Expenditure and Own Price Elasticities for the Dairy Commodity with adjustment for the Zero-Expenditure Problem

	Household Group			
	Rural	Low	Middle	All
Expenditure elasticity	0.8743	0.3397	0.6162	0.3881
	(0.2028)	(0.5450)	(0.5579)	(0.1865)
Marshallian own-price elasticity	-1.3591	-0.8043	-0.2502	-0.5857
	(0.2007)	(0.2484)	(0.5532)	(0.1056)
Hicksian own-price elasticity	-1. 3563	-0.7869	-0.2110	-0.5602
	(0.0551)	(0.0281)	(0.0354)	(0.0123)
t-ratio (δ=0 vs. $\delta \neq$0)	-3.87	-1.96	-2.74	-4.18
(p-value)	(0.000)	(0.050)	(0.006)	(0.000)

Standard errors in parentheses for the elasticity estimates

The magnitudes of decreases in the expenditure elasticities are large enough to change the classification of dairy from luxury to necessity in the case of middle income households. Similar to the findings with regard to expenditure elasticities, on average, adjustment for expenditure censoring has the effect of decreasing the absolute magnitudes of the Marshallian and Hicksian own-price elasticities for dairy.

5.7 Implications of demand model selection for welfare measures

One of the main reasons for estimating demand systems is to facilitate welfare analysis. Quantitative welfare measures are based on demand parameter estimates, and, as the analyses in the previous sections showed, these parameter estimates differ depending on the choice of the demand model. This raises the question: to what extent does this sensitivity of demand parameter estimates to model selection affect welfare measures and hence, policy advice? To answer this question, we use the estimated demand models to estimate welfare effects of an indirect tax reform. In particular, we compare these welfare measures using demand parameters from QUAIDS, nonlinear AIDS, and LA/AIDS, all estimated on the pooled data from all households.

The tax reform evaluated is the zero-rating of the value-added tax (VAT) on meat products. While most of the basic food commodities in South Africa such as grains, milk, fruits and vegetables have zero VAT, meat is not zero-rated. That is, meat is taxed at the standard VAT rate of 14%. The issue of whether or not meat should be zero-rated has been contended for a long time by government and lobby groups, the Congress of South African Trade Unions being most notable among the lobby groups (Watkinson and Makgetla, 2002). In this section, we measure the consumer welfare impacts of adding meat to the list of zero-rated commodities. We use the demand parameter estimates to calculate the indirect utilities, both before and after the hypothesized tax reform. These are then used to compute two money metric welfare measures, namely compensating variation and equivalent variation.

Let $\mathbf{p}_0^h$ denote the price vector before the tax reform, and let u_0^h denote household h's utility given those prices. Let $\mathbf{p}_1^h$ and u_1^h denote the price vector and utility level after the tax reform has been implemented. Denote the expenditure function for price vector $\mathbf{p}^h$ and utility level u^h by $e(\mathbf{p}^h, u^h)$. Compensating variation (CV^h) is defined as the amount of money that a household would have to receive in order the remain at the level of utility u_0^h given the new price vector $\mathbf{p}_1^h$: $\text{CV}^h = e(\mathbf{p}_1^h, u_0^h) - e(\mathbf{p}_0^h, u_0^h)$. Equivalent variation ($\text{EV}^h$), on the other hand, is based on the post-reform utility level u_1^h and asks how much money a household would be indifferent about accepting in lieu of the price change; that is, it is the change in the household's wealth that would be equivalent to the price change in terms of its welfare impact: $\text{EV}^h = e(\mathbf{p}_1^h, u_1^h) - e(\mathbf{p}_0^h, u_1^h)$.

Because of the finding in the previous sections that expenditure endogeneity is not a problem with the pooled data, we choose as our preferred model censored QUAIDS with no adjustment for expenditure endogeneity. We calculate the new price vector, $\mathbf{p}_1^h$, and the CV and EV using the final year of the data (that is, 2004).

Figures 5.1 and 5.2 plot the CV and EV estimates based on the QUAIDS specification for each household. These are negative for every point in the data, indicating that each household experiences a welfare gain as a result of the price decrease. The gains in welfare increase with total expenditure, as would be expected.

To determine the sensitivity of welfare measures to demand model selection, Table 5.17 presents summary statistics of the CV and EV computed using the three model specifications considered in this study, namely QUAIDS, nonlinear AIDS, and LA/AIDS. Based on the QUAIDS model, the removal of a 14% VAT on meat would yield an average compensating variation of R203.32, meaning that the average household would gain R203.32 per month if this tax reform were in effect.[20]

[20] The symbol R stands for Rand, the South African currency.

Figure 5.1 Welfare Gain due to Removal of 14% VAT on Meat: Compensated Variation

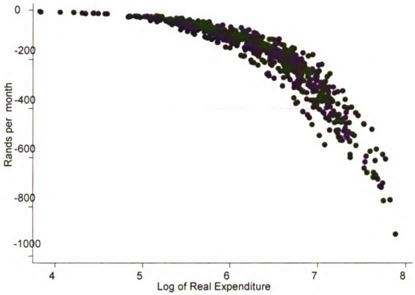
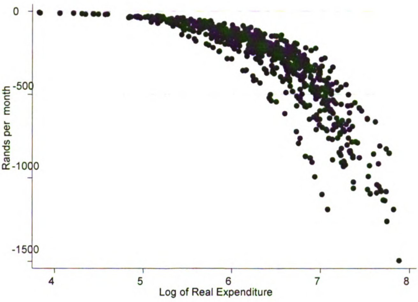


Figure 5.2 Welfare Gain due to Removal of 14% VAT on Meat: Equivalent Variation



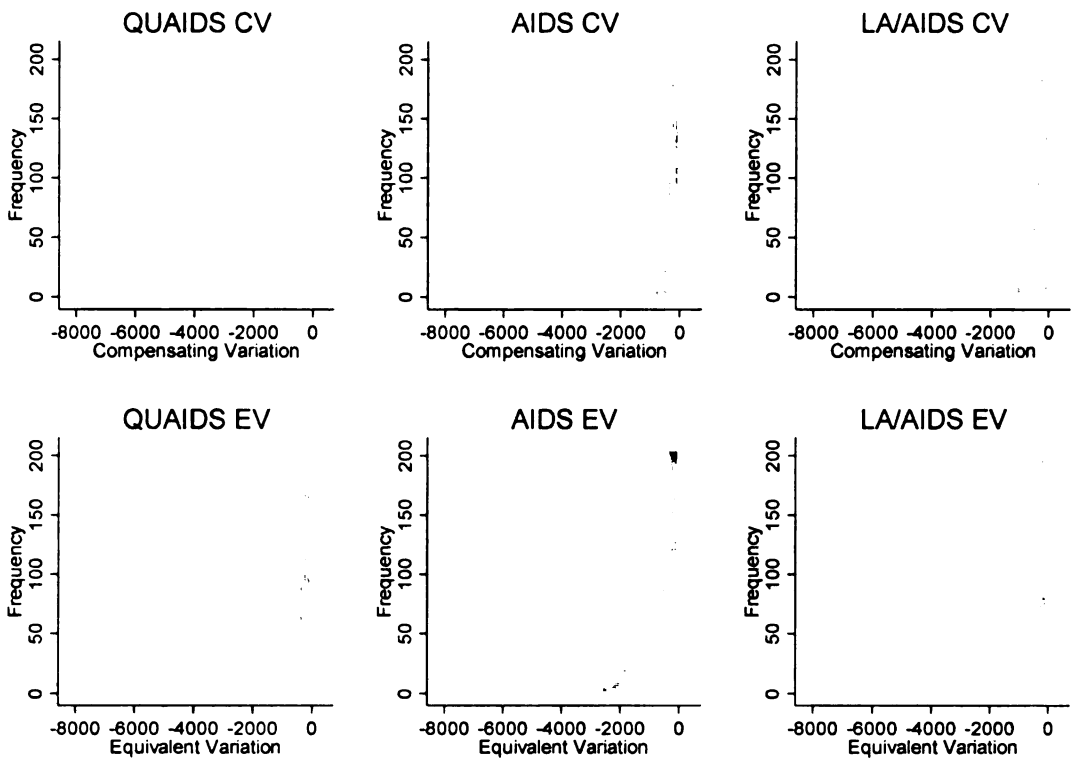
The equivalent variation is R110 higher, implying that the demand for meat exhibits significant income effects. The mean differences in the CV and EV estimates based on the nonlinear AIDS and LA/AIDS models are small. However, compared to the QUAIDS estimates, it is clear that the AIDS and the LA/AIDS models overstate the welfare measures, presumably due to their exclusion of the quadratic expenditure term.

Table 5.17 Average Welfare Effects of Zero-Rating VAT on Meat

	QUAIDS	Nonlinear AIDS	LA/AIDS
Compensating variation	-203.32	-299.64	-306.62
Equivalent variation	-313.93	-644.84	-676.46

Figure 5.3 shows the distribution of the welfare estimates across the sample using the different models. The welfare effects from the three models are all skewed to the left, with the AIDS and LA/AIDS welfare effects estimates being more variable than the QUAIDS estimates. Figures 5.4 and 5.5 indicate the biases introduced when the AIDS and LA/AIDS are used to estimate welfare measures. Based on the equivalent variation, Figure 5.3 plots the difference between the AIDS welfare gain for each household as a proportion of that household's QUAIDS welfare gain. Figure 5.4 compares these biases when the LA/AIDS model is used. It is clear from these figures that for the zero-rating of meat, both AIDS and LA/AIDS models overstate the welfare gains of households, and that this overstatment is larger for richer households. Again, this is consistent with both AIDS and LA/AIDS models not allowing for adequate curvature in the Engel curves.

Figure 5.3 The Distribution of Welfare Gains



To gain a further understanding of the distribution of welfare gains, Table 5.18 summarizes the estimates of compensating and equivalent variations for rural and urban households, as well as for households in different income groups. Based on the compensating variation, the welfare gains of a lower meat price accrue most to urban and high-income households. As was indicated in chapter four, high-income households have a higher share (0.26) of meat products in their budgets than low-income (0.21) and middle-income (0.22) households.

Figure 5.4 Bias in Welfare Gain from Using the Nonlinear AIDS Model

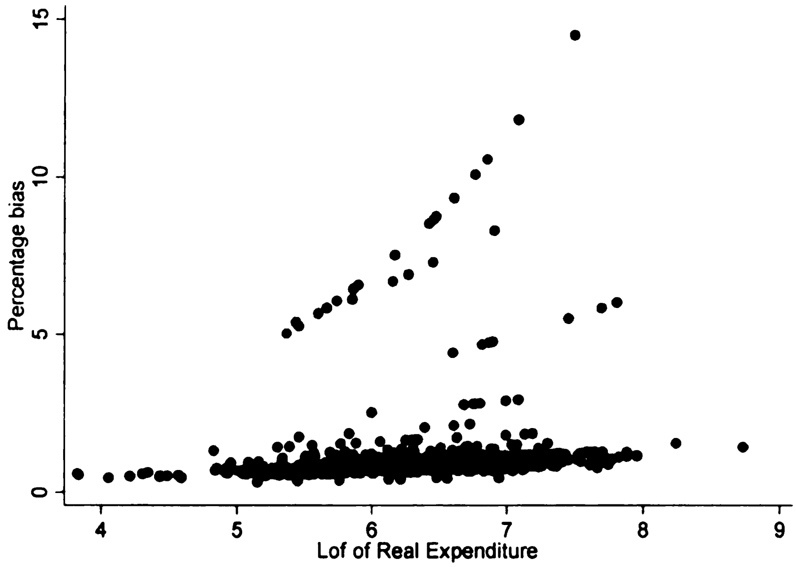


Figure 5.5 Bias in Welfare Gain from Using the LA/AIDS Model

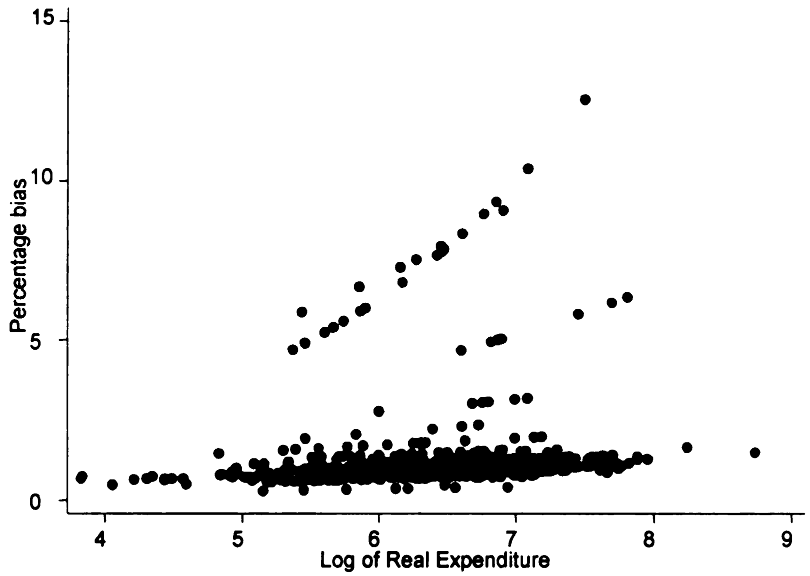


Table 5.18 Average Welfare Gains of Zero-Rating VAT on Meat: Rural-Urban and Income Group Differences

	Regions		Income Groups		
	Rural	**Urban**	**Low**	**Middle**	**High**
Compensating variation	-158.99	-275.00	-138.23	-232.15	-328.85
Equivalent variation	-233.48	-411.24	-408.55	-717.75	-587.97

Based on equivalent variation, middle-income households benefit most from the tax reform. As was shown in figures 5.3 - 5.4, the variability in gains is larger at higher levels of expenditure, which makes the standard deviations of the welfare estimates to be larger for middle and high income groups. On average, all households in rural and urban areas, as well as the three income groups, benefit from the zero-rating of VAT on meat.

5.8 Comparison with previous studies

The majority of food demand studies in South Africa are commodity-specific. Virtually all of these studies are based on aggregate time series of price and expenditure data, do not account for the effect of household demographics, and are not based on theoretically consistent demand models. The exception is the study by Agbola (2003), which estimates price and expenditure elasticities using the LA/AIDS model based on a 1993 cross-sectional survey. In this section, we compare elasticity estimates from this study and those estimated by previous food demand studies in South Africa. Table 5.18 summarizes estimates of price and expenditure elasticities from these previous studies. For ease of reference, we also summarize in Table 5.20 elasticity estimates from all models based on pooled data (for reasons of space, standard errors are left out).

Elliot and van Zyl's (1991) estimates of expenditure elasticities for maize and potatoes, based on log-linear estimations of the budget share equations, are substantially lower than the expenditure elasticity for grains and fruits/vegetables estimated in this study. Part of the reason for these differences is, of course, due to different levels of aggregation. However, they fitted log-linear models on the individual budget share equations (as opposed to estimating these as a system), which implies that they did not account for the interrelatedness of the commodities expenditure shares through, for example, the adding-up constraint. This can lead to biases in the elasticity estimates. The estimates by Niewuwoudt (1998) suffer from the same problem as those by Elliot and van Zyl, in that they are based on individual budget share estimates. Taljaard (2003) used the LA/AIDS to estimate price and expenditure elasticities for different meat types using aggregate, national time series data. Taljaard's estimates of the expenditure elasticities

Table 5.19 Estimates of Elasticities in South Africa: Previous Studies

	STUDY	MODEL	EXPENDITURE ELASTICITY		OWN-PRICE ELASTICITY	
			Rural	**Urban**	**Marsh.**	**Hick.**
GRAINS	Agbola (2003)	LA/AIDS	1.250		-1.730	-1.394
Maize meal	Elliot & van Zyl (1991)	log-log	0.1961	0.0610		
Rice			0.1565	0.1567		
Bread			0.2310	0.1111		
MEAT/FISH	Agbola (2003)	LA/AIDS	1.027		-1.266	-0.919
	Nieuwoudt (1998)		1.34 (blacks)	0.95 (blacks)		
Beef	Hancock (1984)	OLS	0.71		-0.96	
	Taljaard (2003)	LA/AIDS	0.607		-0.544	-0.256
	Cutts & Kirsten (2005)		1.265		-0.594	
	Bowmaker & Nieuwoudt (1990)	OLS	0.96		-1.72	
Chicken	Taljaard	LA/AIDS	0.053		-0.131	-0.116
Pork			0.015		-0.261	-0.260
Mutton			3.642		-0.923	-0.335
FRUITS /VEGETABLES						
Fruits	Agbola (2003)	LA/AIDS	0.717		-0.263	-0.225
Vegetables			0.910		-1.312	-1.195
Potatoes	Elliot and van Zyl (1991)	log-log	0.1597	0.1357		
DAIRY	Agbola (2003)	LA/AIDS	0.898		-1.237	-1.160
Fresh Milk	Nieuwoudt (1998)		0.60 (blacks)	0.50 (blacks) 0.74 (Asians)		
Cheese			0.65 (blacks)	0.65 (blacks) 0.65 (Asians)		

Table 5.20 Comparison of elasticities from different demand model specifications

	Linear Expenditure & Estimation: LA/AIDS		Linear Expenditure, Nonlinear Estimation: AIDS		Quadratic Expenditure, Nonlinear Estimation: QUAIDS			
		Endogeneity controlled		Endogeneity controlled	Without Demographics	With Demographics	Endogeneity controlled	Censored Demand
Expenditure elasticities								
Grains	0.8237	0.7937	0.8288	0.8325	0.4511	0.3881	0.3465	0.2539
Meat, fish	1.1482	1.2671	1.1434	1.2153	0.9572	1.1363	1.1900	1.2209
Fruits, vegetables	0.8783	0.7993	0.8827	0.8408	1.0547	0.9739	0.9405	0.8708
Dairy	1.1799	1.2438	1.1704	1.1720	0.8581	0.6509	0.6020	0.3882
Oils, butter, fats	0.6846	0.7333	0.6899	0.7627	0.6888	0.5572	0.6165	0.5667
Sugar	0.5800	0.5860	0.5874	0.6378	0.3927	0.2924	0.2951	0.3064
Other	1.7038	1.5950	1.6900	1.5208	2.7724	2.9761	2.9649	3.2660
Marshallian/uncompensated own-price elasticities								
Grains	-0.8828	-0.8451	-0.9279	-0.9274	-1.0903	-1.3488	-1.0296	-1.0458
Meat, fish	-1.0822	-1.0364	-1.0996	-1.1129	-0.9561	-1.0711	-1.1077	-1.1058
Fruits, vegetables	-0.8694	-0.8169	-0.9015	-0.8935	-1.0205	-0.8742	-0.9803	-0.9739
Dairy	-0.8545	-0.8441	-0.8797	-0.8775	-0.965	-1.1427	-0.8143	-0.8238
Oils, butter, fats	-1.0293	-1.0522	-1.0668	-1.0845	-1.0223	-1.0695	-1.0604	-1.0679
Sugar	-0.8986	-0.8973	-0.9455	-0.9312	-0.9609	-0.9622	-1.0463	-1.0477
Other	-0.883	-0.9873	-1.2065	-1.224	-3.5341	-0.5572	-4.6426	-4.9186
Hicksian/compensated own-price elasticities								
Grains	-0.6146	-0.5866	-0.658	-0.6563	-0.9434	-0.8393	-0.9032	-0.933
Meat, fish	-0.8172	-0.744	-0.8357	-0.8325	-0.7352	-0.8482	-0.8455	-0.8312
Fruits, vegetables	-0.7146	-0.6761	-0.7459	-0.7454	-0.8346	-0.7068	-0.8087	-0.8082
Dairy	-0.7769	-0.7623	-0.8027	-0.8003	-0.9086	-1.093	-0.7715	-0.7842
Oils, butter, fats	-0.9928	-1.0131	-1.03	-1.0439	-0.9855	-1.0011	-1.0307	-1.0334)
Sugar	-0.8704	-0.8688	-0.9169	-0.9001	-0.9417	-0.8852	-1.032	-1.0333
Other	-0.7133	-0.8284	-1.0382	-1.0726	-3.258	-0.6163	-4.3462	-4.6264

for the individual meat types differ substantially from our estimate for the meat and fish composite, but their average of 1.07 falls below the range of estimates obtained in this study. The main limitations with Taljaard's study are that it is based on highly aggregate data, and does not account for influence of demographic effects on demand behavior.

The results from Agbola (2003) are most comparable to the results in this study for two reasons. First, Agbola's is the only study that covers an exhaustive list of food groups similar to the ones covered here, and second, the KwaZulu-Natal portion of the 1993 nationwide used by Agbola became the first wave of the KIDS panel data used in this study. Agbola's estimates of expenditure elasticities are higher for grains (1.250) compared to this study's estimates. Grains comprise mainly of items that are staple foods in South Africa, so that one can consider an expenditure elasticity in excess of unity (implying that grains are a luxury) to be too high. The Marshallian and Hicksian price elasticity estimates reported by Agbola are also very high for a staple commodity.

The expenditure and price elasticities estimated in this study correct for many of the problems associated with these previous estimates. The QUAIDS model used to estimate expenditure and price elasticities in this study is derived from consumer utility maximization problem (that is, it is theoretically consistent) and allows for more curvature in the Engel curves. The system is estimated in its flexible form, without imposing linearity on the price aggregators, and corrections for econometric problems of expenditure endogeneity and censoring are made. So, this study overcomes many of the problems associated with the previous studies of demand behavior in South Africa, and hence, the price and expenditure elasticity estimates obtained from this study are more accurate and can be used more reliably for policy analysis.

5.9 Chapter summary

This chapter presented and discussed empirical results. Tests for attrition effects showed that attrition does not affect consistent estimation of the coefficients of the demand model. In most of the models estimated, expenditure endogeneity was not found to be problem. It was only in the urban and low income sub-samples that the null hypothesis of expenditure exogeneity was rejected. Correction for expenditure endogeneity was done using the control function approach. The LM tests and the tests for the significance of lambda in the QUAIDS model indicated that a quadratic expenditure term needs to be included in the demand model (that is, these tests favored QUAIDS, as opposed to AIDS, specification). To determine the extent of the bias in parameter estimates from excluding the quadratic expenditure term, the AIDS and LA/AIDS models were also estimated. On average, the expenditure elasticity estimates based on the AIDS and LA/AIDS models were found to be larger than the QUAIDS estimates, consistent with the former models not allowing for adequate curvature in the Engel curves. Since welfare measures depend on the demand parameter estimates, and demand parameter estimates depend on the choice of the demand model, we examined how sensitive the welfare measures are to demand model selection. For this purpose, we compared the estimates of compensating variations and equivalent variations from the QUAIDS, AIDS, and LA/AIDS models resulting from the removal of a 14% VAT on meat. AIDS and LA/AIDS were found to overstate these welfare gains compared to QUAIDS.

Demand behavior was found to significantly differ between rural and urban households, as well as across income groups. Hence, to get a more accurate picture of demand behavior, it is necessary to conduct a disaggregated analysis, as opposed to aggregate analysis which masks these heterogeneous demand behavioral patterns.

Previous estimates of price and expenditure elasticities in South Africa are commodity-specific, and most of them are also based on single equation estimations. The existing theory-based study by Agbola (2003) that examines a range of food groups similar to the ones examined in this study uses the restrictive LA/AIDS model. Our estimates of expenditure and price elasticities correct for many of the problems associated with Agbola's and all other estimates from these previous studies. The QUAIDS model used to estimate expenditure and price elasticities in this study is derived from consumer utility maximization problem (that is, it is theoretically consistent) and allows for curvature in the Engel curves. The system is estimated in its flexible form, without imposing linearity on the price aggregators, and corrections for econometric problems of expenditure endogeneity and censoring are made. Hence, this study provides estimates of price and expenditure elasticities that can be used reliably for policy analysis.

APPENDIX A

ADDITIONAL RESULTS FOR CHAPTER FIVE

Table A1 The estimated reduced forms for expenditure and expenditure-squared

	Income included	**Income and Income-Squared included**	**Income and Income-Squared included**
Variable	**ln x**	**ln x**	**$(\ln x)^2$**
Constant	6.3686 (0.2545)	7.0905 (0.3453)	52.0476 (4.4837)
Price of grains	0.1467 (0.0616)	0.1467 (0.0615)	2.0281 (0.7985)
Price of meat & fish	0.0668 (0.0557)	0.0707 (0.0555)	1.0588 (0.7213)
Price of fruits & vegetables	0.0441 (0.0419)	0.0461 (0.0418)	0.5116 (0.5434)
Price of dairy	-0.0248 (0.0441)	-0.0237 (0.0440)	-0.1381 (0.5716)
Price of oils, butter & fats	0.1099 (0.0572)	0.1078 (0.0570)	1.3129 (0.7410)
Price of sugar	-0.3062 (0.0508)	-0.3050 (0.0506)	-4.0938 (0.6577)
Price of other foods	-0.2015 (0.0590)	-0.2035 (0.0588)	-2.5710 (0.7639)
Household size	0.0467 (0.0031)	0.0465 (0.0031)	0.6197 (0.0403)
Race	-0.4110 (0.0425)	-0.3944 (0.0427)	-5.3974 (0.5554)
Rural	0.0110 (0.0308)	0.0129 (0.0308)	0.2539 (0.3994)
Education	0.0131 (0.0025)	0.0118 (0.0026)	0.1489 (0.0333)
Dummy for 1998	-0.4456 (0.0905)	-0.4460 (0.0904)	-5.8427 (1.1733)
Dummy for 2004	-1.0470 (0.1413)	-1.0427 (0.1411)	-13.6487 (1.8316)
Survey month	-0.0051 (0.0094)	-0.0050 (0.0094)	-0.0781 (0.1220)
Total household income	0.1743 (0.0117)	-0.0409 (0.0706)	-1.0062 (0.9167)
Total household income^2		0.0152 (0.0049)	0.2302 (0.0640)
R^2	*0.4339*	*0.4364*	*0.4423*
F (p-val)	*223.38 (0.000)*	*116.90 (0.0000)*	*116.84 (0.0000)*

[1] Standard errors in parentheses

Table A2 The estimated reduced forms for individual commodity groups

Variable	Grains[1]	Meat/ Fish	Fruits/ Veg.	Dairy	Oils/fats	Sugar	Other
Constant	0.5243	-0.5973	0.6058	-0.4770	0.0907	0.1178	0.7358
	(0.1964)	(0.1873)	(0.1369)	(0.1093)	(0.0631)	(0.0528)	(0.1659)
Price of grains	-0.0037	0.0272	0.0157	-0.0121	0.0023	-0.0103	-0.0192
	(0.0140)	(0.0134)	(0.0098)	(0.0078)	(0.0045)	(0.0038)	(0.0118)
Price of meat/fish	0.0170	-0.0102	-0.0108	-0.0106	0.0043	0.0086	0.0017
	(0.0125)	(0.0119)	(0.0087)	(0.0069)	(0.0040)	(0.0034)	(0.0105)
Price of fruits/Veg.	-0.0115	-0.0032	0.0172	-0.0021	0.0071	-0.0057	-0.0018
	(0.0094)	(0.0090)	(0.0066)	(0.0053)	(0.0030)	(0.0025)	(0.0080)
Price of dairy	-0.0014	0.0121	-0.0139	0.0103	-0.0038	0.0011	-0.0043
	(0.0099)	(0.0094)	(0.0069)	(0.0055)	(0.0032)	(0.0027)	(0.0084)
Price of oils/fats	0.0423	-0.0129	-0.0080	-0.0015	-0.0048	0.0001	-0.0152
	(0.0130)	(0.0124)	(0.0090)	(0.0072)	(0.0042)	(0.0035)	(0.0110)
Price of sugar	-0.0510	0.0239	0.0080	0.0161	-0.0005	0.0001	0.0033
	(0.0124)	(0.0118)	(0.0086)	(0.0069)	(0.0040)	(0.0033)	(0.0105)
Price of other foods	-0.0167	0.0148	-0.0089	0.0154	-0.0012	0.0103	-0.0137
	(0.0135)	(0.0129)	(0.0094)	(0.0075)	(0.0043)	(0.0036)	(0.0114)
Expenditure	0.0545	0.1274	-0.0741	0.1102	0.0120	0.0032	-0.2330
	(0.0495)	(0.0472)	(0.0345)	(0.0275)	(0.0159)	(0.0133)	(0.0418)
Expenditure-squared	-0.0128	-0.0028	0.0031	-0.0064	-0.0022	-0.0022	0.0233
	(0.0036)	(0.0035)	(0.0025)	(0.0020)	(0.0012)	(0.0010)	(0.0031)
Household size	0.0131	-0.0080	0.0001	-0.0004	0.0005	0.0016	-0.0070
	(0.0011)	(0.0010)	(0.0008)	(0.0006)	(0.0004)	(0.0003)	(0.0009)
Race (1 if black)	0.0550	0.0495	-0.0240	-0.0189	-0.0093	0.0024	-0.0547
	(0.0122)	(0.0116)	(0.0085)	(0.0068)	(0.0039)	(0.0033)	(0.0103)
Rural (1 if rural)	0.0379	-0.0381	0.0198	-0.0148	-0.0027	0.0050	-0.0071
	(0.0069)	(0.0066)	(0.0048)	(0.0038)	(0.0022)	(0.0019)	(0.0058)
Education	-0.0005	-0.0009	-0.0001	0.0006	0.0000	-0.0003	0.0012
	(0.0006)	(0.0006)	(0.0005)	(0.0004)	(0.0002)	(0.0002)	(0.0005)
Dummy for 1998	-0.0660	0.0880	-0.0630	0.0342	-0.0291	-0.0215	0.0575
	(0.0213)	(0.0203)	(0.0148)	(0.0118)	(0.0068)	(0.0057)	(0.0180)
Dummy for 2004	-0.2047	0.1045	-0.0003	0.0817	-0.0194	-0.0357	0.0740
	(0.0355)	(0.0338)	(0.0247)	(0.0197)	(0.0114)	(0.0095)	(0.03000
Survey month	0.0019	-0.0021	-0.0070	0.0060	-0.0008	-0.0006	0.0025
	(0.0021)	(0.0020)	(0.0015)	(0.0012)	(0.0007)	(0.0006)	(0.0018)
$\hat{v}$	0.0555	-0.0606	0.0147	-0.0159	-0.0013	0.0047	0.0029
	(0.0154)	(0.0147)	(0.0108)	(0.0086)	(0.0050)	(0.0042)	(0.0130)

1. Standard errors in parentheses

Table A3 Parameter estimates for QUAIDS with and without endogeneity-adjustment

Variable	Eq.[1]	QUAIDS		Endogeneity-adjusted QUAIDS	
		Coeff.	Std. Err.	Coeff.	Std. Err.
Constant	α_1	0.1185	(0.0250)	0.1160	(0.0263)
	α_2	0.2965	(0.0234)	0.3093	(0.0248)
	α_3	0.2328	(0.0167)	0.2265	(0.0176)
	α_4	0.0404	(0.0141)	0.0380	(0.0149)
	α_5	0.0571	(0.0084)	0.0590	(0.0087)
	α_6	0.0151	(0.0071)	0.0160	(0.0073)
	α_7	0.2395	(0.0214)	0.2351	(0.0224)
Expenditure	β_1	-0.0950	(0.0119)	-0.0947	(0.0155)
	β_2	0.0327	(0.0113)	0.0465	(0.0146)
	β_3	-0.0142	(0.0084)	-0.0219	(0.0107)
	β_4	-0.0111	(0.0068)	-0.0128	(0.0088)
	β_5	-0.0215	(0.0038)	-0.0180	(0.0051)
	β_6	-0.0304	(0.0032)	-0.0297	(0.0043)
	β_7	0.1395	(0.0104)	0.1308	(0.0135)
Prices	γ_{11}	0.0220	(0.0118)	0.0225	(0.0130)
	γ_{21}	0.0146	(0.0089)	0.0106	(0.0095)
	γ_{31}	-0.0015	(0.0061)	0.0006	(0.0064)
	γ_{41}	-0.0067	(0.0054)	-0.0061	(0.0057)
	γ_{51}	0.0076	(0.0039)	0.0065	(0.0040)
	γ_{61}	-0.0095	(0.0032)	-0.0100	(0.0034)
	γ_{71}	-0.0265	(0.0083)	-0.0241	(0.0091)
	γ_{22}	-0.0172	(0.0112)	-0.0147	(0.0115)
	γ_{32}	-0.0073	(0.0060)	-0.0078	(0.0061)
	γ_{42}	-0.0028	(0.0052)	-0.0032	(0.0053)
	γ_{52}	-0.0003	(0.0037)	-0.0004	(0.0038)
	γ_{62}	0.0044	(0.0031)	0.0044	(0.0032)
	γ_{72}	0.0087	(0.0078)	0.0111	(0.0081)
	γ_{33}	0.0111	(0.0060)	0.0110	(0.0061)
	γ_{43}	-0.0044	(0.0038)	-0.0042	(0.0038)
	γ_{53}	0.0050	(0.0027)	0.0054	(0.0027)
	γ_{63}	-0.0077	(0.0023)	-0.0077	(0.0023)
	γ_{73}	0.0047	(0.0059)	0.0026	(0.0061)
	γ_{44}	0.0081	(0.0050)	0.0083	(0.0050)
	γ_{54}	-0.0026	(0.0027)	-0.0027	(0.0027)
	γ_{64}	0.0033	(0.0023)	0.0032	(0.0024)
	γ_{74}	0.0051	(0.0052)	0.0047	(0.0054)
	γ_{55}	-0.0045	(0.0039)	-0.0053	(0.0040)
	γ_{65}	0.0012	(0.0024)	0.0013	(0.0025)
	γ_{75}	-0.0063	(0.0037)	-0.0049	(0.0039)
	γ_{66}	0.0013	(0.0030)	0.0015	(0.0031)
	γ_{76}	0.0070	(0.0033)	0.0071	(0.0033)
	γ_{77}	0.0074	(0.0109)	0.0034	(0.0116)
Expenditure-squared	λ_1	-0.0143	(0.0041)	-0.0144	(0.0041)
	λ_2	-0.0002	(0.0038)	-0.0004	(0.0038)
	λ_3	0.0024	(0.0029)	0.0026	(0.0029)
	λ_4	-0.0081	(0.0023)	-0.0081	(0.0023)
	λ_5	-0.0018	(0.0013)	-0.0018	(0.0013)
	λ_6	-0.0037	(0.0011)	-0.0038	(0.0011)
	λ_7	0.0258	(0.0036)	0.0259	(0.0036)

Table A3 (Cont'd)

Household size	z_{11}	0.0099	(0.0007)	0.0099	(0.0009)
	z_{21}	-0.0044	(0.0007)	-0.0053	(0.0009)
	z_{31}	-0.0008	(0.0005)	-0.0003	(0.0006)
	z_{41}	0.0005	(0.0004)	0.0006	(0.0005)
	z_{51}	0.0005	(0.0002)	0.0003	(0.0003)
	z_{61}	0.0013	(0.0002)	0.0013	(0.0003)
	z_{71}	-0.0071	(0.0006)	-0.0066	(0.0008)
Race (1 if black)	z_{12}	0.0814	(0.0093)	0.0815	(0.0107)
	z_{22}	0.0257	(0.0089)	0.0331	(0.0101)
	z_{32}	-0.0189	(0.0065)	-0.0229	(0.0074)
	z_{42}	-0.0293	(0.0053)	-0.0302	(0.0060)
	z_{52}	-0.0096	(0.0031)	-0.0079	(0.0035)
	z_{62}	0.0040	(0.0026)	0.0045	(0.0030)
	z_{72}	-0.0534	(0.0080)	-0.0580	(0.0092)
Rural/Urban (1 if rural)	z_{13}	0.0499	(0.0066)	0.0502	(0.0066)
	z_{23}	-0.0482	(0.0062)	-0.0478	(0.0062)
	z_{33}	0.0193	(0.0045)	0.0190	(0.0045)
	z_{43}	-0.0159	(0.0036)	-0.0160	(0.0036)
	z_{53}	-0.0019	(0.0021)	-0.0017	(0.0022)
	z_{63}	0.0061	(0.0018)	0.0060	(0.0018)
	z_{73}	-0.0093	(0.0055)	-0.0098	(0.0055)
Education of head	z_{14}	-0.0018	(0.0006)	-0.0018	(0.0006)
	z_{24}	0.0005	(0.0005)	0.0002	(0.0006)
	z_{34}	-0.0004	(0.0004)	-0.0002	(0.0004)
	z_{44}	0.0009	(0.0003)	0.0010	(0.0003)
	z_{54}	-9.53e-06	(0.0002)	-0.0001	(0.0002)
	z_{64}	-0.0004	(0.0001)	-0.0004	(0.0002)
	z_{74}	0.0011	(0.0005)	0.0013	(0.0005)
Survey year (=1 if 1998)	z_{15}	-0.0426	(0.0159)	-0.0411	(0.0165)
	z_{25}	0.0498	(0.0156)	0.0549	(0.0160)
	z_{35}	-0.0530	(0.0111)	-0.0561	(0.0114)
	z_{45}	0.0186	(0.0107)	0.0187	(0.0109)
	z_{55}	-0.0256	(0.0061)	-0.0243	(0.0062)
	z_{65}	-0.0161	(0.0051)	-0.0162	(0.0052)
	z_{75}	0.0689	(0.0140)	0.0641	(0.0144)
Survey year (=1 if 2004)	z_{16}	-0.0917	(0.0204)	-0.0866	(0.0218)
	z_{26}	0.0213	(0.0213)	0.0284	(0.0219)
	z_{36}	-0.0192	(0.0157)	-0.0247	(0.0164)
	z_{46}	0.0410	(0.0144)	0.0409	(0.0149)
	z_{56}	-0.0150	(0.0094)	-0.0119	(0.0098)
	z_{66}	-0.0280	(0.0078)	-0.0283	(0.0083)
	z_{76}	0.0915	(0.0185)	0.0822	(0.0195)
Month of survey	z_{17}	0.0027	(0.0021)	0.0027	(0.0021)
	z_{27}	-0.0011	(0.0020)	-0.0010	(0.0020)
	z_{37}	-0.0076	(0.0014)	-0.0077	(0.0014)
	z_{47}	0.0047	(0.0012)	0.0047	(0.0012)
	z_{57}	-0.0008	(0.0007)	-0.0008	(0.0007)
	z_{67}	-0.0007	(0.0006)	-0.0007	(0.0006)
	z_{77}	0.0028	(0.0018)	0.0027	(0.0018)

Table A3 (Cont'd)

Residuals	v_1	-0.0005	(0.0121)
	v_2	-0.0173	(0.0113)
	v_3	0.0097	(0.0082)
	v_4	0.0021	(0.0067)
	v_5	-0.0040	(0.0040)
	v_6	-0.0008	(0.0034)
	v_7	0.0109	(0.0101)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A4 Own and Cross Price Elasticities estimated from the QUAIDS Mmodel with and without Endogeneity Adjustment

I. QUAIDS without adjusting for expenditure endogeneity

Marshallian/Uncompensated own-price elasticities[1,2]

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-1.0296						
	(0.0742)						
MF	0.4959	-1.1077					
	(0.1322)	(0.1586)					
FV	0.2624	-0.0604	-0.9803				
	(0.0773)	(0.0891)	(0.0957)				
DAI	0.0126	-0.0181	-0.0272	-0.8143			
	(0.0309)	(0.0242)	(0.0248)	(0.0788)			
OBF	0.0416	-0.0090	0.0320	-0.0231	-1.0604		
	(0.0180)	(0.0164)	(0.0165)	(0.0504)	(0.0745)		
SUG	-0.0730	0.0152	-0.0263	-0.0231	-0.0081	-1.0463	
	(0.0193)	(0.0220)	(0.0218)	(0.0504)	(0.0548)	(0.0749)	
OTH	0.5649	0.0148	-0.1322	-0.0852	0.4248	1.0110	-4.6426
	(0.1804)	(0.2253)	(0.2176)	(0.0556)	(0.3331)	(0.2576)	(0.6605)

Hicksian/Uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.9032						
	(0.0408)						
MF	0.6222	-0.8455					
	(0.0947)	(0.0306)					
FV	0.3887	0.2018	-0.8087				
	(0.0441)	(0.0606)	(0.0193)				
DAI	0.1390	0.2441	0.1444	-0.7715			
	(0.0401)	(0.0300)	(0.0251)	(0.0100)			
OBF	0.1680	0.2532	0.2036	0.0197	-1.0307		
	(0.0409)	(0.0307)	(0.0220)	(0.0491)	(0.0053)		
SUG	0.0534	0.2774	0.1453	-0.0423	0.0216	-1.0320	
	(0.0552)	(0.0498)	(0.0386)	(0.0623)	(0.0576)	(0.0043)	
OTH	0.6912	0.2770	0.0394	1.7573	0.4546	1.3074	-4.3462
	(0.1412)	(0.1954)	(0.1989)	(0.4879)	(0.3282)	(0.2633)	(0.0181)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A4 (Cont'd)

II. Endogeneity-adjusted QUAIDS model

Marshallian/Uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-1.0458						
	(0.0856)						
MF	0.5888	-1.1058					
	(0.1651)	(0.1943)					
FV	0.2638	-0.0708	-0.9739				
	(0.0814)	(0.0882)	(0.0967)				
DAI	0.0086	-0.0221	-0.0235	-0.8238			
	(0.0371)	(0.0237)	(0.0252)	(0.0977)			
OBF	0.0517	-0.0125	0.0327	0.0127	-1.0679		
	(0.0228)	(0.0187)	(0.0189)	(0.0657)	(0.0755)		
SUG	-0.0777	0.0125	-0.0232	-0.0941	-0.0081	-1.0477	
	(0.0222)	(0.0230	(0.0231)	(0.0645)	(0.0563)	(0.0842)	
OTH	0.6149	0.0254	-0.1574	1.8302	0.4623	1.0634	-4.9186
	(0.1965)	(0.2414)	(0.2333)	(0.5426)	(0.3570)	(0.2785)	(0.7422)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.9330						
	(0.0456)						
MF	0.7017	-0.8312					
	(0.1232)	(0.0342)					
FV	0.3766	0.2038	-0.8082				
	(0.0484)	(0.0572)	(0.0217)				
DAI	0.1215	0.2525	0.1422	-0.7842			
	(0.0485)	(0.0337)	(0.0278)	(0.0118)			
OBF	0.1645	0.2622	0.1984	0.0523	-1.0334)		
	(0.0430)	(0.0293)	(0.0205)	(0.0624)	(0.0064)		
SUG	0.0351	0.2871	0.1424	-0.0545	0.0248	-1.0333	
	(0.0610)	(0.0538)	(0.0416)	(0.0709)	(0.0597)	(0.0053)	
OTH	0.7278	0.3000	0.0083	1.8698	0.4952	1.3587	-4.6264
	(0.1540)	(0.2093)	(0.2134)	(0.5328)	(0.3519)	(0.2861)	(0.0210)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A5 (Part I) Expenditure and Price Elasticities estimated using Data from only Households who do not engage in Own-production

Expenditure elasticities[1]

	GR	MF	FV	DAI	OBF	SUG	OTH
Expenditure elasticity	0.5201	0.8705	1.1284	0.9217	0.4113	0.1425	2.8063
	(0.1670)	(0.1686)	(0.1207)	(0.1863)	(0.1133)	(0.1138)	(0.2318)

Marshallian/Uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.9315						
	(0.0646)						
MF	0.3583	-0.8689					
	(0.1637)	(0.1654)					
FV	0.1872	0.1235	-1.1035				
	(0.0887)	(0.1071)	(0.1198)				
DAI	0.0377	0.0267	-0.0944	-0.7855			
	(0.0449)	(0.0518)	(0.0514)	(0.0942)			
OBF	0.0307	-0.0038	0.0658	-0.0653	-1.2067		
	(0.0181)	(0.0219)	(0.0261)	(0.0474)	(0.0880)		
SUG	-0.0717	-0.0295	0.0153	-0.0717	-0.1134	-1.1248	
	(0.0272)	(0.0329)	(0.0337)	(0.0737)	(0.0715)	(0.1130)	
OTH	0.3962	0.4373	-0.5962	1.1016	1.0061	1.7515	-4.3722
	(0.2350)	(0.2827)	(0.2458)	(0.6089)	(0.4066)	(0.3708)	(0.8207)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.7682						
	(0.0524)						
MF	0.5217	-0.6601					
	(0.1143)	(0.0404)					
FV	0.3506	0.3323	-0.9162				
	(0.0435)	(0.0705)	(0.0200)				
DAI	0.2011	0.2355	0.0930	-0.7214			
	(0.0348)	(0.0344)	(0.0418)	(0.0130)			
OBF	0.1941	0.2050	0.2531	-0.0012	-1.1838		
	(0.0550)	(0.0456)	(0.0329)	(0.0489)	(0.0063)		
SUG	0.0916	0.1793	0.2026	-0.0076	-0.0905	-1.1181	
	(0.0762)	(0.0698)	(0.0502)	(0.0844)	(0.0754)	(0.0053)	
OTH	0.5596	0.6461	-0.4089	1.1657	1.0290	2.0544	-4.0693
	(0.1840)	(0.2433)	(0.2268)	(0.5965)	(0.4007)	(0.3820)	(0.0250)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A5 (Part II) Expenditure and Price Elasticities estimated using Data on all Households including those who Attrited

Expenditure elasticities[1]

	GR	MF	FV	DAI	OBF	SUG	OTH
Expenditure elasticity	0.6048	1.0221	1.0191	1.0803	0.6479	0.4868	2.3860
	(0.0793)	(0.0794)	(0.0597)	(0.0694)	(0.0413)	(0.0400)	(0.0958)

Marshallian/Uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.9888						
	(0.0333)						
MF	0.3560	-0.8343					
	(0.0854)	(0.1079)					
FV	0.1351	-0.0203	-1.0089				
	(0.0488)	(0.0616)	(0.0643)				
DAI	0.0500	-0.0032	-0.0149	-1.1882			
	(0.0224)	(0.0279)	(0.0263)	(0.0909))			
OBF	0.0042	-0.0427	-0.0005	0.1972	-0.8863		
	(0.0122)	(0.0158)	(0.0150)	(0.0422)	(0.0551)		
SUG	-0.0168	0.0097	-0.0176	-0.0257	-0.0790	-0.8336	
	(0.0098)	(0.0127)	(0.0119)	(0.0379)	(0.0389)	(0.0648)	
OTH	0.1990	0.0752	-0.1776	0.4502	0.3779	0.5165	-2.5714
	(0.0890)	(0.1140)	(0.1081)	(0.2323)	(0.1646)	(0.1461)	(0.3248)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.8008						
	(0.0246)						
MF	0.5440	-0.5944					
	(0.0625)	(0.0186)					
FV	0.3230	0.2195	-0.8302				
	(0.0276)	(0.0449)	(0.0105)				
DAI	0.2379	0.2366	0.1638	-1.1083			
	(0.0228)	(0.0240)	(0.0229)	(0.0051)			
OBF	0.1922	0.1971	0.1782	0.2772	-0.8507		
	(0.0210)	(0.0171)	(0.0132)	(0.0405)	(0.0023)		
SUG	0.1712	0.2495	0.1611	0.0543	-0.0434	-0.8104	
	(0.0276)	(0.0237)	(0.0168)	(0.0385)	(0.0390)	(0.0019)	
OTH	0.3870	0.3150	0.0012	0.5302	0.4134	0.7132	-2.3748
	(0.0659)	(0.0964)	(0.0983)	(0.2278)	(0.1627)	(0.1477)	(0.0079)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A6 Parameter Estimates from the AIDS Model with and without Endogeneity Ajustment

Variable	Parameter/ Equation #[1]	AIDS		Endogeneity-adjusted AIDS	
		Coeff.	Std. Err.	Coeff.	Std. Err.
Constant	α_1	0.1398	(0.0242)	0.1378	(0.0256)
	α_2	0.3156	(0.0233)	0.3299	(0.0246)
	α_3	0.2180	(0.0169)	0.2129	(0.0178)
	α_4	0.0353	(0.0142)	0.0343	(0.0149)
	α_5	0.0605	(0.0083)	0.0628	(0.0086)
	α_6	0.0173	(0.0071)	0.0192	(0.0073)
	α_7	0.2135	(0.0208)	0.2031	(0.0218)
Expenditure	β_1	-0.0557	(0.0045)	-0.055	(0.011)
	β_2	0.0331	(0.0042)	0.050	(0.010)
	β_3	-0.0207	(0.0031)	-0.028	(0.008)
	β_4	0.0112	(0.0025)	0.011	(0.006)
	β_5	-0.0165	(0.0014)	-0.013	(0.004)
	β_6	-0.0201	(0.0012)	-0.018	(0.003)
	β_7	0.0687	(0.0038)	0.052	(0.010)
Prices	γ_{11}	0.0167	(0.0117)	0.0171	(0.0126)
	γ_{21}	0.0144	(0.0088)	0.0104	(0.0094)
	γ_{31}	-0.0006	(0.0060)	0.0010	(0.0063)
	γ_{41}	-0.0083	(0.0054)	-0.0083	(0.0056)
	γ_{51}	0.0066	(0.0039)	0.0055	(0.0040)
	γ_{61}	-0.0108	(0.0032)	-0.0119	(0.0033)
	γ_{71}	-0.0181	(0.0081)	-0.0138	(0.0087)
	γ_{22}	-0.0126	(0.0112)	-0.0097	(0.0115)
	γ_{32}	-0.0101	(0.0060)	-0.0105	(0.0061)
	γ_{42}	-0.0052	(0.0052)	-0.0053	(0.0053)
	γ_{52}	-0.0002	(0.0037)	-0.0002	(0.0038)
	γ_{62}	0.0035	(0.0031)	0.0038	(0.0032)
	γ_{72}	0.0102	(0.0077)	0.0114	(0.0080)
	γ_{33}	0.0128	(0.0061)	0.0128	(0.0061)
	γ_{43}	-0.0027	(0.0038)	-0.0027	(0.0038)
	γ_{53}	0.0051	(0.0027)	0.0054	(0.0027)
	γ_{63}	-0.0071	(0.0023)	-0.0072	(0.0023)
	γ_{73}	0.0026	(0.0059)	0.0011	(0.0060)
	γ_{44}	0.0084	(0.0050)	0.0085	(0.0050)
	γ_{54}	-0.0030	(0.0027)	-0.0032	(0.0027)
	γ_{64}	0.0034	(0.0023)	0.0033	(0.0024)
	γ_{74}	0.0074	(0.0052)	0.0076	(0.0053)
	γ_{55}	-0.0045	(0.0039)	-0.0053	(0.0040)
	γ_{65}	0.0008	(0.0024)	0.0008	(0.0025)
	γ_{75}	-0.0048	(0.0037)	-0.0030	(0.0038)
	γ_{66}	0.0022	(0.0031)	0.0029	(0.0032)
	γ_{76}	0.0080	(0.0033)	0.0083	(0.0033)
	γ_{77}	-0.0053	(0.0107)	-0.0116	(0.0111)
Household size	z_{11}	0.0096	(0.0007)	0.0096	(0.0009)
	z_{21}	-0.0045	(0.0007)	-0.0055	(0.0009)
	z_{31}	-0.0006	(0.0005)	-0.0002	(0.0006)
	z_{41}	0.0004	(0.0004)	0.0004	(0.0005)
	z_{51}	0.0005	(0.0002)	0.0003	(0.0003)
	z_{61}	0.0013	(0.0002)	0.0011	(0.0003)
	z_{71}	-0.0067	(0.0006)	-0.0057	(0.0008)

Table A6 (Cont'd)

Race (1 if black)	z_{12}	0.0852	(0.0093)	0.0858	(0.0107)
	z_{22}	0.0254	(0.0088)	0.0338	(0.0101)
	z_{32}	-0.0192	(0.0065)	-0.0230	(0.0073)
	z_{42}	-0.0266	(0.0052)	-0.0265	(0.0060)
	z_{52}	-0.0091	(0.0031)	-0.0071	(0.0035)
	z_{62}	0.0051	(0.0026)	0.0064	(0.0030)
	z_{72}	-0.0607	(0.0080)	-0.0694	(0.0092)
Rural/Urban (1 if rural)	z_{13}	0.0489	(0.0066)	0.0492	(0.0066)
	z_{23}	-0.0483	(0.0062)	-0.0478	(0.0062)
	z_{33}	0.0196	(0.0045)	0.0193	(0.0045)
	z_{43}	-0.0165	(0.0036)	-0.0165	(0.0036)
	z_{53}	-0.0020	(0.0021)	-0.0019	(0.0022)
	z_{63}	0.0059	(0.0018)	0.0059	(0.0018)
	z_{73}	-0.0075	(0.0056)	-0.0081	(0.0056)
Education of head	z_{14}	-0.0018	(0.0006)	-0.0018	(0.0006)
	z_{24}	0.0005	(0.0005)	0.0001	(0.0006)
	z_{34}	-0.0004	(0.0004)	-0.0002	(0.0004)
	z_{44}	0.0010	(0.0003)	0.0010	(0.0003)
	z_{54}	0.0000	(0.0002)	-0.0001	(0.0002)
	z_{64}	-0.0004	(0.0001)	-0.0004	(0.0002)
	z_{74}	0.0011	(0.0005)	0.0015	(0.0005)
Survey year (=1 if 1998)	z_{15}	-0.0438	(0.0159)	-0.0425	(0.0166)
	z_{25}	0.0375	(0.0156)	0.0444	(0.0160)
	z_{35}	-0.0453	(0.0112)	-0.0487	(0.0116)
	z_{45}	0.0280	(0.0107)	0.0287	(0.0109)
	z_{55}	-0.0262	(0.0061)	-0.0247	(0.0062)
	z_{65}	-0.0143	(0.0052)	-0.0138	(0.0052)
	z_{75}	0.0641	(0.0141)	0.0566	(0.0146)
Survey year (=1 if 2004)	z_{16}	-0.0945	(0.0204)	-0.0901	(0.0221)
	z_{26}	0.0038	(0.0213)	0.0140	(0.0220)
	z_{36}	-0.0086	(0.0159)	-0.0144	(0.0167)
	z_{46}	0.0532	(0.0146)	0.0541	(0.0150)
	z_{56}	-0.0166	(0.0094)	-0.0136	(0.0099)
	z_{66}	-0.0242	(0.0078)	-0.0227	(0.0084)
	z_{76}	0.0869	(0.0188)	0.0727	(0.0199)
Month of survey	z_{17}	0.0025	(0.0021)	0.0026	(0.0021)
	z_{27}	-0.0023	(0.0020)	-0.0022	(0.0020)
	z_{37}	-0.0069	(0.0015)	-0.0069	(0.0015)
	z_{47}	0.0058	(0.0012)	0.0059	(0.0012)
	z_{57}	-0.0009	(0.0007)	-0.0008	(0.0007)
	z_{67}	-0.0006	(0.0006)	-0.0006	(0.0006)
	z_{77}	0.0023	(0.0018)	0.0021	(0.0018)
Residuals 1: expenditure	v_{11}			-0.0014	(0.0123)
	v_{21}			-0.0199	(0.0114)
	v_{31}			0.0088	(0.0083)
	v_{41}			-0.0001	(0.0067)
	v_{51}			-0.0046	(0.0041)
	v_{61}			-0.0029	(0.0035)
	v_{71}			0.0200	(0.0104)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A7 Own and Cross Price Elasticities estimated from the AIDS Model with and without Endogeneity-Adjustment

I. AIDS without adjusting for expenditure endogeneity

Marshallian/Uncompensated own-price elasticities[1]

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.9279						
	(0.0356)						
MF	0.0618	-1.0996					
	(0.0267)	(0.0472)					
FV	0.0103	-0.0511	-0.9015				
	(0.0184)	(0.0258)	(0.0339)				
DAI	-0.0231	-0.0239	-0.0146	-0.8797			
	(0.0164)	(0.0223)	(0.0218)	(0.0774)			
OBF	0.0235	-0.0030	0.0303	-0.0464	-1.0668		
	(0.0118)	(0.0160)	(0.0151)	(0.0411)	(0.0733)		
SUG	-0.0318	0.0142	-0.0395	0.0518	0.0156	-0.9455	
	(0.0097)	(0.0135)	(0.0130)	(0.0357)	(0.0459)	(0.0622)	
OTH	-0.0432	0.0370	0.0191	0.1104	-0.0858	0.1537	-1.2065
	(0.0246)	(0.0333)	(0.0332)	(0.0787)	(0.0694)	(0.0612)	(0.1029)

Hicksian/Uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.6580						
	(0.0350)						
MF	0.2797	-0.8357					
	(0.0268)	(0.0471)					
FV	0.1767	0.1310	-0.7459				
	(0.0184)	(0.0258)	(0.0338)				
DAI	0.0390	0.0440	0.0499	-0.8027			
	(0.0164)	(0.0223)	(0.0218)	(0.0773)			
OBF	0.0739	0.0522	0.0825	0.0075	-1.0300		
	(0.0118)	(0.0160)	(0.0150)	(0.0411)	(0.0731)		
SUG	0.0142	0.0613	0.0082	0.1010	0.0635	-0.9169	
	(0.0097)	(0.0135)	(0.0130)	(0.0357)	(0.0459)	(0.0623)	
OTH	0.0509	0.1399	0.1166	0.2111	0.0121	0.2513	-1.0382
	(0.0246)	(0.0333)	(0.0332)	(0.0787)	(0.0694)	(0.0612)	(0.1031)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A7 (Cont'd)

II. Endogeneity-adjusted AIDS model

Marshallian/Uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.9274						
	(0.0395)						
MF	0.0498	-1.1129					
	(0.0278)	(0.0476)					
FV	0.0148	-0.0559	-0.8935				
	(0.0192)	(0.0263)	(0.0346)				
DAI	-0.0231	-0.0250	-0.0142	-0.8775			
	(0.0172)	(0.0226)	(0.0218)	(0.0778)			
OBF	0.0201	-0.0039	0.0324	-0.0486	-1.0845		
	(0.0123)	(0.0164)	(0.0152)	(0.0414)	(0.0761)		
SUG	-0.0350	0.0153	-0.0399	0.0498	0.0144	-0.9312	
	(0.0101)	(0.0135)	(0.0130)	(0.0359)	(0.0474)	(0.0641)	
OTH	-0.0310	0.0392	0.0120	0.1127	-0.0534	0.1586	-1.2240
	(0.0258)	(0.0338)	(0.0335)	(0.0803)	(0.0715)	(0.0619)	(0.1037)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.6563						
	(0.0357)						
MF	0.2680	-0.8325					
	(0.0282)	(0.0470)					
FV	0.1814	0.1290	-0.7454				
	(0.0191)	(0.0261)	(0.0338)				
DAI	0.0391	0.0441	0.0498	-0.8003			
	(0.0172)	(0.0225)	(0.0218)	(0.0774)			
OBF	0.0705	0.0521	0.0842	0.0053	-1.0439		
	(0.0122)	(0.0163)	(0.0151)	(0.0414)	(0.0751)		
SUG	0.0110	0.0616	0.0075	0.0990	0.0625	-0.9001	
	(0.0102)	(0.0136)	(0.0130)	(0.0359)	(0.0474)	(0.0650)	
OTH	0.0631	0.1437	0.1088	0.2134	0.0450	0.2565	-1.0726
	(0.0261)	(0.0339)	(0.0337)	(0.0804)	(0.0716)	(0.0620)	(0.1051)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

[2]. Standard errors in parentheses

Table A8 Parameter Estimates for QUAIDS: Rural-Urban Differences

Variable	Eq.[1]	RURAL		URBAN		URBAN	
				Endog.-corrected		No endg. correc.	
		Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Constant	α_1	0.2480	(0.0346)	0.0322	(0.0372)	0.0505	(0.0352)
	α_2	0.2657	(0.0292)	0.3064	(0.0414)	0.2883	(0.0393)
	α_3	0.2538	(0.0220)	0.2196	(0.0263)	0.2184	(0.0251)
	α_4	0.0056	(0.0181)	0.0715	(0.0248)	0.0598	(0.0240)
	α_5	0.0497	(0.0108)	0.0751	(0.0146)	0.0744	(0.0141)
	α_6	0.0324	(0.0098)	0.0005	(0.0098)	0.0043	(0.0095)
	α_7	0.1449	(0.0250)	0.2947	(0.0424)	0.3041	(0.0404)
Expenditure	β_1	-0.1282	(0.0149)	-0.0701	(0.0223)	-0.0509	(0.0191)
	β_2	0.0821	(0.0137)	-0.0649	(0.0225)	-0.0831	(0.0191)
	β_3	-0.0080	(0.0101)	-0.0061	(0.0164)	-0.0070	(0.0141)
	β_4	0.0027	(0.0079)	-0.0074	(0.0142)	-0.0198	(0.0121)
	β_5	-0.0111	(0.0046)	-0.0323	(0.0082)	-0.0332	(0.0066)
	β_6	-0.0356	(0.0042)	-0.0292	(0.0053)	-0.0243	(0.0043)
	β_7	0.0980	(0.0112)	0.2100	(0.0248)	0.2181	(0.0214)
Prices	γ_{11}	0.0073	(0.0170)	0.0260	(0.0171)	0.0156	(0.0157)
	γ_{21}	0.0262	(0.0121)	-0.0063	(0.0139)	-0.0003	(0.0131)
	γ_{31}	-0.0057	(0.0080)	0.0126	(0.0097)	0.0136	(0.0092)
	γ_{41}	-0.0002	(0.0070)	0.0067	(0.0089)	0.0118	(0.0085)
	γ_{51}	0.0046	(0.0052)	0.0093	(0.0063)	0.0081	(0.0061)
	γ_{61}	-0.0144	(0.0047)	0.0022	(0.0041)	0.0005	(0.0041)
	γ_{71}	-0.0178	(0.0108)	-0.0505	(0.0141)	-0.0492	(0.0132)
	γ_{22}	-0.0311	(0.0140)	0.0097	(0.0190)	0.0089	(0.0187)
	γ_{32}	-0.0016	(0.0073)	-0.0132	(0.0105)	-0.0142	(0.0104)
	γ_{42}	-0.0034	(0.0063)	-0.0001	(0.0096)	-0.0031	(0.0096)
	γ_{52}	0.0024	(0.0048)	-0.0010	(0.0066)	0.0001	(0.0063)
	γ_{62}	0.0071	(0.0044)	0.0010	(0.0043)	0.0029	(0.0042)
	γ_{72}	0.0004	(0.0093)	0.0099	(0.0144)	0.0058	(0.0142)
	γ_{33}	0.0046	(0.0073)	0.0239	(0.0112)	0.0244	(0.0111)
	γ_{43}	-0.0046	(0.0044)	-0.0102	(0.0076)	-0.0113	(0.0076)
	γ_{53}	0.0056	(0.0031)	0.0008	(0.0054)	0.0002	(0.0052)
	γ_{63}	-0.0075	(0.0029)	-0.0046	(0.0037)	-0.0038	(0.0036)
	γ_{73}	0.0092	(0.0067)	-0.0093	(0.0110)	-0.0089	(0.0109)
	γ_{44}	0.0029	(0.0057)	0.0119	(0.0102)	0.0085	(0.0101)
	γ_{54}	0.0003	(0.0031)	-0.0187	(0.0056)	-0.0165	(0.0053)
	γ_{64}	0.0061	(0.0029)	-0.0019	(0.0039)	0.0001	(0.0038)
	γ_{74}	-0.0011	(0.0060)	0.0122	(0.0098)	0.0104	(0.0098)
	γ_{55}	-0.0079	(0.0047)	0.0067	(0.0079)	0.0077	(0.0070)
	γ_{65}	0.0012	(0.0031)	-0.0023	(0.0044)	-0.0047	(0.0040)
	γ_{75}	-0.0061	(0.0046)	0.0052	(0.0068)	0.0051	(0.0065)
	γ_{66}	-0.0007	(0.0041)	-0.0024	(0.0053)	-0.0042	(0.0052)
	γ_{76}	0.0081	(0.0043)	0.0081	(0.0052)	0.0092	(0.0050)
	γ_{77}	0.0072	(0.0126)	0.0243	(0.0214)	0.0276	(0.0209)
Expenditure-squared	λ_1	-0.0267	(0.0052)	0.0142	(0.0072)	0.0149	(0.0072)
	λ_2	0.0158	(0.0047)	-0.0415	(0.0073)	-0.0417	(0.0073)
	λ_3	0.0034	(0.0034)	0.0059	(0.0054)	0.0057	(0.0054)
	λ_4	-0.0045	(0.0027)	-0.0136	(0.0047)	-0.0139	(0.0047)
	λ_5	0.0007	(0.0016)	-0.0050	(0.0025)	-0.0050	(0.0025)
	λ_6	-0.0050	(0.0014)	-0.0030	(0.0016)	-0.0029	(0.0016)
	λ_7	0.0163	(0.0039)	0.0430	(0.0085)	0.0429	(0.0085)

Table A8 (Cont'd)

Household size	z_{11}	0.0102	(0.0009)	0.0156	(0.0013)	0.0149	(0.0012)
	z_{21}	-0.0051	(0.0008)	-0.0032	(0.0014)	-0.0025	(0.0013)
	z_{31}	-0.0008	(0.0006)	-0.0027	(0.0009)	-0.0027	(0.0009)
	z_{41}	0.0002	(0.0005)	-0.0016	(0.0008)	-0.0011	(0.0008)
	z_{51}	0.0003	(0.0003)	0.0002	(0.0005)	0.0002	(0.0004)
	z_{61}	0.0012	(0.0002)	0.0023	(0.0003)	0.0021	(0.0003)
	z_{71}	-0.0062	(0.0006)	-0.0105	(0.0014)	-0.0108	(0.0013)
Education of head	z_{14}	0.0016	(0.0007)	-0.0024	(0.0009)	0.0031	(0.0008)
	z_{24}	-0.0009	(0.0007)	-0.0014	(0.0010)	0.0007	(0.0008)
	z_{34}	-0.0002	(0.0005)	-0.0008	(0.0006)	0.0008	(0.0006)
	z_{44}	-0.0007	(0.0004)	0.0016	(0.0006)	-0.0020	(0.0005)
	z_{54}	0.0000	(0.0002)	0.0000	(0.0003)	0.0000	(0.0003)
	z_{64}	0.0004	(0.0002)	-0.0005	(0.0002)	0.0006	(0.0002)
	z_{74}	-0.0001	(0.0005)	0.0035	(0.0010)	-0.0032	(0.0009)
Survey year (=1 if 1998)	z_{15}	-0.0681	(0.0209)	0.0319	(0.0240)	0.0414	(0.0236)
	z_{25}	0.0906	(0.0192)	0.0172	(0.0260)	0.0077	(0.0261)
	z_{35}	-0.0605	(0.0140)	-0.0439	(0.0179)	-0.0451	(0.0176)
	z_{45}	0.0287	(0.0128)	0.0003	(0.0198)	-0.0070	(0.0193)
	z_{55}	-0.0188	(0.0074)	-0.0539	(0.0115)	-0.0523	(0.0108)
	z_{65}	-0.0208	(0.0068)	-0.0106	(0.0076)	-0.0068	(0.0072)
	z_{75}	0.0489	(0.0163)	0.0591	(0.0262)	0.0621	(0.0261)
Survey year (=1 if 2004)	z_{16}	-0.1247	(0.0264)	0.0285	(0.0337)	0.0433	(0.0320)
	z_{26}	0.0668	(0.0255)	-0.0182	(0.0388)	-0.0331	(0.0386)
	z_{36}	-0.0320	(0.0190)	-0.0022	(0.0291)	-0.0014	(0.0278)
	z_{46}	0.0439	(0.0167)	0.0294	(0.0297)	0.0183	(0.0282)
	z_{56}	-0.0011	(0.0109)	-0.0771	(0.0189)	-0.0795	(0.0173)
	z_{66}	-0.0336	(0.0099)	-0.0250	(0.0127)	-0.0195	(0.0122)
	z_{76}	0.0806	(0.0213)	0.0647	(0.0359)	0.0718	(0.0347)
Month of survey	z_{17}	0.0024	(0.0027)	0.0061	(0.0031)	0.0063	(0.0031)
	z_{27}	0.0017	(0.0025)	-0.0034	(0.0032)	-0.0035	(0.0032)
	z_{37}	-0.0091	(0.0019)	-0.0050	(0.0021)	-0.0049	(0.0021)
	z_{47}	0.0052	(0.0015)	0.0026	(0.0019)	0.0025	(0.0019)
	z_{57}	-0.0010	(0.0008)	-0.0010	(0.0011)	-0.0010	(0.0011)
	z_{67}	-0.0019	(0.0008)	0.0013	(0.0007)	0.0014	(0.0007)
	z_{77}	0.0027	(0.0020)	-0.0007	(0.0034)	-0.0007	(0.0034)
Residuals	v_1			0.0237	(0.0150)		
	v_2			-0.0239	(0.0157)		
	v_3			-0.0004	(0.0108)		
	v_4			-0.0153	(0.0096)		
	v_5			-0.0016	(0.0059)		
	v_6			0.0058	(0.0038)		
	v_7			0.0117	(0.0160)		

[1]. The commodities represented by the different equation numbers are: 1 = grains; 2 = meat and fish; 3 = fruits and vegetables; 4 = dairy; 5 = oils, butter, and other fats; 6 = sugar; 7 = other foods

Table A9 Own and Cross Price Elasticities: Rural-Urban Differences

I. RURAL HOUSEHOLDS

Marshallian/Uncompensated own-price elasticities[1]

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.8399						
	(0.0815)						
MF	-0.0280	-0.6579					
	(0.1044)	(0.1789)					
FV	-0.0944	0.5106	-0.9542				
	(0.1168)	(0.1844)	(0.1577)				
DAI	0.0498	0.0329	-0.0708	-0.8600			
	(0.0537)	(0.1266)	(0.0531)	(0.1294)			
OBF	0.0556	0.0031	0.0085	-0.1951	-0.8278		
	(0.0348)	(0.0709)	(0.0350)	(0.0784)	(0.1352)		
SUG	0.0552	-0.1203	0.0064	-0.1353	-0.1607	-1.1789	
	(0.0327)	(0.0498)	(0.0349)	(0.0685)	(0.0880)	(0.1463)	
OTH	-0.9685	2.2556	-0.5096	2.4149	1.5436	2.0453	-5.4848
	(0.4150)	(0.4872)	(0.4592)	(0.8329)	(0.6892)	(0.7210)	(1.1737)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.5878						
	(0.0468)						
MF	0.2242	-0.6526					
	(0.0906)	(0.0499)					
FV	0.1577	0.5159	-0.7788				
	(0.0808)	(0.1536)	(0.0282)				
DAI	0.3019	0.0382	0.1046	-0.8129			
	(0.0709)	(0.1216)	(0.0545)	(0.0182)			
OBF	0.3077	0.0084	0.1840	-0.1480	-0.8106		
	(0.0575)	(0.0835)	(0.0423)	(0.0781)	(0.0087)		
SUG	0.3073	-0.1149	0.1819	-0.0882	-0.1435	-1.1685	
	(0.0738)	(0.0824)	(0.0587)	(0.0806)	(0.0919)	(0.0052)	
OTH	-0.7164	2.2609	-0.3341	2.4620	1.5608	2.5031	-5.0271
	(0.3702)	(0.4424)	(0.4318)	(0.8157)	(0.6810)	(0.7361)	(0.0402)

Table A9 (Cont'd)

I. URBAN HOUSEHOLDS

Marshallian/Uncompensated own-price elasticities[1]

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.7474						
	(0.0998)						
MF	-0.0751	-0.5248					
	(0.1282)	(0.1889)					
FV	-0.0686	0.4852	-0.9598				
	(0.1185)	(0.1942)	(0.1568)				
DAI	-0.0004	0.1449	-0.0876	-0.7245			
	(0.0631)	(0.1441)	(0.0625)	(0.1383)			
OBF	0.0631	0.0035	0.0097	-0.2214	-0.8409		
	(0.0370)	(0.0811)	(0.0368)	(0.0951)	(0.1440)		
SUG	0.0715	-0.1611	0.0102	-0.1912	-0.1396	-1.1547	
	(0.0386)	(0.0580)	(0.0406)	(0.0835)	(0.0945)	(0.1502)	
OTH	-0.8574	2.0804	-0.4953	2.1835	1.4593	1.9692	-5.1792
	(0.3894)	(0.4748)	(0.4280)	(0.7891)	(0.6475)	(0.6798)	(1.1354)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.5191						
	(0.0474)						
MF	0.1532	-0.4934					
	(0.1059)	(0.0502)					
FV	0.1597	0.5166	-0.7835				
	(0.0855)	(0.1660)	(0.0291)				
DAI	0.2279	0.1762	0.0887	-0.6640			
	(0.0671)	(0.1331)	(0.0539)	(0.0196)			
OBF	0.2914	0.0349	0.1860	-0.1609	-0.8227		
	(0.0566)	(0.0900)	(0.0412)	(0.0925)	(0.0100)		
SUG	0.2998	-0.1297	0.1864	-0.1306	-0.1214	-1.1492	
	(0.0802)	(0.0904)	(0.0652)	(0.0954)	(0.0986)	(0.0061)	
OTH	-0.6291	2.1118	-0.3191	2.2440	1.4774	2.4152	-4.7331
	(0.3464)	(0.4315)	(0.4010)	(0.7719)	(0.6393)	(0.6967)	(0.0417)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

Table A10 Parameter Estimates of the QUAIDS Model: Income-Groups Differences

	Eq.[1]	Low (End.Corrected)		Middle		High	
Variable		**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**	**Coeff.**	**Std. Err.**
Constant	α_1	0.1616	(0.0632)	0.1156	(0.0467)	0.1166	(0.0339)
	α_2	0.3455	(0.0571)	0.3004	(0.0409)	0.2922	(0.0377)
	α_3	0.2041	(0.0417)	0.2333	(0.0297)	0.2292	(0.0248)
	α_4	0.0427	(0.0278)	0.0552	(0.0296)	0.0267	(0.0221)
	α_5	0.0652	(0.0194)	0.0487	(0.0113)	0.0702	(0.0162)
	α_6	0.0206	(0.0200)	0.0172	(0.0114)	0.0144	(0.0080)
	α_7	0.1603	(0.0480)	0.2297	(0.0362)	0.2507	(0.0367)
Expenditure	β_1	-0.0527	(0.0377)	-0.1207	(0.0230)	-0.0609	(0.0180)
	β_2	0.0606	(0.0329)	0.0762	(0.0222)	-0.0278	(0.0193)
	β_3	-0.0465	(0.0254)	-0.0321	(0.0166)	0.0034	(0.0130)
	β_4	-0.0136	(0.0162)	0.0044	(0.0151)	-0.0398	(0.0117)
	β_5	-0.0173	(0.0118)	-0.0265	(0.0055)	-0.0152	(0.0081)
	β_6	-0.0288	(0.0127)	-0.0362	(0.0055)	-0.0130	(0.0038)
	β_7	0.0982	(0.0299)	0.1349	(0.0183)	0.1534	(0.0201)
Prices	γ_{11}	0.0197	(0.0272)	0.0164	(0.0215)	0.0209	(0.0162)
	γ_{21}	0.0074	(0.0189)	0.0167	(0.0155)	0.0146	(0.0133)
	γ_{31}	0.0070	(0.0130)	0.0054	(0.0107)	-0.0060	(0.0087)
	γ_{41}	0.0065	(0.0097)	-0.0077	(0.0103)	-0.0108	(0.0087)
	γ_{51}	0.0049	(0.0080)	-0.0016	(0.0052)	0.0117	(0.0073)
	γ_{61}	-0.0196	(0.0081)	-0.0120	(0.0051)	-0.0033	(0.0036)
	γ_{71}	-0.0259	(0.0181)	-0.0172	(0.0141)	-0.0271	(0.0129)
	γ_{22}	-0.0155	(0.0218)	-0.0170	(0.0185)	-0.0190	(0.0185)
	γ_{32}	-0.0079	(0.0120)	-0.0078	(0.0099)	-0.0103	(0.0095)
	γ_{42}	0.0006	(0.0084)	-0.0004	(0.0094)	0.0013	(0.0093)
	γ_{52}	-0.0064	(0.0073)	0.0027	(0.0048)	0.0045	(0.0073)
	γ_{62}	0.0081	(0.0074)	0.0037	(0.0048)	-0.0012	(0.0037)
	γ_{72}	0.0138	(0.0150)	0.0021	(0.0125)	0.0100	(0.0137)
	γ_{33}	-0.0021	(0.0123)	0.0165	(0.0102)	0.0163	(0.0093)
	γ_{43}	-0.0085	(0.0063)	-0.0062	(0.0070)	-0.0017	(0.0066)
	γ_{53}	0.0056	(0.0052)	0.0054	(0.0035)	0.0079	(0.0052)
	γ_{63}	-0.0082	(0.0053)	-0.0087	(0.0035)	-0.0030	(0.0027)
	γ_{73}	0.0141	(0.0115)	-0.0045	(0.0095)	-0.0032	(0.0100)
	γ_{44}	-0.0024	(0.0069)	0.0132	(0.0095)	0.0199	(0.0102)
	γ_{54}	0.0042	(0.0044)	-0.0003	(0.0035)	-0.0148	(0.0062)
	γ_{64}	0.0023	(0.0046)	0.0019	(0.0036)	-0.0006	(0.0035)
	γ_{74}	-0.0026	(0.0085)	-0.0005	(0.0091)	0.0068	(0.0095)
	γ_{55}	-0.0011	(0.0072)	-0.0136	(0.0049)	0.0003	(0.0082)
	γ_{65}	0.0015	(0.0051)	0.0017	(0.0034)	0.0016	(0.0035)
	γ_{75}	-0.0086	(0.0074)	0.0057	(0.0049)	-0.0114	(0.0073)
	γ_{66}	0.0054	(0.0072)	0.0009	(0.0047)	0.0013	(0.0037)
	γ_{76}	0.0105	(0.0075)	0.0125	(0.0050)	0.0051	(0.0041)
	γ_{77}	-0.0011	(0.0214)	0.0020	(0.0173)	0.0199	(0.0195)
Expenditure-squared	λ_1	-0.0090	(0.0081)	-0.0249	(0.0083)	0.0042	(0.0085)
	λ_2	0.0078	(0.0071)	0.0104	(0.0079)	-0.0238	(0.0091)
	λ_3	-0.0004	(0.0055)	0.0002	(0.0059)	0.0091	(0.0061)
	λ_4	-0.0049	(0.0034)	-0.0028	(0.0053)	-0.0216	(0.0056)
	λ_5	-0.0031	(0.0025)	-0.0032	(0.0019)	0.0015	(0.0038)
	λ_6	-0.0066	(0.0026)	-0.0057	(0.0019)	0.0036	(0.0018)
	λ_7	0.0163	(0.0066)	0.0261	(0.0066)	0.0270	(0.0097)

Table A10 (Cont'd)

Household size	z_{11}	0.0097	(0.0022)	0.0091	(0.0012)	0.0102	(0.0011)
	z_{21}	-0.0061	(0.0019)	-0.0042	(0.0011)	-0.0043	(0.0012)
	z_{31}	0.0007	(0.0015)	-0.0010	(0.0008)	0.0000	(0.0008)
	z_{41}	0.0014	(0.0009)	0.0002	(0.0007)	0.0003	(0.0007)
	z_{51}	0.0000	(0.0007)	0.0000	(0.0003)	0.0012	(0.0005)
	z_{61}	0.0009	(0.0007)	0.0013	(0.0003)	0.0014	(0.0002)
	z_{71}	-0.0065	(0.0016)	-0.0054	(0.0009)	-0.0089	(0.0012)
Race (1 if black)	z_{12}	0.0994	(0.0361)	0.0646	(0.0203)	0.0817	(0.0102)
	z_{22}	-0.0335	(0.0324)	0.0096	(0.0188)	0.0338	(0.0110)
	z_{32}	0.0253	(0.0246)	-0.0308	(0.0140)	-0.0289	(0.0074)
	z_{42}	-0.0342	(0.0159)	-0.0291	(0.0131)	-0.0323	(0.0065)
	z_{52}	-0.0232	(0.0111)	-0.0088	(0.0049)	-0.0089	(0.0047)
	z_{62}	0.0098	(0.0116)	0.0044	(0.0050)	0.0047	(0.0023)
	z_{72}	-0.0435	(0.0274)	-0.0100	(0.0158)	-0.0503	(0.0112)
Rural/Urban (1 if rural)	z_{13}	0.0242	(0.0176)	0.0596	(0.0112)	0.0402	(0.0090)
	z_{23}	-0.0159	(0.0157)	-0.0480	(0.0104)	-0.0425	(0.0095)
	z_{33}	0.0041	(0.0119)	0.0138	(0.0077)	0.0265	(0.0064)
	z_{43}	-0.0241	(0.0077)	-0.0093	(0.0071)	-0.0130	(0.0054)
	z_{53}	0.0054	(0.0054)	-0.0061	(0.0027)	0.0023	(0.0040)
	z_{63}	0.0095	(0.0056)	0.0031	(0.0027)	0.0035	(0.0019)
	z_{73}	-0.0031	(0.0132)	-0.0130	(0.0086)	-0.0170	(0.0095)
Education of head	z_{14}	-0.0005	(0.0015)	-0.0024	(0.0010)	-0.0015	(0.0007)
	z_{24}	0.0010	(0.0013)	0.0008	(0.0009)	-0.0006	(0.0008)
	z_{34}	0.0006	(0.0010)	-0.0001	(0.0007)	-0.0004	(0.0005)
	z_{44}	0.0008	(0.0006)	0.0010	(0.0006)	0.0011	(0.0004)
	z_{54}	-0.0007	(0.0005)	0.0003	(0.0002)	-0.0001	(0.0003)
	z_{64}	-0.0009	(0.0005)	-0.0001	(0.0002)	-0.0004	(0.0001)
	z_{74}	-0.0002	(0.0011)	0.0005	(0.0008)	0.0020	(0.0008)
Survey year (=1 if 1998)	z_{15}	-0.0226	(0.0349)	-0.0477	(0.0281)	-0.0291	(0.0232)
	z_{25}	0.0561	(0.0321)	0.0932	(0.0265)	0.0284	(0.0260)
	z_{35}	-0.0832	(0.0245)	-0.0580	(0.0194)	-0.0500	(0.0170)
	z_{45}	0.0224	(0.0174)	0.0277	(0.0212)	0.0140	(0.0186)
	z_{55}	-0.0103	(0.0118)	-0.0289	(0.0080)	-0.0408	(0.0126)
	z_{65}	-0.0210	(0.0122)	-0.0234	(0.0080)	-0.0097	(0.0065)
	z_{75}	0.0587	(0.0271)	0.0371	(0.0229)	0.0872	(0.0248)
Survey year (=1 if 2004)	z_{16}	-0.0670	(0.0472)	-0.0841	(0.0359)	-0.0772	(0.0310)
	z_{26}	0.0339	(0.0449)	0.0539	(0.0352)	0.0053	(0.0366)
	z_{36}	-0.0715	(0.0359)	-0.0226	(0.0262)	-0.0087	(0.0255)
	z_{46}	0.0121	(0.0246)	0.0440	(0.0275)	0.0548	(0.0270)
	z_{56}	0.0123	(0.0193)	-0.0096	(0.0122)	-0.0431	(0.0197)
	z_{66}	-0.0251	(0.0204)	-0.0383	(0.0119)	-0.0181	(0.0100)
	z_{76}	0.1054	(0.0380)	0.0567	(0.0300)	0.0870	(0.0329)
Month of survey	z_{17}	0.0036	(0.0045)	0.0030	(0.0036)	0.0018	(0.0029)
	z_{27}	-0.0027	(0.0041)	0.0023	(0.0033)	-0.0019	(0.0030)
	z_{37}	-0.0113	(0.0031)	-0.0068	(0.0024)	-0.0072	(0.0021)
	z_{47}	0.0060	(0.0020)	0.0034	(0.0023)	0.0045	(0.0017)
	z_{57}	-0.0009	(0.0014)	-0.0008	(0.0009)	-0.0005	(0.0013)
	z_{67}	-0.0011	(0.0014)	-0.0008	(0.0009)	-0.0001	(0.0006)
	z_{77}	0.0064	(0.0034)	-0.0003	(0.0027)	0.0032	(0.0031)

Table A10 (Cont'd)

Residuals	v_1	-0.0296	(0.0275)		
	v_2	-0.0137	(0.0238)		
	v_3	0.0403	(0.0182)		
	v_4	0.0145	(0.0119)		
	v_5	-0.0090	(0.0086)		
	v_6	-0.0125	(0.0091)		
	v_7	0.0101	(0.0204)		

[1]. The commodities represented by the different equation numbers are: 1 = grains; 2 = meat and fish; 3 = fruits and vegetables; 4 = dairy; 5 = oils, butter, and other fats; 6 = sugar; 7 = other foods

Table A11 Own and Cross Price Elasticities: Income-Group Differences

I. LOW-INCOME (corrected for expenditure endogeneity)

Marshallian/Uncompensated own-price elasticities1

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.8889						
	(0.1868)						
MF	0.9211	-2.4796					
	(0.8679)	(1.4364)					
FV	0.2240	-0.3664	-0.9468				
	(0.2429)	(0.3283)	(0.2343)				
DAI	0.0373	-0.0298	-0.0351	-0.9815			
	(0.0995)	(0.1521)	(0.0407)	(0.3596)			
OBF	0.0645	-0.1135	0.0490	0.2605	-0.9035		
	(0.0841)	(0.1289)	(0.0650)	(0.3146)	(0.1712)		
SUG	-0.1019	0.1108	-0.0401	-0.1536	-0.0860	-1.1357	
	(0.0842)	(0.1439)	(0.0669)	(0.3242)	(0.2126)	(0.4463)	
OTH	0.6457	-1.0430	0.1735	2.6821	1.5030	2.5366	-7.2895
	(0.6901)	(1.0044)	(0.8469)	(2.0815)	(1.3510)	(1.1148)	(3.6996)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.8111						
	(0.2406)						
MF	0.9988	-2.0963					
	(0.6338)	(0.1314)					
FV	0.3018	0.0169	-0.8119				
	(0.1503)	(0.2523)	(0.0937)				
DAI	0.1150	0.3535	0.0999	-0.9617			
	(0.2278)	(0.1878)	(0.0801)	(0.0268)			
OBF	0.1422	0.2698	0.1839	0.2803	-0.8781		
	(0.2019)	(0.1281)	(0.0447)	(0.3036)	(0.0200)		
SUG	-0.0242	0.4941	0.0948	-0.1338	-0.0606	-1.1337	
	(0.2931)	(0.2456)	(0.1559)	(0.3348)	(0.2213)	(0.0221)	
OTH	0.7234	-0.6597	0.3084	2.7019	1.5285	2.7980	-7.0282
	(0.4609)	(0.8818)	(0.7552)	(2.0583)	(1.3337)	(1.1491)	(0.0617)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

Table A11 (Cont'd)

II. MIDDLE-INCOME

Marshallian/Uncompensated own-price elasticities1

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-1.1924						
	(0.1987)						
MF	0.8583	-1.6434					
	(0.3035)	(0.4451)					
FV	0.3283	-0.2673	-0.8690				
	(0.1384)	(0.1412)	(0.1374)				
DAI	0.1018	-0.0900	-0.0255	-0.7310			
	(0.0998)	(0.0803)	(0.0649)	(0.1742)			
OBF	0.0009	0.0026	0.0398	-0.0171	-1.2419		
	(0.0324)	(0.0274)	(0.0194)	(0.0554)	(0.0966)		
SUG	-0.1135	0.0658	-0.0434	-0.0265	-0.0297	-1.0992	
	(0.0409)	(0.0497)	(0.0423)	(0.1104)	(0.0865)	(0.1241)	
OTH	0.8653	-0.6195	0.0072	0.4973	0.9365	1.3480	-5.1092
	(0.3258)	(0.4322)	(0.4018)	(1.0297)	(0.4669)	(0.3776)	(1.3480)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.8111						
	(0.2406)						
MF	0.9988	-2.0963					
	(0.6338)	(0.1314)					
FV	0.3018	0.0169	-0.8119				
	(0.1503)	(0.2523)	(0.0937)				
DAI	0.1150	0.3535	0.0999	-0.9617			
	(0.2278)	(0.1878)	(0.0801)	(0.0268)			
OBF	0.1422	0.2698	0.1839	0.2803	-0.8781		
	(0.2019)	(0.1281)	(0.0447)	(0.3036)	(0.0200)		
SUG	-0.0242	0.4941	0.0948	-0.1338	-0.0606	-1.1337	
	(0.2931)	(0.2456)	(0.1559)	(0.3348)	(0.2213)	(0.0221)	
OTH	0.7234	-0.6597	0.3084	2.7019	1.5285	2.7980	-7.0282
	(0.4609)	(0.8818)	(0.7552)	(2.0583)	(1.3337)	(1.1491)	(0.0617)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

Table A11 (Cont'd)

II. HIGH-INCOME

Marshallian/Uncompensated own-price elasticities1

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.8770						
	(0.0668)						
MF	0.0429	-0.5898					
	(0.1826)	(0.1550)					
FV	-0.0569	0.4815	-1.2072				
	(0.1945)	(0.2525)	(0.2764)				
DAI	-0.0092	-0.1292	0.0725	-1.1427			
	(0.0595)	(0.0980)	(0.0833)	(0.3081)			
OBF	0.0501	0.0765	0.0149	0.0040	-0.9912		
	(0.0376)	(0.0757)	(0.0572)	(0.1806)	(0.1720)		
SUG	-0.0029	-0.0339	0.0004	-0.0901	0.0463	-0.9203	
	(0.0179)	(0.0356)	(0.0278)	(0.0838)	(0.0719)	(0.1075)	
OTH	-0.2797	1.4025	-0.8377	4.1039	-0.5546	-0.7364	-3.9698
	(0.4866)	(0.5807)	(0.5542)	(1.1784)	(1.1056)	(1.0043)	(1.2688)

Hicksian/uncompensated own-price elasticities

	GR	MF	FV	DAI	OBF	SUG	OTH
GR	-0.6515						
	(0.0671)						
MF	0.2685	-0.4964					
	(0.1212)	(0.0724)					
FV	0.1686	0.5749	-1.0054				
	(0.1296)	(0.1895)	(0.0351)				
DAI	0.2163	-0.0358	0.2742	-1.1407			
	(0.1196)	(0.1366)	(0.1080)	(0.0215)			
OBF	0.2756	0.1699	0.2167	0.0060	-0.9524		
	(0.0538)	(0.0837)	(0.0519)	(0.1746)	(0.0123)		
SUG	0.2226	0.0595	0.2022	-0.0881	0.0851	-0.8952	
	(0.0785)	(0.0888)	(0.0527)	(0.0905)	(0.0768)	(0.0051)	
OTH	-0.0542	1.4959	-0.6360	4.1060	-0.5159	-0.3545	-3.5879
	(0.4205)	(0.5115)	(0.5204)	(1.1586)	(1.0938)	(1.0100)	(0.0502)

[1]. GR = grains; MF = meat and fish; FV = fruits and vegetables; DAI = dairy; OBF = oils, butter, and other fats; SUG = sugar; OTH = other foods

CHAPTER SIX

SUMMARY AND CONCLUSIONS

6.1 Summary

This study analyzed food expenditure patterns in South Africa, taking into account differences in preferences across rural and urban households, as well as across income groups. The study was motivated by the need to provide an accurate analysis of food demand in South Africa, and to provide demand behavioral parameters that can be used for welfare analyses. Analyses of the panel data indicated that there is a need to allow for more curvature in the Engel curve relationships than is permitted by the Working-Leser form, which allows for only a linear relationship between budget shares and expenditure. The implication of this finding is that popular functional forms such as AIDS do not provide an accurate picture of demand behavior in South Africa. Given this finding, this study estimated the QUAIDS model developed by Banks *et al.* (1997), which is a generalization of the AIDS model that allows for a quadratic relationship between budget shares and expenditure. The QUAIDS model was used to estimate demand functions for seven food groups— grains, meat and fish, fruits and vegetables, dairy, oils and fats, sugar, and all other foods. The demand functions were estimated using panel data on household food consumption collected as part of the KwaZulu-Natal Income Dynamics Study (KIDS), with surveys carried out in 1993, 1998, and 2004.

The QUAIDS model was estimated in its flexible form, without imposing linearity on the price aggregators, as is commonly done in empirical estimation of AIDS. Expenditure endogeneity was explicitly tested, and corrected for using the control function approach where necessary. Expenditure censoring, which was found to be a problem in the dairy commodity, was corrected for using a two-step procedure suggested by Shonkwiler and Yen (1999).

Because the existing food demand studies in South Africa are based on the LA/AIDS model, the study also estimated the food demand system using AIDS and LA/AIDS. The expenditure and price elasticity estimates from these models were compared. It was found that the exclusion of the quadratic expenditure term has a larger effect on the magnitudes of expenditure elasticities than price elasticities. On average, the QUAIDS and AIDS expenditure elasticity estimates differed by about 53%, with most of the AIDS estimates being larger. However, these differences were much larger in the case of commodities that were found to require a quadratic expenditure term. The AIDS expenditure elasticity estimates for grains, sugar, and dairy, which all required an a quadratic expenditure term, were 113%, 101%, and 80%, respectively, larger than the QUAIDS estimates. The standard errors of both the QUAIDS and AIDS expenditure elasticity estimates were all significant at the 1% level, so that the decision to choose between QUAIDS and AIDS models could not be made based on the statistical significance of the estimated expenditure elasticities. The estimates of price and expenditure elasticities from the AIDS and LA/AIDS models were found to be very similar, and so were their standard errors.

In addition to comparing the elasticity estimates, the study also examined the bias in welfare measures that can result if these are based on the restrictive AIDS and LA/AIDS models. This was accomplished by measuring the welfare gains from the zero-rating of the value-added tax (VAT) on meat products. Two common measures of welfare were computed, namely, the compensating variation and equivalent variation. Similar to the findings in the case of expenditure elasticities, it was found that AIDS and the LA/AIDS models overstate the welfare gains from zero-rating meat. Consistent with both AIDS and LA/AIDS models not allowing for adequate curvature in the Engel curves, their overstating of welfare gains was found to be larger for richer households (that is, the bias in welfare gains increases with total household expenditure).

Apart from the empirical contributions explained above, this study also makes methodological contribution. The study build on the work of Bank *et al.* (1997) to develop an LM test that can be used to determine whether or not a quadratic (QUAIDS) or a linear (AIDS) demand model specification is appropriate. In particular, the implication of corollary 2 in Banks *et al.* (p.533) is that a utility-derived demand system that is rank 3 and exactly aggregable cannot have coefficients on both the linear and the quadratic expenditure terms that are independent of prices. In other words, if a rank 3 exactly aggregable demand system that is derived from utility theory has a coefficient on the linear expenditure term that is independent of prices, then it *must* have a coefficient on the quadratic expenditure term that is price dependent. The study noted that since the QUAIDS model, which is rank 3 and exactly aggregable, has a coefficient on the linear expenditure term that is independent of prices, then whether or not it is the appropriate functional form can be tested for based on the statistical significance of prices in the coefficient on the quadratic expenditure term. The test was developed and implemented in this study, and provided statistical evidence in support of the QUAIDS functional form. The overall conclusion from this test was consistent with that based on the likelihood ratio test for significance of lambda in the QUAIDS model, although the results of the two tests sometimes differed in the individual budget share equations. These differences in conclusions in the individual budget share equations (regarding which equations are linear and which are not) was not necessarily unexpected, given the differences in approach followed in constructing the two tests.

Given the highly unequal distribution of income in South Africa, and the large divide in wealth distribution between urban and rural areas, it is expected that preferences will differ significantly among households in different socioeconomic and demographic groups. Hence, pooling data across all households is likely to obscure important information on variability in demand behavior. To determine the impact of this household heterogeneity on demand, this study analyzed the demand behavior of rural and urban households separately, as well as households in different income groups.

For some food groups, the difference in the estimated expenditure elasticities between rural and urban samples was found to be substantial. For instance, a 1% increase in total food expenditure for urban households leads to only about 0.02% increase in the budget share of meat and fish, but to about 1.73% increase in the case of rural households. High income households respond to a 1% increase in the price of meat and fish by decreasing consumption by 0.59%, while low income households decrease consumption by about 2.48%. These results provide further evidence on the need to undertake disaggregated analysis of demand behavior in a country with highly diverse socioeconomic and demographic household groups like South Africa.

6.2 Conclusions

This study is the first to apply the QUAIDS model to analyze food consumption patterns in South Africa. The study implemented recent advances in econometric methods designed to enhance demand estimation with micro-level data. Hence, the price and expenditure elasticities estimated in this study are an improvement of the previous estimates for South Africa, the majority of which are commodity-specific and based on single equation estimation. Furthermore, most of these previous studies use aggregate data, and therefore, do not allow for the impact of demographic effects on consumption. The exception is a study Agbola (2003), that uses a 1993 cross-sectional dataset to estimate demand functions for a broad range of food groups similar to the one estimated in this study. However, Agbola uses the restrictive LA/AIDS model. As shown in chapter five, the magnitudes of demand elasticities can vary widely depending on the functional form used. For instance, the AIDS and LA/AIDS expenditure elasticities tend to be larger than the QUAIDS expenditure elasticities; the standard errors of these elasticity estimates were both significant at the 1% level. This dependence of demand behavioral parameters on model selection is particularly important because of its implication for policy advice. We showed, through the example of a tax reform on meat products, that welfare measures such as equivalent variation can differ significantly between QUAIDS and AIDS. The expenditure and price elasticities estimated in this study correct for many of the problems associated with these previous estimates, and therefore, can be used more reliably in calculating welfare measures.

The finding of substantial consumption differences between rural and urban households, as well as among households in different income-groups has important welfare implications. These results indicate that the design of anti-poverty and nutrient enhancement policies needs to be region-specific and should be based on accurate and comprehensive food poverty studies. No systematic differences in the absolute magnitudes of the expenditure elasticities and own-price elasticities were found (for instance, it could not be said that households are more responsive to expenditure changes than price changes); this means that a combination of income and price policies may be more effective in influencing consumption patterns than those based solely on one and not the other. The usual challenge remains one of finding optimal policy designs that make food more affordable to vulnerable groups in the long run, while not adversely affecting the development of a food marketing system that stimulates production incentives and income growth. One of the keys to the design of such policies is that they be based on thorough and robust statistical analyses.

As with all empirical models, there are some limitations to this study that should form the basis for future research. The estimation of the QUAIDS model in this study was based on pooled maximum likelihood estimation. One possibility is to extend the QUAIDS model in ways that address the problem of unobserved heterogeneity taking advantage of the panel structure of this dataset, such as random effects or fixed effects.

REFERENCES

Aasnes, J., E. Biorn, and T. Skjerpen. "Engel Functions, Panel Data, and Latent Variables." *Econometrica* 61(6)(1993):1395-1422

Abdulai, A. "Household Demand for Food in Switzerland. A quadratic Almost Ideal Demand System." *Swiss Journal of Economics and Statistics* 138(1) (2002):1-18

Abdulai, A. "Economies of Scale and Demand for Food in Switzerland: Parametric and Non-Parametric Analysis." *Journal of Agricultural Economics* 54(2)(July 2003):247-267.

Abdulai, A. and D. Aubert "A cross-section analysis of household demand for food and nutrients in Tanzania." *Agricultural Economics* 31(2004):67-79

Adato, M., M.R. Carter, and J. May. "Sense in Sociability? Social Exclusion and Persistent Poverty in South Africa." Staff paper 477, University of Wisconsin, Department of Agricultural and Applied Economics, Madison, WI, September 2004.

Agbola, F.W. "Estimation of Food Demand Patterns in South Africa Based on a Survey of Households." *Journal of Agricultural and Applied Economics* 35(3)(December 2003):662-670

Aguero, J.M. and B.W. Gould. "Household Composition and Brazilian Food Purchases: An Expenditure System Approach." *Canadian Journal of Agricultural Economics* 51(2003):323-345.

Anderson, G. and R.W. Blundell. "Estimation and Hypothesis Testing in Dynamic Singular Equation Systems." *Econometrica* 50(6)(1982):1559-1579.

Amemiya, T. "Multivariate Regression and Simultaneous Equation Models When the Dependent Variable Are Truncated Normal." *Econometrica* 42(November 1974):999-1012.

Amemiya, T. *Advanced Econometrics*. Cambridge: Harvard University Press, 1985.

Arndt, C. "Demand for Herbicide Use in Corn: An Entropy Approach Using Micro-Level Data." *Journal of Agricultural and Resource Economics* 24(1)(1999):2004-221.

Arndt, C., S. Liu, and P.V. Preckel. "On Dual Approaches to Demand Systems Estimation in the Presence of Binding Quantity Constraints." *Applied Economics* 31(1999):999-1008.

Banks, J.; Blundell, R., and A. Lewbel. "Tax Reform and Welfare Measurement: Do We Need Demand System Estimates?" *The Economic Journal* 106(September 1996):1227-1241.

Banks, J.; Blundell, R., and A. Lewbel. "Quadratic Engel Curves and Consumer Demand." *The Review of Economics and Statistics* 79(4)(November 1997):527-539.

Barten, A.P. "Family Composition, Price and Expenditure Patterns." In *Econometric Analysis for National Planning*, Hart, P.E., G. Mills, and J.K. Whittaker eds., Butterworth: London, 1964.

Barten, A.P. "Maximum Likelihood Estimation of Complete System of Demand Equations." *European Economic Review* 1(1969):7-73.

Becketti, S., W. Gould, L. Lillard, and F. Welch. "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation." *Journal of Labor Economics* 6(4)(October 1988):472-492.

Belete, A., C.O. Igodan, C. M'Marete, and W. van Averbeke. "Analysis of rural household consumption expenditure in South Africa: the case of food plot holders in the Tyefu irrigation scheme in the Eastern Cape Province." *Agrekon* 38(2):194-201

Blundell, R.W., A. Duncan, and K. Pendakur. "Semiparametric Estimation and Consumer Demand." *Journal of Applied Econometrics* 13(5)(September-October 1998):435-461.

Blundell, R.W. and C. Meghir. "Bivariate Alternatives to the Tobit Model." *Journal of Econometrics* 34(1987):179-200.

Blundell, R.W., P. Pashardes, and G. Weber. "What Do We Learn About Consumer Demand Patterns from Micro Data." *American Economic Review* 83(3) (June1993):570-597.

Blundell, R.W. and J.L. Powell. "Endogeneity in Semiparametric Binary Response Models." *Review of Economic Studies* 71(2004):655-679.

Blundell, R.W. and J.M. Robin. "Estimation in Large and Disaggregated Demand Systems: An Estimator for Conditionally Linear Systems." *Journal of Applied Econometrics* 14(1999):209-232.

Blundell, R.W. and J.M. Robin. "Latent Separability: Grouping Goods without Weak Separability." *Econometrica* 68(1)(2000):53-84.

Bollino, C.A., F. Perali, and N. Rossi. "Linear Household Technologies." *Journal of Applied Econometrics* 15(3)(May-June 2000): 275-87.

Bouis, H. "The Effect of Income on Demand for Food in Poor Countries: Are our Data Bases giving us Reliable Estimates?" *Journal of Development Economics* 44(1994):199-226.

Bouis, H., and L. Haddad. "Are Our Estimates of calorie-income elasticities too high? A recalibration of the plausible range." *Journal of Development Economics* 39(2)(1992):333-364.

Bowmaker, P.A., and W.L Nieuwoudt. "Demand equations for selected South African agricultural commodities." *Agrekon* 29(1)(1990):46–50.

Browning, R., and J.M. Collado. "The Response of Expenditures to Anticipated Income Changes: Panel Data Estimates." *American Economic Review* 91(3)(June 2001):681-692.

Browning and Collado. "Habits and Heterogenity in Demands: A Panel Data Analysis." Centre for Applied Microeconometrics, Institute of Economics CAM working Paper No. 2004-18, University of Copenhagen, Copenhagen, 2004.

Brownig, M.J., and C. Meghir. "The Effects of Male and Female Labour Supply on Commodity Demands." *Econometrica* 59 (1991):925-951.

Buse, A. "Evaluating the Linearized Almost Ideal Demand System." *American Journal of Agricultural Economics* 76(4)(November 1994):781-793.

Carpentier A. and H Guyomard. Unconditional Elasticities in Two-Stage Demand Systems: An Approximate Solution." *American Journal of Agricultural Economics* 83(1)(February 2001):222-229

Carter, M.R., and J. May. "Poverty, Livelihood and Class in Rural South Africa." *World Development* 27(1)(January 1999):1-20.

Carter, M.R., and J. May. "One Kind of Freedom: Poverty Dynamics in Post-Apartheid South Africa." *World Development* 29(12)(December 2001):1987-2006.

Christensen, M.L. "Heterogeneity in Consumer Demands and the Income Effect." Institute of Economics, University of Copenhagen, mimeo (2004).

Christensen, L.R., D.W. Jorgenson, and L.J. Lau. "Transcendental Logarithmic Utility Functions." *American Economic Review* 65 3(1975)(June): 367-383.

Chung, C.F. "Modelling Demand Systems with Demographic Effects Based on the Modifying Function Approach." *Economics Letters* (73)(2001):269-274.

Cox, T.L., and M.K. Wohlgenant. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *American Journal of Agricultural Economics* 68(4)(November 1986):908-19.

Craig, J.G. "Some Statistical Models for Limited Dependent Variables with Applications to the Demand for Durable Goods." *Econometrica* 39(1971):829-844.

Davis, G.C. "The Generalized Composite Commodity Theorem: Stronger Support in the Presence of Data Limitations." *The Review of Economics and Statistics* 85(2):476-480.

Deaton, A. "Quality, Quantity, And Spatial Variation of Price." *American Economic Review* 78(3)(June 1988):418-430.

Deaton, A. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore, MD: John Hopkins University Press, 1997.

Deaton, A., and M. Irish. "A Statistical Model for Zero Expenditures in Household Budgets." *Journal of Public Economics* 23(1984):312-326.

Deaton, A., and J. Muellbauer. *Economics and Consumer Behavior*. Cambridge University Press, 1980a.

Deaton, A., and J. Muellbauer. "An Almost Ideal Demand System." *American Economic Review* 70(3)(June 1980b):312-336.

Dhar, T., J.P. Chavas, and B.W. Gould. "An Empirical Assessment of Endogeneity Issues in Demand Analysis for Differentiated Products." *American Journal of Agricultural Economics* 85(3)(August 2003):605-617.

Döckel, J.A., and J.A. Groenewald. "The demand for food in South Africa." *Agrekon* 9(4)(1970):15-20.

Dong, D., Shonkwiler, J.S., and O. Capps, Jr. "Estimation of Demand Functions Using Cross-Sectional Household Data: The Problem Revisited." *American Journal of Agricultural Economics* 80(3)(August 1998):466-473.

Dynan, K.E. "Habit Formation in Consumer Preferences: Evidence from Panel Data." *American Economic Review* 90(3)(2000):391-406.

Eales, J.S. and L.J. Unnevehr. "Demand for Beef and Chicken Products: Separability and Structural Change." *American Journal of Agricultural Economics* 70(1988):521-532.

Edgerton, D.L. "Weak Separability and the Estimation of Elasticities in Multistage Demand Systems." *American Journal of Agricultural Economics* 79(1997):62-79.

Elliot, M.B., and J. van Zyl. "The effect of income on the market for carbohydrates in South Africa." *Agrekon* 30(4)(1991):269-71.

Fitsgerald, J., P. Gottschalk, and R. Moffitt. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources* 33(2)(Spring 1998):251-299.

Food Pricing Monitoring Committee. Report of the Food Pricing Monitoring Committee, National Department of Agriculture, Pretoria, South Africa, 2003.

Gorman, W.M. "Tricks With Utility Functions." In *Essays in Economic Analysis*, M. Artis and R. Nobay eds. Cambridge: Cambridge University Press, 1976.

Gorman, W.M. "Some Engel Curves." In *The Theory and Measurement of Consumer Behaviour*, A. Deaton ed. Cambridge, UK: Cambridge University Press, 1981.

Gould, W., J. Pitblado, and W. Sribney. *Maximum Likelihood Estimation with Stata.* College Station, Texas: Stata Corporation.

Gould, W. and H.J. Villarreal. "An Assessment of the Current Structure of Food Demand in Urban China." *Agricultural Economics* 34(2006):1-16.

Hancock, P.J., W. L Nieuwoudt, and M.C. Lyne. "Demand analysis of meats in South Africa." *Agrekon* 23(3)(1984):26–29.

Heckman, J.J. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variable Models Using Simulated ML and GMM Estimation Methods." *Annals of Economic and Social Measurement* 5(1976):475-492.

Heien, D., and C.R. Wessells. "Demand System Estimation with Microdata: A Censored Regression Approach." *Journal of Business and Statistics* 8(3)(July 1990):365-371.

Hausman, J.A. "Specification Tests in Econometrics." *Econometrica* 46(1978):1251-1271

Hausman, J.A., and W. E. Taylor. "Panel Data and Unobservable individual effects." *Econometrica* 31(1986):93-118.

Human Science Research Council (January 2004). *Food Security in South Africa: Key Policies Issues for the Medium Term* (Online). Available from: www.sarpn.org.za/documents/d0000685/Food_security_SA_January2004.pdf

Integrated Food Security Strategy (IFSS). National Department of Agriculture Policy Document. Pretoria, South Africa, 2002.

Jooste, A. "Beef trade preferences and intensities in the Southern African Customs Union." *Agrekon* 35(3)(1996):123–130.

Jorgenson, D.W., L.J. Lau, and T.M. Stoker. "The Transcendental Logarithmic Model of Aggregate Consumer Behavior." In R. Basman and G. Rhodes eds., *Advances in Econometrics*, vol.1, Greenwich, Ct: JAI Press, 1982.

Koch, S.F. "South African Households: Are Expenditure Patterns Almost Ideal?" Paper Presented at the 9th Annual Conference on Econometric Modelling for Africa, June 30 – July 02, 2004, Cape Town, South Africa.

Labeaga, J.M., and A. López. "A Study of Petro Consumption Using Spanish Panel Data." *Applied Economics* 29(6)(1997):795-802.

LaFrance, J.T. "When is Expenditure 'Exogenous' in Separable Demand Models?" *Western Journal of Agricultural Economics* 16(1991):49-62

Laurent, C. and E. Comon. "Behavioral Heterogeneity and the Income Effect." *The Review of Economics and Statistics* 85(3)(2003):653-669.

Lecocq, S. and J. Robin. "Estimating Demand Response with Panel Data." Empirical *Economics* (2006). Forthcoming.

Lee, L., and M. M. Pitt. "Microeconometric Demand System with Binding Nonnegativity Constraints: The Dual Approach." *Econometrica* 54(5)(September 1986):1237-1242.

Leibbrandt, M., and I. Woolard. "A comparison of poverty in South Africa's nine provinces." *Development Southern Africa* 16(1)(March 1999):37-35.

Leibbrandt, M., G. Fields, and P. Cichello. "Earnings and Employment Dynamics for Africans in Post-apartheid South Africa: A Panel Study of KwaZulu-Natal." *Journal of African Economies* (Forthcoming).

Lesser, C.E.V. "Forms of Engel Functions." *Econometrica* 31(4)(October 1963): 694-703

Lewbel, A. "A Unified Approach to Incorporating Demographic or Other Effects into Demand Systems." *Review of Economic Studies* 52(1)(January 1985):1-18.

Lewbel, A. "The Rank of Demand Systems: Theory and Nonparametric Estimation." *Econometrica* 59(1991):711-730.

Lewbel, A. "Aggregation Without Separability: A Generalized Composite Commodity Theorem." *American Economic Review* 86(3):524-542.

Lewbel, A. "Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D." *Econometrica* 65 (1997): 93-118.

Liu, K.E. "Food Demand in Urban China: An Empirical Analysis Using Micro Household Data." Unpublished Ph.D. Dissertation, Ohio State University, Columbus, Ohio, 2003.

Loubser, M. (1990). "Income Elasticities of the demand for consumer goods and services." Bureau of Market Research No. 175, University of South Africa, Pretoria.

Lubbe, W.F. "Decomposition of price time series components of the red meat industry for efficient policy and marketing strategies." *Agrekon* 29(4)(1990): 296 - 304.

Maluccio, J. "Attrition in the KwaZulu Natal Income Dynamics Study, 1993-1998." IFPRI FCND Discussion Paper No. 95, October 2000.

May, J., and I. Woolard. "Measurement and experience". In J. May (Ed.), *Poverty and inequality in South Africa: report prepared for the office of the executive deputy president and the inter-ministerial committee on poverty and inequality.* Durban, Praxis Publishing, 1998.

Meenakshi, J.V., and R. Ray. "Regional Differences in India's Food and Expenditure Pattern: A Complete Demand Systems Approach." *Journal of International Development* 11(Jan/Feb 1999): 47-74.

Meyerhoefer, C.D. "The Econometrics of Non-Standard Structural Demand Modeling: Applications to Transition Country Data." Unpunblished Ph.D. Dissertation, Cornell University, Ithaca, New York, 2002.

Meyerhoefer, C.D., C.K. Ranney, and D.E. Sahn. "Consistent Estimation of Censored Demand Systems using Panel Data." *American Journal of Agricultural Economics* 87(3)(August 2005):660-672.

Moffit, R., J. Fitzgerald, and P. Gottschalk. "Sample Attrition in Panel Data: The Role of Selection on Observables." Annale d'Economie et de Statistique 55-56(1999)

Molina, J.S. and A. I. Gil. "The Demand Behavior of Consumers in Peru: A Demographic Analysis using the QUAIDS." *Journal of Developing Areas* 39(1)(Spring 2005):191-206.

Moschini, G. "A Non-nested Test of Separability for Flexible Functional Forms." *The Review of Economics and Statistics* 74(2)(1992):365-369.

Moschini, G. "Units of Measurement and the Stone Index in Demand System Estimation." *American Journal of Agricultural Economics* 77(February 1995):63-68.

Moschini, G., D. Moro, and R.D. Green. "Maintaining and Testing Separability in Demand Systems." *American Journal of Agricultural Economics* 76(February 1994):61-73.

Moro, D. and P. Sckokai. "Heterogenous Preferences in Household Food Consumption in Italy. *European Review of Agricultural Economics* 27(3)(2000):305-323.

Muellbauer, J. "Community Preferences and the Representative Consumer." *Econometrica* 44(1976): 525-543.

Muellbauer, J. "Testing the Barten Model of Household Composition Effects and the Cost of Children." *Economic Journal* 87(347)(September 1977):800-808.

Nayga, Jr., R. M. and O. Capps, Jr. "Tests of Weak Separability in Disaggregated Meat Products." *American Journal of Agricultural Economics.* 76(4) (November 1994):800-808.

Nelson, J.A. "Quality Variation and Quantity Aggregation in Consumer Demand for Food." *American Journal of Agricultural Economics* 73(4)(November 1991):1204-12.

Nieuwoudt, W.L. "The demand for livestock products in South Africa for 2000, 2010 and 2020: Part 1." *Agrekon* 37(2)(1998).

Orcutt, G.H., H.W. Watts and J.B. Edwards (1968). "Data Aggregation and Information Loss." *American Economic Review* 58(4)(1968):773-787.

Park, J. and G.C. Davis. "The Theory and Econometrics of Health Information in Cross-Sectional Nutrient Demand Analysis." *American Journal of Agricultural Economics* 83(4)(November 2001): 840-851.

Piggot, N.E. "The Nested PIGLOG Model: An Application to U.S. Food Demand." *American Journal of Agricultural Economics* 85(1)(February 2003):1-15.

Pitt, M.M.; M.R. Rosenzweig, and Md. N. Hassan. "Productivity, Health, and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries." *American Economic Review* 80(5)(December 1990):1139-1156.

Poi, B. P. "From the Help Desk: Demand System Estimation." *Stata Journal* 2(4)(2002):403-410.

Pollak, R. A., and T.J. Wales. "Estimation of the Linear Expenditure System." *Econometrica* 37(4)(October 1969): 611-627.

Pollak, R.A., and T.J. Wales. "Estimation of Complete Demand Systems from Household Budget Data: The Linear and Quadratic Expenditure Systems." *American Economic Review* 68(3)(June 1978):348-359.

Pollak, R.A. and T.J. Wales. "Comparison of the Quadratic Expenditure System and Translog Demand Systems with Alternative Specifications of Demographic Effects." *Econometrica* 48(3)(April 1980):595-612.

Pollak, R.A. and T.J. Wales. "Demographic Variables in Demand Analysis." *Econometrica* 49(6)(November 1981):1533-1551.

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

Poonyth, D., R. Hassan, and J.F. Kirsten. "Random Coefficients Analysis of Changes in Meat Consumption Preferences in South Africa." *Agrekon* 40 (2001):426-437

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

Ravallion, M. "Income Effects of Malnutrition." *Economic Development and Cultural Change* 38(3):489-515.

Reed, A.J., J.W. Levendahl, and C. Hallahan. "The Generalized Composite Commodity Theorem and Food Demand Estimation." *American Journal of Agricultural Economics* 87(1):28-37.

Shonkwiler, J.S., and S.T. Yen. "Two-Step Estimation of a Censored System of Equations." *American Journal of Agricultural Economics* 81(4)(November 1999):972-982.

Skoufias, E. "Is the Calorie-Income Elasticity Sensitive to Price Changes? Evidence from Indonesia." *World Development* 31(7)(July 2003):1291-1307.

Statistics South Africa (SSA). *Bulletin of Statistics* 33(1-2)(1999). Pretoria, South Africa.

Statistics South Africa (Stats SA) (2000). Website: www.statssa.gov.za

Stone, R. "Linear Expenditure Systems and Demand Analysis: An application to the Pattern of British Demand." *The Economic Journal* 64(225)(September 1954):511-27.

Subramanian, S., and A. Deaton. "The Demand for Food and Calories." *Journal of Political Economy* 104(1)(1996):133-162.

Taljaard, P.R., Z.G. Alemu, and H.D. van Schalkwyk. "The Demand for Meat in South Africa: An Almost Ideal Estimation." *Agrekon* 43(4)(December 2004):430-443.

Theil, H. "The Information Approach to Demand Analysis." *Econometrica* (33)(1965):67-87.

Thomas, D., E. Frakenberg, and J.P. Smith. "Lost but not Forgotten: Attrition and Follow-up in the Indonesia Family Health Survey." *Journal of Human Resources* 36(4)(Autumn 2001):556-592.

Tobin, J. "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26(1)(January 1958):24-36.

Trade and Industrial Policy Strategies (TIPS). Website: www.tips.org.za

Varian, H.R. *Microeconomic Analysis*. New York, N.Y.: W.W. Norton & Company, Inc., 1992.

Wales, T.J., and A.D. Woodland. "Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints." *Journal of Econometrics* 21(3)(1983):263-285.

Watkinson, E. and N. Makgetla. "COSATU looks at the Food Security Crisis." *South African Labour Bulletin* 6(2)(April 2002).

Woolard, I., S. Klasen, and M. Leibbrandt. "Income mobility and household dynamics in South Africa: The case of KwaZulu-Natal." *Labour Markets and Social Frontiers* 2(2002):5-11.

Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*. Cincinnati, OH: South-Western, 2000.

Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

World Bank. GNI Per Capita 2004, Atlas Method and PPP. At: www.worldbank.org

World Development Report (2002), The World Bank.

World Health Organizations (1995). Website: www.who.int

van Zyl, J., and J. Kirsten. "Food Security in South Africa." *Agrekon* 31(4)(December 1992):170-183.

Yen, S.T., K. Kan, and S. Su. "Household demand for fats and oils: two-step estimation of censored demand system." *Applied Economics* 34(14)(September 2002):1799-1806.