CHARLES UNIVERSITY IN PRAGUE
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Master Thesis

2016 Šarlota Smutná

CHARLES UNIVERSITY IN PRAGUE
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Šarlota Smutná

Household food demand in the Czech Republic: coherent demand system dealing with selectivity

MASTER THESIS

Prague 2016

Author: Bc. Šarlota Smutná, MSc.
Supervisor: Mgr. Milan Ščasný, Ph.D.

Academic year: 2015/2016

I would like to thank Dr. Milan Ščasný for supervising this thesis, for careful guidance, for patience, and helpful comments. I would like to also thank the Environment Center of Charles University in Prague for providing a data.

I hereby declare, that I have written this thesis using only literature and other sources listed in bibliography. Furthermore, I declare that I have not used this thesis to acquire another academic degree. Lastly, I declare that individual entries from the household budget survey have been analysed in accordance with the Agreement on the data usage signed between the Environment Center of Charles University in Prague and the Czech Statistical Office. I acknowledge and agree with lending and publishing of the thesis.

In Prague, 12 May 2016 Šarlota Smutná

Abstract

Demand for food is widely studied topic in applied econometrics. Demand systems are the most useful models to evaluate demand and estimate the income and price elasticities. Different demand systems used for food demand are discussed in this thesis. Almost Ideal Demand System (AIDS) is the most popular among researchers thanks to conformity with economic theory, simple estimation, and flexibility with respect to non-linearity of Engel's curves or to control for socio-demographic or structural variables of household. Reporting of zero consumption by respondents when analysing demand on budget survey data requires special treatment, as censoring leads to the selectivity problem and hence biased estimates. Several techniques to treat the selectivity in order to obtain unbiased estimate of demand elasticities are discussed. Specifically, the Heien and Wessels, Shonkwiler and Yen, and Cosslett's semi-parametric corrections are incorporated into the AIDS model and empirically compared among each other. Since homogeneity and symmetry conditions are not fulfilled in this case, income and price elasticities of food demand are estimated by the unrestricted version of QUAIDS model which suits the budget survey data of Czech households the best with the correction for the selectivity by Shonkwiler and Yen's estimator. The estimated income elasticities of demand for all food and non-alcoholic beverages categories are higher than unity. The own-price elasticities are for all categories negative and vary across them.

Keywords food, demand, AIDS, selectivity

Abstrakt

Poptávka po potravinách je v aplikované ekonometrii často zkoumaným tématem. Analyzována bývá především prostřednictvím koherentních poptávkových systémů, v rámci kterých se odhadují i cenové a důchodové elasticity. Tato práce popisuje nejčastěji využívané poptávkové systémy pro odhad poptávkových elasticit potravin. AIDS („Almost Ideal Demand System“) je jedním z nejpopulárnějších systémů díky jeho souladu s ekonomickou teorií a díky jeho flexibilitě zahrnout socio-demografické proměnné nebo pracovat s daty, která nesplňují předpoklad lineárních Engelových křivek. Speciální pozornost vyžadují nulové výdaje za spotřebu, které se často objevují v datech ze statistiky rodinných účtů. Takto cenzorovaná data jsou často spojena s problémem selektivity, jehož neřešení vede ke zkresleným odhadům. Diskutováno je tedy několik způsobů, jak problém selektivity v poptávkových modelech řešit. Konkrétně byl poptávkový systém AIDS rozšířen o korekci selektivity vycházející ze studií Heien a Wesselse, Shonkwilera a Yena a z Cosslettovy semiparametrické metody. Tyto různé druhy korekce jsou mezi sebou empiricky porovnány. Systém pro poptávku po potravinách využívající individuální data ze statistiky českých rodinných účtů nesplňuje podmínku homogenity a symetrie, ani linearitu Engelových křivek. Výsledné cenové a důchodové elasticity poptávky po potravinách jsou proto odhadnuty pomocí modelu QUAIDS s korekcí selektivity dle metody Shonkwilera a Yena, u kterého mohou být teoretické restrikce modelu porušeny. Důchodové elasticity jsou větší než 1 pro všechny komodity potravin a nealkoholických nápojů popsané QUAIDS modelem. Vlastní cenové elasticity jsou negativní pro všechny komodity systému, přičemž je jejich hodnota proměnlivá.

Klíčová slova potraviny, poptávka, AIDS, selektivita

Contents

List of Tables

List of Figures

Master Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author:	**Bc. Šarlota Smutná**	Supervisor:	Mgr. Milan Ščasný, Ph.D.
E-mail:	sarlota.smutna@gmail.com	E-mail:	milan.scasny@czp.cuni.cz
Specialization:	Economic Theories	Defense planned:	June 2015

Proposed topic

Household food demand in the Czech Republic: coherent demand system dealing with selectivity

Motivation

Food represents one of the most basic needs in human life and therefore it is one of the most studied topics in the field of demand analysis. There are several demand systems available to analyse the consumers' behaviour. The widely popular and one of the most used demand systems is the AIDS – Almost Ideal Demand System. Particular problem can arise when working with budget survey data. This is the only one type of data (beside time series and scanner data) where zero observations occur. The significant percentage of zeros can appear when studying demand for specifically defined food items, for example closely specified type of meat (beef) or milk (low-fat milk). This leads to the censoring of the data. Moreover, the censoring is given by the consuming (positive observations) and non-consumig (zero observations) households. Hence, it is given by some specific rule which means that the uncensored data is not random sample. This aspect is called selectivity. Both, selectivity and censoring, need special attention, otherwise the estimation would lead to the biased results. There are several available techniques how to deal with censored data in demand systems. One of the main approaches is the Heckman two-step procedure, which has been also implemented in demand systems in the literature

The aim of this thesis is therefore to study different demand systems that are used for food demand analysis together with examining different approaches which deal with zero observations in order to analyse food demand of Czech households.

Hypotheses

1. Different techniques dealing with selectivity given by the censoring of data give similar results.

2. The estimates of elasticities are biased when the selectivity of data is not treated.

3. Food demand elasticities are inelastic in income.

4. The elasticities differ for different segments of households distinguished according to the income.

Methodology

The household food demand will be analysed for the case of Czech Republic. The demand elasticities (income, own-price, and cross-price elasticities) will be the main outcome of the analysis. The data is collected by the Czech Statistical Office but it will be provided by the Environment Center of Charles University in Prague. It consists of 3000 Czech households and includes all the incomes and expenditures information as well as socio-demographic information. The aim is to estimate demand for food divided in subcategories such as dairy or bakery products. Next step will be to choose some of these categories and estimate the demand for more specific items included in the categories. The reason is that for such items there is higher probability to deal with problem of censoring. The choice of the demand system will be made arbitrarily according to available literature and particular characteristics of different systems such as AIDS or LES - Linear Expenditure System. Concerning the censoring techniques, the attempt is to apply a variety of them within the one chosen system and compare their results between each other.

Expected Contribution

The output of this thesis beside the food demand analysis for Czech households should be the comparison of different approaches to deal with censoring implemented in the demand system estimation. The results for Czech data will be unique as there are only few complex studies held on them which are focusing on food until now.

Outline

1. Introduction: Motivation
2. Literature review
   (a) Demand systems used for food demand analysis
   (b) Aspect of selectivity
   (c) Analyses conducted on the Czech data
3. Empirical analysis
   (a) Use of different selectivity approaches to estimate food demand system for Czech Republic
   (b) Comparison and discussion of results
4. Conclusion

Core bibliography

[1] Takeshi Amemiya. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica*, 42(6):pp. 999–1012, 1974. ISSN 00129682. URL http://www.jstor.org/stable/1914214.

[2] Angus S Deaton and John Muellbauer. An almost ideal demand system. *American Economic Review*, 70(3):312–26, 1980.

[3] Diansheng Dong, Brian W. Gould, and Harry M. Kaiser. Food demand in mexico: An application of the amemiya-tobin approach to the estimation of a censored food system. *American Journal of Agricultural Economics*, 86(4):1094–1107, 2004. URL http://EconPapers.repec.org/RePEc:oup:ajagec:v:86:y:2004:i:4:p:1094-1107.

[4] Richard Green and Julian M Alston. Elasticities in aids models: a clarification and extension. *American Journal of Agricultural Economics*, 73(3):874–875, 1991.

[5] James J Heckman. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153–162, January 1979. URL http://ideas.repec.org/a/ecm/emetrp/v47y1979i1p153-61.html.

[6] Dale Heien and Cathy Roheim Wessels. Demand Systems Estimation with Microdata: A Censored Regression Approach. *Journal of Business & Economic Statistics*, 8(3):365–71, July 1990. URL http://ideas.repec.org/a/bes/jnlbes/v8y1990i3p365-71.html.

Author | Supervisor

Chapter 1

Introduction

Food represents one of the most basic needs in human life and therefore it is one of the most studied topics in the field of demand analysis. The interest can lie on the food in general, on the food category (meat products, for instance) or on the specific food item (pork). The most useful parameter of the consumer demand is an elasticity. Different types of elasticities are distinguished and express the sensibility of the consumers with respect to the changes on the market such as prices of goods. The interest lies on them because of their broad variety of applications.

Different statistical models have been developed to analyse the demand and to estimate the elasticities. Single equation framework or demand systems are used. The demand systems are more complex and are assumed to be more closed to the reality. They relate all the goods in the consumer basket, which are at the same level. For example, all the food categories or all the food sub-items regrouped in one food category such as meat products or milk products. Since 1940s various demand systems have been proposed. The commonly used in the past and nowadays is the Almost Ideal Demand System and its derivations.

When estimating such models one specific difficulty can appear. Among different types of data that can be used, the budget survey data can contain part of observations equal to zero. By the zero, it is meant the consumer who does not consume given item and thus the appropriate quantity or expenditure is equal to zero. Such data with significant part of zeros is censored. The zeros can be explained by different means as the consumers have different motivations to not consume the given food item. The censoring of data is then given by different stochastic mechanisms as well. Usually, only the positive observations can be used in the estimating process, then the sample used is not random but censored by this specific stochastic mechanism. Such estimation can lead to biased results. Various treatment techniques have been developed during the time such as Heckman's estimator.

There are two principal goals of this thesis in the theoretical part. Firstly, the thesis aims to study and describe the demand systems which are commonly use to analyse the food demand. Secondly, the distinct approaches to deal with the selectivity problem and censoring of the data will be examined as well. There are not many techniques which have been already incorporated in the demand systems and clearly described. The goal is to choose the most fitting demand system and apply on it different censoring techniques and compare these techniques between each other empirically. The empirical goal is to analyse the food demand of Czech households through the chosen demand system with treatment of censoring. The data for the analysis is collected by the Czech Statistical Office and is provided by the Environment Center of Charles University in Prague. It is budget survey data consisting of income, expenditure and socio-demographic information about approximatively 3000 households. The data from 2013 is used.

The remainder of the thesis is organized as follows. The next chapter brings the literature review. It summarizes available demand systems which are used for food demand and it presents the treatment techniques dealing with the selectivity problem. The Chapter 3 describes the methodology together with incorporation of treatment techniques in the demand system, the data that is worked with is described as well. Further, the results of the empirical part of the thesis are presented - the analysis of the demand upon the budget survey data in case of the Czech Republic. The Chapter 4 discusses the results and findings regarding the different selectivity techniques and finally the Chapter 5 concludes the thesis.

Chapter 2

Literature review

2.1 Motivation

Elasticities are the most frequent and the most important outcome of demand analyses. They express the sensitivity of the consumer on the change of some particular parameter of the good (its price), on the change on the market (price of other goods), or on the change of consumer's conditions (income). Principally, the own-price e_{p_i}, the cross-price e_{p_j} and the income elasticity e_M are distinguished respectively. Let x stands for the quantity of good i demanded, p is price and M is income, then the different elasticities are defined as:

$$e_{p_i} = \frac{\%\Delta x_i}{\%\Delta p_i} \tag{2.1}$$

$$e_{p_j} = \frac{\%\Delta x_i}{\%\Delta p_j} \tag{2.2}$$

$$e_M = \frac{\%\Delta x_i}{\%\Delta M} \tag{2.3}$$

The interest to estimate elasticities lies on their wide usage. For example, the company producing a good can evaluate how the sales will change with the price change. The governments can evaluate the impacts of change of tax system on the sales and thus on the new tax revenues. The income elasticities could help to see what basket of food is available for low income or no-income people in developing countries and provide the important information when building up food policies.

In order to estimate the elasticity, the single equation models or more complex demand systems are used. In the single equation framework, the attention is paid on the one given food. As an illustration, it can be the linear equation in the logarithmic form estimated by the Ordinary Least Squares (OLS):

$$\ln q = X\beta + \epsilon, \tag{2.4}$$

where q is the demanded quantity of food item of interest, X is the matrix of explanatory variables, β are the coefficients to estimate and ϵ are disturbances. Among explanatory variables, there are usually included the prices, the income and various socio-demographic information. In case of log-log model where the explanatory variables are also taken in natural logarithm, the elasticities are then expressed directly by the β coefficients linked to income and prices.

On the other hand, the demand systems allow for interaction of the food of interest and other food items at the same level. The outcome is not the behaviour of the customer towards one good only but towards the different goods or their groups. Usually, the goods are linked through the budget of the household or individual. Obviously, the demand systems are more demanding than single equation models but they are probably more close to reality thanks to their complexity. The widely popular and one of the most used demand systems is the AIDS – Almost Ideal Demand System and some of its derivations such as linear or quadratic form of AIDS. In the next section, the main demand systems which are used for food analysis beside the AIDS are presented.

For the analysis, generally the three basic types of data can be employed - time series, budget survey and scanner data. Firstly, the time series are aggregated data on the macroeconomic level, it means usually a sum or average over all consumers in one year. It is commonly calculated by the national statistical offices in most of the states. For the purpose of demand analysis this data has very important disadvantage, there is no information about individual consumers. Secondly, the scanner data became to be available around 1980 thanks to technological development. It is based on the recording of all the transactions with goods at the counters in the stores. It gives the information about prices and quantities of sold food. On the other hand, it does not record any information concerning particular consumer. Usually, the socio-demographic characteristics of inhabitants living in the area around the store are used. This type of data can be used, for example, to examine the spatial allocation of the demand or the effect of advertisements. Thirdly, the budget survey data is very common because of having relatively immense advantage as it provides socio-demographic characteristics of the consumers involved in the survey. When analysing this type of data, it is necessary to be careful when making the conclusions as the data concerns usually a representative sample of the target population only. At the same time, the budget survey data is the only type of the data where the zero observations occur. By the zero, it is meant the zero consumption of the given food. The significant percentage of zeros can appear when studying demand for specifically defined food items, for example, closely specified type of meat (beef) or milk (low-fat milk). Lastly, the budget survey or another survey data can be collected on year or different basis which can give the panel or pseudo-panel data. This type of data is to some extent a combination of survey data and time series.

Having a significant part of observations equal to zero (non-consuming households) in the dataset and other observations continuous over positive values, the data is then called censored, more precisely left-censored.[1] The motivation or explanations of reported zeros are several. For illustration, the consumers are not interested in the item at all (vegetarians and meat products) or the consumers buy given item seasonally (sugar or salt are not purchased on the daily base). Obviously, the consuming and non-consuming households are distinguished by some specific rule. It means that the censored observations (the consuming households) do not create a random sample. This fact is called the selectivity. Consequently, the censoring of the data usually causes the biasedness of the estimated results. Hence, such a dataset need a special attention. The treatment depends on the mechanism how the censoring is produced. The different mechanisms and their treatments are discussed in one of the sections below.

Therefore, when conducting the demand analysis on budget survey data, it is important to treat the estimating process because of the selectivity problem and incorporate these techniques in the statistical model. As the goal of this thesis is to compare empirically different treatment techniques applied on the suitable demand system, firstly the different existing demand systems are described. Secondly, the different treatment techniques are explained. In the empirical part the implementation of these techniques in the chosen demand system is discussed and the analysis is conducted.

2.2 Demand systems

To analyse the consumers' demand behaviour there is single equation framework and there are demand systems. Both methods are used and bring different advantages.

The single equation models are simple as they are aimed on the analysed item and give it in relation with other items arbitrary chosen. They are partially isolated from the overall conditions of the market. They focus particularly on the estimation of elasticity. It is reproached them that they do not take enough into account the consumer theory.

On the other hand, the demand systems are usually derived from the conditions described in microeconomic theory of consumer. They are trying to be complex as they are aiming to be comprised of the sum of all demand functions which form the consumer basket or the compacted part of the basket. The first demand system based on the consumer preferences is the Linear Expenditure System (LES) derived in 1947-48 by Klein and Rubin which was firstly implemented by Stone (1954). Since this paper, the variety of systems has been developed.

The nowadays systems focus on the complex problem and are built on simultaneous equations which incorporate mutual dependences between large number of commodities in consumer basket. The goal is still to estimate the elasticity, the income and

[1] The data can be also truncated, the difference between these two terms is precise later.

price elasticity. There are two main features to be accomplished by demand systems. They are supposed to be flexible enough to be able to differ between income elasticity for particular income segments of population. The reason is that generally with the increasing income, the income elasticity becomes to decrease. Next, they are supposed to be able to deal with zero demand in order to not lose observations. (Kumar et al., 2011)

Beside LES which is not used so frequently, the very common systems are the AIDS and its modifications such as linear approximation LA/AIDS or quadratic form QUAIDS. It is one of the most spread and applied method to examine the consumer demand for various agricultural commodities during last four decades. The authors employing the AIDS usually appreciate the simplicity, clearness and consistency with microeconomic theory of consumers' behaviour. The other models used nowadays as well are: the Rotterdam Demand System, the Quadratic Expenditure Demand System, the Translog Demand System or CBS and NBR models which are derived on the base of the AIDS and Rotterdam model and are called according to Netherlands Central Bureau of Statistics and National Bureau of Research, the employers of the authors of the model - Keller van Driel and Neves.

Unfortunately, it is not defined which model is appropriate to use in particular situations. It always lies on the author to consider what is the goal of his work, what type of date he works with and then finally decide what model to choose. Obviously, the models stay to be the approximations of the real relationships on the market between consumer and goods and between particular goods with each other. The closeness of these approximations is always as good as the chosen model is appropriate in the given situation and as the model is able to explain the reality.

2.2.1 Linear Expenditure System - LES

The goal of Klein and Rubin (1947) was to propose a cost-of-living index which is dependent on the measurable prices only and based on the properties of the demand functions. The index is defined as a ratio of two incomes. The income in the denominator is the income of base period. The income in the numerator is the smallest value one would need to buy such a basket of goods at current prices in order to reach the same rate of utility as in the base period. One of the derived equations served to Stone as a base for building the LES.

The objective of Stone (1954) was "to derive a practical system of demand equations which possess properties usually considered desirable from the standpoint of elementary economic theory." There are three such properties:

- **Additivity**: sum of expenditures for different items has to be equal to the total expenditures $\sum_i p_i q_i = \mu$, where p_i is price and q_i quantity of item i, μ stands for total expenditures

- **Homogeneity**: sum of all price elasticities and income elasticity has to be equal to zero $\sum_j e^i_{p_j} + e^i_M = 0$
- **Symmetry of the substitution matrix (Slutsky condition)**: the demand effects has to be symmetric on both sides, if the substitution term is s, then $s_{ij} = s_{ji}$.

Stone (1954) derived the system of equations which directly fulfils these properties from its definition. The general form of the system is (following the notation of the article):

$$pq = b\mu + Bp \quad \text{with} \quad B = (bi - I)c \tag{2.5}$$

where p and q are vectors of prices and quantities of different items j, respectively, b is a vector of constants such that $\sum_j b_j = 1$, it practically expresses the proportion of item j on total expenditures, c is a vector of constants, i is a unit vector and I is a unit matrix. Actually, the constants in b have to satisfy $0 < b_j < 1$ which ensures the Slutsky condition and excludes inferior and complementary goods.[2]

If c is the null vector, the model is called to have a naïve form. The vector of constants b can be then estimated by OLS regressing different $p_j q_j$ on total expenditures μ. To estimate the non zero vectors of constants b and c of the so called sophisticated form, the author has proposed to use the iterative procedure where the initiate estimate of b will be used to estimate c. The estimate of c will be used to estimate b more precisely and then the whole process will be repeated until receiving the stable results with given precision. The two forms of the system can be mixed, different forms can be assigned to different commodities according to their nature in one system.

Stone (1954) examined this system on British data over the years 1920-1938 to analyse the system of six large groups of commodities. He compared the calculated expenditures by the system with the real expenditures by the sample correlation coefficient between these two. By the appropriate combination of sophisticated and naïve form, he reached the coefficients larger than 0.90 for every group of commodities. He also tried to forecast the consumption for 1952 by analysing the data from 1900 with satisfactory results taking into account the influence of wars and big changes in the consumer behaviour over 50 years.

Berges and Casellas (2002) have applied the LES to estimate the own-price, total food expenditure and income elasticities of different food categories for two different groups of households in Argentina. The households have been divided between those above and below the poverty guideline as the aim of the study was to discuss the impact of public policies that redistribute the income to poor households on food demand and nutrient and calories intake. The authors give several advantages why they have

[2]See Stone (1954) for more details.

chosen this type of demand system: "it has a straightforward and reasonable interpretation, it is one of the few systems that automatically satisfy all the demand theoretical restrictions and it can be derived from a specific utility function." Since the original purpose of iterative approach of estimating the system, the advanced methods are already known such as Maximum Likelihood estimator (MLE). Moreover, the authors have used the approach of Shonkwiler and Yen (1999) to deal with censored observation, so the system was estimated differently in two stages. The study has shown that lower-income households eat much more meat (traditionally given) and the diet of higher-income households is much more diversified and includes the ready-to-eat meals. The elasticities have been significantly different for two groups of households and that encourages the implementation of food policy for poor households.

Raper et al. (2002) have also applied the LES to examine consumer behaviour of households labelled with poverty and non-poverty status. The aim was once again to estimate the price and expenditures elasticities of two different income groups for nine aggregated food categories. The difference is that they focused on the effect of the demographic variables in this case such as family size, race of the head of the household and the presence of children. The authors have chosen this type of the demand system even though they have pointed out some disadvantages. Firstly, the LES does not allow goods to play a role of complements, substitutes or to be inferior but the authors argue that analysing such broad food categories it is not possible to distinguish such types of goods. Secondly, this system works with constant budget shares that could be restrictive assumption for different income groups but the authors estimate the system for two income groups separately to overcome this difficulty. The demographic variables under the assumption that the quantities purchased are their linear combination were implemented in the model through one of the estimated coefficients. The problem of zero observations was solved by the Heckman's technique[3] incorporated in the system and the final system was estimated by the procedure of nonlinear Seemingly Unrelated Regressions (SUR). The authors' conclusion is that the consumer behaviour of households is different in dependence on the poverty and non-poverty status and on the demographic variables as well. They have also expressed the elasticities using the information on these variables. The contribution is that such information could improve the effectiveness of food programmes and to evaluate the impact of changes in these programmes thanks to the estimated elasticities.

2.2.2 Quadratic expenditure system - QES

The quadratic expenditure system is a generalization of the linear expenditure system, which is quadratic in total expenditures. The system was described by Howe et al.

[3]The technique is described later.

(1979). The class of quadratic demand functions has general form:

$$h^i(P,\mu) = \frac{1}{g^2}(\alpha_i - \frac{g_i}{g}\alpha)(\mu - f)^2 + \frac{g_i}{g}(\mu - f) + f_i, \tag{2.6}$$

where f, g, α are homogeneous functions of degree one, μ stands for total expenditures and P is a vector of prices. The authors present two specific form of these functions:

$$p_i h^i(P,\mu) = p_i b_i + a_i(\mu - \sum p_k b_k) + (c_i - a_i)\alpha \prod p_k^{-c_k}(\mu - \sum p_k b_k)^2 \tag{2.7}$$

and

$$p_i h^i(P,\mu) = p_i b_i + a_i(\mu - \sum p_k b_k) + (c_i p_i - a_i \sum p_k c_k) \prod p_k^{-2a_k}(\mu - \sum p_k b_k)^2, \tag{2.8}$$

where a, b, c are vectors of constants such that for (2.7) $\sum a_k = 1$ and $\sum c_k = 1$, for (2.8) $\sum a_k = 1$ only. To obtain shares w_i on the left-hand side of the equations (2.7) and (2.8), the right-hand side has to be divided by μ. This system in the form of (2.8) was applied on the time series U.S. data from 1929-1975 for estimation of the marginal budget shares and own price elasticities for four categories: food, clothing, shelter and miscellaneous. The estimation process was MLE.

Similarly, Pollak and Wales (1978) estimated the QES of the form of (2.7) for budget survey data. The QES of this form is equivalent to the LES in the case of $a_i = c_i$ for all i. They analysed marginal budget shares and own price elasticities for three categories: food, clothing and miscellaneous on the UK budget survey data from 1968 and 1972. They compared the QES and LES through log-likelihood ratio statistics and the result is that the QES is a significant improvement of the LES.

Pollak and Wales (1980) described and compared two different techniques of incorporating the demographic variables, specifically number of children and their age in this case, into QES and Translog demand system (discussed below). The two techniques are called demographic scaling and demographic translating and are based on the dependence of some parameter(s) on the demographic variables. Then the form of dependence has to be specify. Regarding the demographic translation, the original demand system

$$x_i = \bar{h}^i(P,\mu) \tag{2.9}$$

is modified in

$$h^i(P,\mu) = d_i + \bar{h}^i(P, \mu - \sum p_k d_k) \tag{2.10}$$

where x are quantities, p are prices, μ is expenditure and d are translation parameters that depend on the demographic variables η only. $D^i(\eta) = d_i$ are the functions relating the parameters d with the variables η. The authors give an example of linear demographic translating:

$$D^i(\eta) = d_i^* + \sum_r \delta_{ir}\eta_r. \tag{2.11}$$

Similarly, in case of demographic scaling the original demand system (2.6) is replaced by the

$$h^i(P, \mu) = m_i \bar{h}^i(p_1 m_1, p_2 m_2, ..., p_n m_n, \mu) \tag{2.12}$$

and the linear demographic scaling has a form of

$$M^i(\eta) = 1 + \sum_r \epsilon_{ir}\eta_r. \tag{2.13}$$

The authors compares these two techniques incorporated in QES and Translog demand system on the U.K. household survey data for 1968 and 1972. The analysis was conducted on three categories: food, clothes and others. According to log-likelihood the authors state that for their data the composition of households matter, they have significant impact on the consumer behaviour, and the scaling method reaches higher log-likelihood values than translating method regardless the form of the demand system.

The QES was applied, for example, in the study of Chern and Wang (1994) to evaluate behaviour of Chinese consumers. The motivation was the period of economic reforms in China dating from 1978 that led to the free market and its expansion. Some food products were under rationing, the grains and oils, the most important products, were rationed even until 1993. During this period, the income of habitants increased together with GNP and the question was how this increase will influence food demand. The demand system was built for 10 most important food commodities: grains, oil, fresh vegetables, dried vegetables, pork, beef, poultry, eggs, fish and fresh fruits. These products were chosen according to the Chinese dietary tradition and the share on living expenditures. For example, pork and grain expenditures built more than 15 % of living expenditures. The data used for this study consisted of average expenditures and quantities of food items for 28 Chinese cities and provinces for the period of 1985-1990. The system was estimated by iterative nonlinear SUR and the marginal budget shares, own-price and expenditure elasticities were determined. The drawn conclusions are that ”China has begun to follow the traditional path of economic development as observed in other developed and developing countries”, the increase in income implied the decrease of the share of total expenditure destined for food so the Engel’s law hold, but almost the half of the new income will be still spent for food and the income elasticities remain high. Lastly, authors estimated the QES and LES as well and their comparison led to the result that the two systems are equivalent regarding the estimated results in this case and both of them perform well.

2.2.3 Rotterdam demand system

Theil (1965) built up the proposal of the demand system on the probabilistic approach together with information theory. He started with simple question: "If we select at random a dollar of a consumer's or a company's budget, what is the chance that it will be spent on the ith commodity?" The answer is actually equal to the budget share of the commodity i, w_i. The shares have to also satisfy the nonnegativity and they sum up to one as probabilities. He suggested the general form of the demand function:

$$\frac{w_i + (w_i)_{-1}}{2}\Delta \ln q_i = A_i + B_i \Delta \ln \bar{\mu} + \sum_j C_{ij} \Delta \ln \bar{p}_j + u_i, \tag{2.14}$$

where some terms are closely specified, $\bar{\mu} = \frac{\mu}{p}$ is real income and p is price index, $\bar{p}_j = \frac{p_j}{p^*}$ is deflated price of the commodity i such as $\Delta \ln p^* = \sum_i B_i \ln p_i$. The Δ stands for first difference. On the other hand, Barten (1964) used specified form to analyse Dutch time series data:

$$\Delta \ln q_i(t) = \alpha_i + \sum_j \epsilon_i j \Delta \ln p_j(t) + \eta_i \Delta \mu(t) + c_i \sum_j w_j \Delta \ln p_j(t) + u_i(t) \tag{2.15}$$

estimated by SUR. Hence, these two papers are assigned to the definition of Rotterdam demand system which is called according to authors' site.

The consumers have started to think about how the nutritions' composition of food impacts their health during time. Capps Jr and Schmitz (1991) had besides others two goals - to build a theoretical framework taking into account the nutrition and health factors in demand analysis and apply it on the demand for beef, pork, poultry and fish with respect to the information about cholesterol's content. The Rotterdam demand system served as a starting point, they worked with the following form of the system (combination of two previous forms):

$$\frac{w_i + w_{it}}{2}\Delta \ln q_{it} = \alpha_i + b_i \left[\Delta \ln \mu_{it} - \sum_k \frac{w_k + w_{kt}}{2} \Delta \ln p_{kt}\right] + \sum_j c_{ij} \Delta \ln p_{jt} + \epsilon_{ij}, \tag{2.16}$$

where b_i is marginal budget share of commodity i and c_{ij} is compensated cross-price elasticity of i with respect to j weighted by budget share of i, these are coefficients to estimate. To ensure properties derived from economic theory, the restrictions can be imposed such that:

- additivity: $\sum_i b_i = 1$, $\sum_i c_{ij} = 0$
- homogeneity: $\sum_j c_{ij} = 0$ for $\forall i$
- symmetry $c_{ij} = c_{ji}$.

To implement the information about cholesterol content, the authors replace α_i coefficient by $\gamma_i \ln CHOL_{it}$ with restriction $\sum \gamma_i = 0$. This system was applied on the U.S. annual time series data for 1966-1988. For the estimation the iterative Zellner estimation procedure was used with dropping out the equation of the fish demand. Authors found that "there exists sample evidence to indicate that cholesterol information, with a half-year lag, is a statistically significant determinant in the consumption of pork, poultry, and fish."

Recently, the Rotterdam demand system was employed by Anwar et al. (2012) to analyse the food demand in Pakistan. Their motivation was a lack of the studies of price elasticities in Pakistan. The model was conducted for nine major food products: wheat, rice, milk, mutton, chicken, apples, mango, potato and onion; to estimate own-price, cross-price and expenditure elasticities. The estimation was made by SUR technique on the household survey data from 2007-2008.

Another example is the article of Selvanathan and Selvanathan* (2004) that investigated the demand for alcohol in Australia through this demand system as well. The motivation was to explain the rapid decrease in alcohol consumption during last decades caused mainly by fall of the beer consumption. They edited the Rotterdam model to incorporate effects of economic changes and demographic effects conditional on the alcohol market.

2.2.4 Translog demand system

Christensen et al. (1975) have developed later called Translog demand system. After the works of Stone and Theil or Barten, their motivation was "to develop tests of the theory of demand that do not employ additivity or homotheticity as part of the maintained hypothesis and to exploit the existence of an indirect utility function defined on total expenditure and the prices of all commodities implied by a complete model of consumer demand." They describe the direct transcendental logarithmic (translog, in short) utility functions which can be used to characterize the system of indirect demand functions and indirect translog utility functions useful for the system of direct demand functions. These two systems are dual for each other. The direct translog utility function is derived as:

$$\ln U = \ln U(x_1, x_2, ..., x_m) \tag{2.17}$$

$$-\ln U = \alpha_0 + \sum_i \alpha_i \ln x_i + \frac{1}{2} \sum_i \sum_j \beta_{ij} x_i x_j \tag{2.18}$$

which leads to

$$w_j = \frac{p_j x_j}{\mu} = \frac{\alpha_j + \sum \beta_{ij} \ln x_i}{\sum(\alpha_k + \sum \beta_{ki} \ln x_i)} \tag{2.19}$$

and the indirect translog utility function is described as:

$$\ln V = \ln V(\frac{p_1}{\mu}, \frac{p_2}{\mu}, ..., \frac{p_m}{\mu}) \tag{2.20}$$

$$\ln V = \alpha_0 + \sum_i \alpha_i \ln \frac{p_i}{\mu} + \frac{1}{2} \sum_i \sum_j \beta_{ij} \frac{p_i}{\mu} \frac{p_j}{\mu} \tag{2.21}$$

which leads to

$$w_j = \frac{p_j x_j}{\mu} = \frac{\alpha_j + \sum \beta_{ij} \ln \frac{p_i}{\mu}}{\sum(\alpha_k + \sum \beta_{ki} \ln \frac{p_i}{\mu})} \tag{2.22}$$

where $\sum \alpha_i = -1$ can be normalized in both cases. Then the authors give a series of parameters restrictions which can be tested and thus it can be evaluated if the functions fulfil the characteristics of the theory of demand such symmetry and further the additivity and homotheticity, for instance. The empirical test of the theory of demand is conducted as well on the U.S. time series data from 1929-1972 for three-commodity system: services of consumers' durables, nondurable goods, other services and for both indirect and direct translog utility functions. To estimate the system, it is necessary to specify the distribution of the disturbances added at the right-hand side of either (2.19) or (2.22). Then, the $(k-1)$ equations are necessary to estimate the demand system for k categories by MLE. In the same time, it does not matter which equation is excluded. Then the results are tested against the estimated coefficients obtained under restrictions through the test statistic based on the likelihood ratio. The authors have finally found that the evidence and the theory of demand are inconsistent.

Nicol (1989) has extended the model by the third-order Taylor series expansion term of the indirect translog utility function:

$$\ln V = \alpha_0 + \sum_i \alpha_i \ln \frac{p_i}{\mu} + \frac{1}{2} \sum_i \sum_j \beta_{ij} \frac{p_i}{\mu} \frac{p_j}{\mu} + \frac{1}{6} \sum_i \sum_j \sum_m \beta_{ijm} \frac{p_i}{\mu} \frac{p_j}{\mu} \frac{p_m}{\mu} \tag{2.23}$$

thus

$$w_j = \frac{p_j x_j}{\mu} = \frac{\alpha_j + \sum \beta_{ij} \ln \frac{p_i}{\mu} + \frac{1}{2} \sum \sum \beta_{ijm} \ln \frac{p_i}{\mu} \ln \frac{p_m}{\mu}}{\sum(\alpha_k + \sum \beta_{ik} \ln \frac{p_i}{\mu} + \frac{1}{2} \sum \sum \beta_{ikm} \ln \frac{p_i}{\mu} \ln \frac{p_m}{\mu})}. \tag{2.24}$$

The goal of this study was to test the validity of the presence of this third-order term on the evidence from Canadian data. The household survey data from 1978 and 1982 were stratified by the family size and for each out of 4 groups the Translog demand system was estimated for three categories: durable goods, nondurable goods and services. The demand system was estimated by MLE on two equations. The author has affirmed that the division of the data according to the family size is necessary as the hypothesis of having equal parameters was rejected. The importance of third-order term was statistically proved as well.

One possibility how to incorporate demographic variables in the Translog demand system was already described above based on Pollak and Wales (1980). Yen et al. (2002) developed the coefficient term α_i to involve the demographic variables in the Translog system:

$$\alpha_i = \sum_k \alpha_{ik} z_k \tag{2.25}$$

where z_k are demographic variables with $z_1 = 1$. Yen et al. (2002) have focused on the demand for fats and oils among U.S. households because of significant increase in the consumption during last 50 years (more than twice higher). Second authors' motivation was the fact that most of the studies on this topic were conducted on the time series data and the aggregated data does not permit to distinguish between fats used in industry, used for production of another food items or directly consumed. The analysis was conducted on the U.S. household survey data from 1987-88 for five products: butter, margarine, shortening, cooking oil and salad dressing. Because of frequent zeros in the data, for example, the 38.25 % of households do not consume 3 out of these 5 products, the treatment censoring technique of Shonkwiler and Yen (1999) was incorporated in the Translog demand system. Price elasticities, total expenditure elasticities and elasticities with respect to demographic variables were estimated as well.

For example, the Translog demand system was recently used in the study of Davis et al. (2009) as well. The authors used this model to estimate the demand elasticities for fluid milk products on the U.S. market.

2.2.5 Almost Ideal Demand System - AIDS

Deaton and Muellbauer (1980) have proposed the Almost Ideal Demand System, the AIDS. The derivation departs from the model of PIGLOG preferences (price-independent generalized logarithmic preferences) which allow to sum over all consumers or households in the one representative. The model is based on the equations

$$w_i = \alpha_i + \sum_j \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{\mu}{P}\right), \tag{2.26}$$

where

$$\ln P = \alpha_0 + \sum_k \alpha_k \ln p_k + \frac{1}{2} \sum_j \sum_k \gamma_{kj} \ln p_k \ln p_j. \tag{2.27}$$

The w_i stands for the share of item i on the total expenditures, the p are the prices, the μ is the value of total expenditures as before and P is the price index. The α, β, γ are the coefficients to be estimated and they are supposed to fulfil the following requirements:

$$\sum_{i=1} \alpha_i = 1 \quad \sum_{i=1} \gamma_{ij} = 0 \quad \sum_{i=1} \beta_i = 0 \tag{2.28}$$

$$\sum_j \gamma_{ij} = 0 \tag{2.29}$$

$$\gamma_{ij} = \gamma_{ji}. \tag{2.30}$$

The requirements are developed to fulfil the necessary conditions from the microeconomic theory - additivity, homogeneity and symmetry of substitution matrix. The additivity is fulfilled by (2.28). The homogeneity is accomplished by (2.29). The third requirement, the Slutsky condition is fulfilled by (2.30). The advantage is, that the conditions can be tested upon the coefficients estimated values and the model does not have to be automatically restricted by them. It is possible that the real maximization problem (or the evidence) does not follow the theory. Moreover, considering the budget shares constant, the coefficients have straightforward interpretation. Having $\frac{\mu}{P}$ constant, the γ_{ij} represents the effect of 1 percent change in p_j on the budget share i times one hundred, thus γ_{ij} illustrates the change in relative prices. The change in real expenditures is represented by β_i, positive β_i stands for luxuries and negative for necessities.

The authors recommend to estimated the system of equations by the Maximum Likelihood estimator. They add that the model could be estimated also by the OLS in the case the price index P is known.

One of the modifications of the model uses the Stone index as the price index: $\ln P^* = \sum w_k \ln p_k$. It can be written that $P \simeq \phi P^*$ and thus the base equation has the form of

$$w_i = (\alpha_i - \beta_i \ \ln \phi) + \sum_j \gamma_{ij} \ \ln p_j + \beta_i \ \ln \left(\frac{x}{P^*} \right). \tag{2.31}$$

This modification is called linear approximation of the AIDS, shortly LA/AIDS. Instead of the Stone index, the Laspeyer's index can be used as well, for instance.

The authors applied their new demand system to the real data as well. They estimated the demand on the British time series data from 1954-1974 for eight commodities' categories: food, clothing, housing services, fuel, drink and tobacco, transport and communication services, other goods, and other services. Their results are according to their words "credible and in line with other studies." They have also compared the performance of AIDS and LA/AIDS and estimates of both are very close so the LA/AIDS is a good approximation of AIDS as P^* and P were never different by more than 0.008 in their case. Similarly, the values of the log-likelihood of two models were basically equal.

Blanciforti and Green (1983) proposed the dynamic form of the AIDS model. They have extended the model by the habits of consumers through the coefficient α:

$$\alpha_i = \alpha_i^* + \alpha_i^{**} q_{it-1}$$

where the habit is defined by the quantity of items i purchased in the previous periods. In this framework, it is important to take into account the characteristics of disturbances and potential problem with autocorrelation of errors.

Heien and Pompelli (1988) extended the AIDS model by incorporating the socio-demographic variables d_j. They included them through the coefficient α as well:

$$\alpha_i = \rho_{i0} + \sum_{j=1}^{S} \rho_{ij} d_j, \quad i = 1, ..., n \tag{2.32}$$

In this case, the condition of additivity has to be adjusted:

$$\sum_{i=1}^{n} \rho_{i0} = 1 \qquad \sum_{i=1}^{n} \rho_{ij} = 0$$

They applied this version of the AIDS on U.S. household survey data from 1977 to estimate the price and expenditures elasticities for specific types of beef: steak, roast, and ground beef.

Finally, the last of the known version of the AIDS is the quadratic AIDS (QUAIDS) which breaks the linearity in expenditures. This model assumes that the relation between income and expenditures is not linear. Between others, Mittal et al. (2010) used this model specification to analyse food demand in India. The QUAIDS model proposed by Banks et al. (1997) has the form:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \ \ln p_j + \beta_i \ \ln\left(\frac{\mu}{P}\right) + \lambda_i \big(\prod_i p_i^{\beta_i}\big)^{-1} \ \ln\left(\frac{\mu}{P}\right)^2 \tag{2.33}$$

and the additivity condition is extended by

$$\sum_i \lambda_i = 0. \tag{2.34}$$

Mittal et al. (2010) used extended version of this model, she incorporated the dummy variable indicating the rural or urban location of the household and the QUAIDS model was the second step of two-stage procedure. Nevertheless, she applied this model to the household expenditure survey data of India from different periods from 80's till 2000. The system was built for 6 commodities: cereals, pulses, fruit and vegetables, milk, edible oil, and sugar. The economy in India is in the rapid development that's why the appropriate demand analyses are important for proposing adequate policies.

2.3 Selectivity problem

It was already mentioned that basically three types of data are used to analyse demand and these are time series, budget surveys and scanner data. Nowadays, the budget survey data is very popular and common as it provides the information about socio-demographic characteristics of the consumers directly. Moreover, it is the only type of data where the zeros occur. By the zero, it is meant that the household or the individual reports the no consumption of the given food.

The no consumption reported can have several explanations. Firstly, it is important to distinguish whether the consumer is interested in the given food, it means he figures on its market or not. In the case a consumer is not at the market the zero stands for no utility from the item. Secondly, the consumer can figure on the market but can still report zero if the zero is the corner solution of his utility maximization problem. For example, the price is unavailable at the moment or the substitutes are more convenient. Thirdly, the length of the survey can play an important role. When the period is too short for a given food, it is possible that the item is not purchased in this period as it is simply needed once a longer period of time. This situation is called as the problem of the infrequency purchase, in the literature. Finally, it is always possible that the data is misreported.

When the observations of the variable are given by significant portion of zeros and the rest of values is continuous over strictly positive values, the variable is called limited. The data is called censored in the situation where the information on the side of the dependent variable is lost. Then there is the truncated data where the information on both sides is lost - the dependent and independent variable or variables are limited. Let consider the example of simple demand analysis where the dependent variable is the quantity of the purchased food that is limited by zero. In this case the data is censored. On the other hand, when the dependent variable is the logarithmic form of the quantity, the zero quantity cannot be used and the whole observation has to be deleted, then it is truncated data. To avoid such a loss of the information, it is possible to assign to these observations some minimal value which is not in the dataset otherwise and hence the date is censored. This is the reason why the attention is paid here on the censored data problem only.

In case of the censored data, the classical estimation procedures face problems and can be biased. The variety of estimators or tools had been developed over time to overcome possible biasedness.

Notation

To make the following description of the different models and estimators easier and clear enough, the notation is synchronized. In this section, it is quickly introduced to keep the text coherent and comprehensible. Let consider the analysis of demand for the type of food A, then

- y^* is the real quantity of A purchased
- y is the observed variable representing the quantity of A purchased

Next, it is supposed that demand of A is given by the following relation:

$$y^* = X\beta + \epsilon,$$

where X is the matrix (i x k) of k variables explaining the demand and of i observations and ϵ is the vector of error terms with the assumption of $\mathrm{E}[\epsilon] = 0$. The interest lies on the consistent estimation of β which is the vector of k coefficients. To estimate the β, the censored data y is only at the disposition. It is also useful to distinguish participation equation and outcome equation. Participation equation gives the mechanism to clarify whether the consumer participates in the market or not (he purchases the A or not). Outcome equation gives the mechanism to determine what quantity of A is purchased. These two mechanisms can be assumed to be identical or different.

The representative model assuming identical stochastic process for both equations is the tobit model described below. Further models assume that the participation and the outcome equation are given by the different stochastic mechanisms, basically different sets of variables are valuable for the equations:

$$\begin{aligned}
\text{participation equation: } y_1^* &= X_1\beta_1 + u \\
\text{outcome equation: } y_2^* &= X_2\beta_2 + v \\
\text{observation: } y_2 &= \begin{cases} X_2\beta_2 + v & y_1^* > 0 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

The main characteristic of models working with this decision in two steps (firstly decide if to buy or to not buy, if yes, secondly decide about the quantity) is the question of correlation of error terms u and v. In the following, both models, considering the zero and nonzero correlation of u and v are discussed.

The model assuming zero correlation of disturbances is usually in the literature called two-part model. The selection models are those who are allowing for the nonzero correlation of the error terms. They are called selection models as the sample with which it is worked is not chosen randomly but it is given by the rule. This rule is represented by the participation equation and the models follow the above given structure. Once again, when imposing assumption on the distribution of error terms (usually, the bivariate normal with zero means), this simple sample selection model can be estimated by the MLE.

2.3.1 Identical stochastic process: Tobit model

Tobin (1958) proposed his nowadays called tobit model. The important assumption in this framework is that 'the explanatory variables influence both, the probability of limit response and size of non-limit response.' Applying this on the problem of demand analysis, it says that the participation and outcome equation are given by the same stochastic mechanism. It means that the same variables to the same extent determine whether the consumer participates in the market and in the same time what quantity he is buying. The model can be summarized as:

$$y^* = X\beta + \epsilon$$
$$y = \begin{cases} X\beta + \epsilon & y^* > 0 \\ - & y^* \leq 0. \end{cases}$$

The final model $y = X\beta + \epsilon$ can be simply estimated by the MLE technique when the assumption of distribution of errors is made, usually the normal distribution with zero mean is assumed.

2.3.2 Different stochastic process with no correlation: Double hurdle model

Double-hurdle model belongs to the models not allowing for the correlation between the two error terms u and v. It was firstly described by Cragg (1971) and is of the form of:

$$\text{participation equation: } y_1^* = X_1\beta_1 + u$$
$$\text{outcome equation: } y_2^* = X_2\beta_2 + v$$
$$\text{observation: } y_2 = \begin{cases} X_2\beta_2 + v & y_1^* > 0 \wedge y_2^* > 0 \\ 0 & \text{otherwise.} \end{cases}$$

This form is quite specific as it considers the condition of $y_2^* > 0$, hence the infrequency purchase. When assuming the disturbances u and v being normally distributed and no correlated, the likelihood function to estimate this model is:

$$L_N = \prod_0 1 - \Phi\left(\frac{X_1\beta_1}{\sigma_u}\right)\Phi\left(\frac{X_2\beta_2}{\sigma_v}\right)\prod_+ \Phi\left(\frac{X_1\beta_1}{\sigma_u}\right)\Phi\left(\frac{X_2\beta_2}{\sigma_v}\right) \prod_+ \frac{1}{\sigma_v}\phi\left(\frac{y_2 - X_2\beta_2}{\sigma_v}\right)\Phi\left(\frac{X_2\beta_2}{\sigma_v}\right)^{-1} \quad (2.35)$$

where σ_u^2 is the variance of u and σ_v^2 is the variance of v. It is obvious that the reliability of the estimate is dependent on the distributional assumption of disturbances and also on their homoscedasticity as usually in case of the MLE.

2.3.3 Different stochastic process with correlation: Sample selection model

As already stated, the sample selection model consists of participation and outcome equation as stated above and at the same time the correlation between u and v is nonzero. The observations are results of specific rule, thus the sample is not random. Once again, assuming the bivariate normal distribution of disturbances, the model can be estimated by the MLE (Amemiya, 1984):

$$L_N = \prod_0 \Phi\left(\frac{-X_1\beta_1}{\sigma_u}\right)\prod_+ \Phi\left(\frac{\frac{X_1\beta_1}{\sigma_u} + \rho(y_2 - X_2\beta_2)/\sigma_v}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{(y_2 - X_2\beta_2)}{\sigma_v}\right)\frac{1}{\sigma_v}, \quad (2.36)$$

where ρ is the correlation coefficient of u and v and their covariance-variance matrix has following form

$$\begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \quad (2.37)$$

with $\sigma_u^2 = 1$ normalized. This process of estimation of sample selection model is sometimes denoted in the literature as FIML (full-information maximum likelihood). Alternatively, there exists also the Heckman's estimator denoted as LIML (limited-information maximum likelihood).

Parametric estimation: Heckman estimator

One of the most used tool to deal with selectivity problem is the Heckman estimator proposed by Heckman (1976). Basically, it is the alternative estimator of the sample selection model. It departs from the derivation of the bias which would result from the estimation of the outcome equation from the observed data using the OLS technique. Assuming that the errors u and v follow the joint density and hence can be rewritten as:

$$v = \frac{\sigma_{uv}}{\sigma_u}u + \xi = \sigma_{uv}u + \xi$$

because of the normalization that $\sigma_u = 1$ where σ_{uv} is the covariation of u and v and σ_u is the standard deviation of u, the bias is then derived as follows

$$E[y_2|X_2, y_1^* > 0] = X_2\beta_2 + E[v|X_2, y_1^* > 0],$$

with

$$\begin{aligned} E[v|X_2, y_1^* > 0] &= E[v|X_2, u > -X_1\beta_1] = E[\frac{\sigma_{uv}}{\sigma_u}u + \xi|u > -X_1\beta_1] \\ &= \frac{\sigma_{uv}}{\sigma_u}E[u|u > -X_1\beta_1] = \frac{\sigma_{uv}}{\sigma_u}\frac{\phi(-\frac{X_1\beta_1}{\sigma_u})}{1 - \Phi(-\frac{X_1\beta_1}{\sigma_u})} \\ &= \frac{\sigma_{uv}}{\sigma_u}\frac{\phi(-\frac{X_1\beta_1}{\sigma_u})}{\Phi(\frac{X_1\beta_1}{\sigma_u})} = \frac{\sigma_{uv}}{\sigma_u}\lambda\left(-\frac{X_1\beta_1}{\sigma_u}\right). \end{aligned}$$

The last term $\lambda(\cdot)$ is the Mills ratio which is unfortunately unknown as it requires the information about the vector of coefficients of the participation equation β_1 and $\sigma_u = 1$.

Heckman (1976) proposed to firstly estimate the participation equation by the probit model given the $\sigma_u = 1$ and obtain the $\hat{\beta}_1$. Then, the estimation of the Mills ratio can be calculated $\hat{\lambda}$ for each observation. Finally, the outcome equation can be estimated by the OLS technique using the nonzero observations only and including among the regressors the $\hat{\lambda}$.

This estimator of the sample selection model cannot be efficient because the errors v are by the definition heteroscedastic.

Semi-parametric estimation: Cosslett estimator

The problem of previous estimators is that in some way they rely on the distributional assumptions. When those assumptions are violated, the estimates are biased. That is the advantage of the semi-parametric and non-parametric techniques as they impose less or no restrictions on the distribution.

To see the performance of semi-parametric methods, the Cosslett (1991) approach has been chosen as a representative because of its simplicity and clearness of the application. It was already shown that the bias of the OLS estimation of the outcome equation arises form the term $E[v|u > -\mathbf{X_1}\beta_1] = \Psi\ (\mathbf{X_1}\beta_1)$. The functional form of the Ψ is unknown.

The first step is to estimate the $\hat{\beta}_1$ with unknown distribution F of the error term of the participation equation u. Cosslett (1991) proposes in his article two basic methods of the estimation. Firstly, he gives his own approach (Cosslett, 1983) and secondly he indicates Klein and Spady (1993). The first article presents the 'distribution-free maximum likelihood estimator of the binary choice model' and the second one inputs the kernel estimation of the density. Basically, the kernel estimation can be simply used to determine the distribution F. Another example of obtaining the distribution which is simple to implement is the Gallant and Nychka (1987) approach, the estimate is continuous in this case. Having the distribution $\hat{F}(u)$, the next step is to estimate the participation equation by the MLE using $\hat{F}(u)$ instead of posing distributional assumption.

The second step is to approximate the bias of the OLS estimate as in Heckman approach:

$$y_2 = X_2\beta_2 + \epsilon + \Psi(X_1\beta_1). \tag{2.38}$$

Cosslett (1991) proposes to use a series of dummy variables such as

$$y_{2,i} = X_{2,i}\beta_2 + \epsilon_i + \sum_{j=1}^{J} \lambda_j I(i \in I(j)) \tag{2.39}$$

where

$$I(j) = \{i|u^*_{j-1} < -X_{1,i}\hat{\beta}_1 < u^*_j \text{ and } y_{1,i} = 1\}, \quad \lambda_j = E[v|u > u^*_j]. \tag{2.40}$$

By the estimation of the distribution $\hat{F}(u)$, the step function is obtained. Its steps are indicating the number of intervals J and their boards u^*_j. To be precise, $u^*_0 = -\infty$, $u_{J+1} = \infty$, $I(J+1)$ is an empty set and $I(1)$ does not have to be non-empty, other subsets have to be non-empty. In case of the continuous function of $\hat{F}(u)$, the number of intervals and thus dummies has to be given arbitrarily as there is no guidance how to determine it. Practically, it is important to find experimentally such number which produces the stable estimate.

Among the regressors, the whole set of dummy variables is used, the constant term is not then identified. In such a case, it is possible to use the constant term proposed by Heckman (1990) or its variation introduced by Andrews and Schafgans (1998). Both constant terms are by these methods post-estimation determined and depend on chosen coefficients which are not clearly given. On the other hand, Hussinger (2008) uses the coefficient of the last dummy variable (λ_J) as the constant term which seems to be reasonable upon its experimental results.

2.4 Aim of empirical part

The goal of this theoretical part was to describe the main demand systems used for the food demand analysis and to present the techniques, which are dealing with significant part of zeros in the data set as a potential problem for further estimations. We have seen principal representatives of food demand systems such as LES, Rotterdam demand system, Translog demand system and AIDS. For example, the LES is more useful for analysis of broad food categories as it does not permit to distinguish for luxurious or necessary goods. On the other hand, the AIDS model is the most frequently used demand system since it has been defined till nowadays, the various versions included. The system of equations is clear in its definition, the interpretation of coefficients is straightforward and the different types of elasticities can be calculated upon its results. These are the reasons why the AIDS has been chosen for further analysis in this thesis.

Even the demand systems that are estimated upon budget survey data can meet the problems with frequent zeros in the data set. As the zero observations has to be often omitted, the results would be affected by the sample selection bias when considering that the zero observations are determined by some specific rule and thus the sample cannot be considered as a random sample. The sample selection problem was explicated in the previous section. The goal of the next empirical part of this thesis is firstly to describe various incorporation of censoring treatment techniques into the AIDS model. Secondly, explore the food demand analysis of Czech households through different censored versions of AIDS and compare these empi-rical results between each other to find out which technique is the most suitable and thus describes well the food demand in Czech Republic.

Chapter 3

Data analysis

3.1 Methodology

3.1.1 Estimation of AIDS model

The AIDS model or its modifications can be estimated by the MLE technique as suggested by the authors of the model. To derive the likelihood function, the distributional assumption has to be made. As in the most applications, the multivariate normal distribution of disturbances u_i, where $i = 1, ..., K$ is the number of equations or groups of commodities, is assumed:

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_K \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u_1} & \sigma_{u_1,u_2} & \cdots & \sigma_{u_1,u_K} \\ \sigma_{u_2,u_1} & \sigma^2_{u_2} & \cdots & \sigma_{u_2,u_K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_K,u_1} & \sigma_{u_K,u_2} & \cdots & \sigma^2_{u_K} \end{pmatrix}\right), \tag{3.1}$$

$$\text{equivalently: } \boldsymbol{u} \sim N(\boldsymbol{\mu}, \boldsymbol{E}). \tag{3.2}$$

From the probability density function

$$f_u(u_1, .., u_K) = \frac{1}{\sqrt{|\boldsymbol{E}|(2\pi)^K}} \cdot \exp\Big\{-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})'\boldsymbol{E}^{-1}(\boldsymbol{u}-\boldsymbol{\mu})\Big\} \tag{3.3}$$

the log-likelihood function can be derived for N observations and K equations

$$L_N = -\frac{N}{2}\ln(|\boldsymbol{E}|) - \frac{NK}{2}\ln(2\pi) - \frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})'\boldsymbol{E}^{-1}(\boldsymbol{u}-\boldsymbol{\mu}). \tag{3.4}$$

In case of the AIDS models the variance-covariance matrix $\boldsymbol{E}$ is singular because of the additivity across the system of equations ($\sum_i w_i = 1$). Thus the MLE is applied on the $(K-1)$ equations. Barten (1969) showed that the choice of the excluded equation does not have any influence on the results. Modifying the last term of equation (3.4), the concentrated log-likelihood function to estimate $(K-1)$ equations can be obtained such as in Poi (2002):

$$L_N = -\frac{N}{2}\ln(|S|) - \frac{N(K-1)}{2}\ln(2\pi) - \frac{N(K-1)}{2} \tag{3.5}$$

where

$$S \equiv \frac{1}{N}\sum_{h=1}^{N}\hat{u}_h\hat{u}_h', \quad \hat{u}_h \equiv [w_{1h} - \hat{w}_{1h}, ..., w_{K-1,h} - \hat{w}_{K-1,h}]. \tag{3.6}$$

On the basis of this function, the coefficients of $(K-1)$ equations can be estimated and the rest is determined through the additivity conditions. The equations can directly have the form including the price index P:

$$w_i = (\alpha_i - \beta_i\alpha_0) + \sum_j \gamma_{ij}\ln p_j + \beta_i[\ln(\mu) - \sum_j \alpha_j \ln p_j - \frac{1}{2}\sum_j\sum_k \gamma_{jk}\ln p_j \ln p_k] + u_i. \tag{3.7}$$

Second possibility is to estimate the restricted model by the conditions of homogeneity and symmetry. Comparing these two estimates one can test whether the evidence is in the accordance with the consumer theory.

The elasticities of the AIDS model can be estimated as (Green and Alston, 1990, 1991):

- income elasticity: $e_M^i = 1 + \frac{\beta_i}{w_i}$
- Marshallian price elasticity: $e_{p_j}^i = -\delta_{ij} + \frac{1}{w_i}(\gamma_{ij} - \beta_i\alpha_j - \beta_i\sum_k \gamma_{kj} \ \ln p_k)$.

The elasticities of the QUAIDS model can be estimated as (Banks et al., 1997):

- income elasticity: $e_M^i = 1 + \frac{1}{w_i}(\beta_i + 2\lambda_i\big(\prod_i p_i^{\beta_i}\big)^{-1}\ln(\frac{\mu}{P}))$
- Marshallian price elasticity: $e_{p_j}^i = -\delta_{ij} + \frac{1}{w_i}\Big\{ - \Big(\beta_i + 2\lambda_i\big(\prod_i p_i^{\beta_i}\big)^{-1}\ln(\frac{\mu}{P})\Big)$ $\Big(\alpha_j + \sum_k \gamma_{kj}\ln p_k\Big) - \lambda_i\beta_j(\prod_i p_i^{\beta_i})^{-1}(\ln(\frac{\mu}{P}))^2 + \gamma_{ij}\Big\}$.

The δ_{ij} is the Kronecker's delta which is equal to 1 if $i = j$ and equal to 0 otherwise. To estimate Hicksian price elasticities, the following relation can be used:

$$e_{p_j}^{i,C} = e_{p_j}^i + e_M^i w_j. \tag{3.8}$$

3.1.2 Incorporation of censoring techniques

All the articles known to the author, that are aiming to construct the censored system, are dealing with the selectivity problem directly. Actually, the selection model nests the two-part model. Moreover, the proposed censored systems which are supposed to be solved in one-step by the maximum likelihood are not used. These models are involving numerous complex integrals and are thus very demanding. The two-step estimation procedures are more popular as they are simpler to be applied.

Heien and Wessels' or Heckman's estimator

Heien and Wessels (1990) have proposed a two-step approach to estimate censored demand system. Concretely, it was AIDS model in their case. Their technique is simple to apply that is why it gained the popularity among researchers.

The probit model is estimated for each household and each equation in the system. Let denote the vector of explanatory variables for the dichotomous choice as $\boldsymbol{Z}$ and the corresponding coefficients η's. The consistent estimates of η's are obtained and the Inverse Mills Ratio (IMR) can be calculated:

- for consuming households: $IMR = \frac{\phi(Z'_{ih}\hat{\eta}_i)}{\Phi(Z'_{ih}\hat{\eta}_i)}$
- for non-consuming households: $IMR = \frac{\phi(Z'_{ih}\hat{\eta}_i)}{1-\Phi(Z'_{ih}\hat{\eta}_i)}$

where Φ is standard normal cumulative distribution function and ϕ is the standard normal probability distribution function. The IMR is added as an explanatory variable into each equation, in the example of AIDS model:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \ \ln p_j + \beta_i \ \ln\left(\frac{\mu}{P}\right) + \delta_i IMR_i. \tag{3.9}$$

This technique reminds the Heckman two-step estimator but Heckman's estimator is estimated on the subsample of consuming households only. Such system no longer satisfies the additivity as $\sum_i w_i \neq 0$. To preserve this condition, the authors proposed to define the last term of excluded equation as:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \ \ln p_j + \beta_i \ \ln\left(\frac{\mu}{P}\right) - \sum_i^{K-1} \delta_i IMR_i. \tag{3.10}$$

Thus this system can be estimated as usual QU/AIDS model with one excluded equation under the additivity condition of coefficients. The formulas to calculate different types of elasticities remain the same.

Although, we will see in the next subsection that this estimator was proved to be inconsistent, it is included in this thesis. Its performance is rather similar to other estimators in the case of small level of censoring. Moreover, most of the studies are still using this estimator. This frequent usage is the main motivation to include this estimator in our comparison as well.

Shonkwiler and Yen's estimator

Shonkwiler and Yen (1999) showed that the technique of Heien and Wessels (1990) is theoretically inconsistent and practically, its performance is poor according to Monte Carlo simulations for high levels of censoring. Therefore they have proposed their own two-stage estimator, which performance was appropriate. The true parameters in the Monte Carlo experiment were always included at least in the 95 % confidence interval of estimated coefficients. In general, this method supposes to estimate the probit model for each household and each food item at first, where the explanatory variables are denoted as $\boldsymbol{Z}$ and corresponding coefficients $\boldsymbol{\eta}$. In the second stage, the dependent variable y_{ih}, $i = 1, ..., K$, $h = 1, ..., N$ is reformulated as:

$$y_{ih} = \Phi(Z'_{ih}\hat{\eta}_i)f(X_{ih}\beta_i) + \delta_i\phi(Z'_{ih}\hat{\eta}_i) + \epsilon_{ih} \tag{3.11}$$

where X are the explanatory variables in the second stage together with their coefficients β, and δ is another coefficient to estimate. This second stage can be estimated by MLE or SUR. To be concrete, for the AIDS model, the second stage looks like (Ecker and Qaim, 2011):

$$w^*_{ih} = \Phi(Z'_{ih}\hat{\eta}_i)w_{ih} + \delta_i\phi(Z'_{ih}\hat{\eta}_i) + \epsilon_{ih} \tag{3.12}$$

where w_{ih} has already known form from equation (2.26) or equivalently (3.7). Newly, the w^*_i do not sum up to one. It means that the covariance-variance matrix of disturbances is no longer singular. Hence, the system should be estimated on all K equations and the additivity condition is no longer imposed. Moreover, Yen et al. (2002) gives the form of concentrated log-likelihood function:

$$L_N = \frac{NK}{2}\ln(2\pi + 1) - \frac{N}{2}\ln(|S|), \tag{3.13}$$

notation remains the same as in the equation (3.5).

The elasticities upon this modification of AIDS model can be estimated as:

- income elasticity: $e^i_M = 1 + \frac{1}{w^*_i}\Phi(Z'_{ih}\hat{\eta}_i)\beta_i$
- Marshallian price elasticity: $e^i_{p_j} = -\delta_{ij} + \frac{1}{w^*_i}\Phi(Z'_{ih}\hat{\eta}_i)(\gamma_{ij} - \beta_i\alpha_j - \beta_i\sum_k \gamma_{kj}\ \ln p_k)$

and for QUAIDS model (Ecker and Qaim, 2011):

- income elasticity: $e^i_M = 1 + \frac{1}{w^*_i}\Phi(Z'_{ih}\hat{\eta}_i)(\beta_i + 2\lambda_i\big(\prod_i p_i^{\beta_i}\big)^{-1}\ln(\frac{\mu}{P}))$
- Marshallian price elasticity: $e^i_{p_j} = -\delta_{ij} + \frac{1}{w^*_i}\Phi(Z'_{ih}\hat{\eta}_i)\Big\{ - \Big(\beta_i + 2\lambda_i\big(\prod_i p_i^{\beta_i}\big)^{-1} \ln(\frac{\mu}{P})\Big)\Big(\alpha_j + \sum_k \gamma_{kj}\ln p_k\Big) - \lambda_i\beta_j(\prod_i p_i^{\beta_i})^{-1}(\ln(\frac{\mu}{P}))^2 + \gamma_{ij}\Big\}$.

Cosslett's semi-parametric estimator

In contrast to Heckman's estimator, Cosslett (1991) does not pose the distributional assumption on the error terms in the first stage when the consumer decides whether to buy the goods or not. Instead of estimating probit model in the first stage, he proposes to estimate the distribution $F(u)$ non-parametrically. In this case, the Gallant and

Nychka (1987) approach is used. Then, according to Cosslett (1991), the participation equation (or the first stage) is given by:

$$E(y_i|Z_i, X_i; u_i > -Z_i\eta_i) = f(X_i\beta_i) + \Psi(Z_i'\eta_i), \tag{3.14}$$

in the same notation as in the subsection above, where the unknown function $\Psi(\cdot)$ is approximated by the set of dummy variables (equation (2.39)). It is also clear that (Shonkwiler and Yen, 1999):

$$E(y_i|Z_i, X_i; u_i \leq -Z_i'\eta_i) = 0. \tag{3.15}$$

Hence, it implies (Shonkwiler and Yen, 1999):

$$E(y_i|Z_i, X_i) = (1 - F(-Z_i'\eta_i)) \cdot (f(X_i\beta_i) + \Psi(Z_i'\eta_i)) + F(-Z_i'\eta_i) \cdot 0 \tag{3.16}$$

$$E(y_i|Z_i, X_i) = (1 - F(-Z_i'\eta_i)) \cdot (f(X_i\beta_i) + \Psi(Z_i'\eta_i)) \tag{3.17}$$

as

$$P[d_i = 1|Z_i] = 1 - P[u_i \leq -Z_i'\eta_i] = 1 - F(-Z_i'\eta_i) \tag{3.18}$$

where $F(\cdot)$ is the distribution of error terms in the first stage estimated by the Gallant and Nychka (1987) method. Following the same reasoning as in the previous subsection, the QU/AIDS model has form of

$$w_{ih}^* = (1 - F(-Z_{ih}'\hat{\eta}_i))(w_{ih} + \Psi(Z_{ih}'\hat{\eta}_i)) + \epsilon_{ih} \tag{3.19}$$

and can be estimated upon the whole set of equations maximizing the concentrated log-likelihood function (3.13). The formulas for elasticities remain the same as well, except that the $\Phi(Z_{ih}'\hat{\eta}_i)$ has to be substituted by $(1 - F(-Z_{ih}'\hat{\eta}_i))$.

3.1.3 Standard errors

To derive standard errors of estimated coefficients, the variance-covariance matrix has to be defined properly. In every case, the QU/AIDS model or censored QU/AIDS, the special attention has to be paid on the estimation of this matrix. For example, in case of Shonkwiler and Yen (1999) estimator, the error terms are heteroscedastic, thus the usual estimate of this matrix has to be adjusted, otherwise it would be incorrect. Moreover, the solutions are rather complex. To overcome all these different model-specific problems, the bootstrap method to estimate standard errors will be applied to all models. Specifically, the paired bootstrap method will be used. It means that the group of dependent and independent variables is fixed for every $h = 1, ..., N$. Then, the observations are resampled with recurrence to obtain new sample. This procedure is repeated in order to obtain B=500 different samples. For each of this samples b, the estimation procedure is conducted to get the values of coefficients, let say $\boldsymbol{\varphi}_b$. Then, the standard error of a coefficient φ_1 can be derived as:

$$SE(\hat{\varphi}_1) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\varphi}_{1,b} - \bar{\hat{\varphi}}_{1,b})^2}. \tag{3.20}$$

The bootstrap method to derive standard errors of estimated coefficients and of calculated elasticities have been also used by Brosig (1998) and Janský (2014).

3.1.4 Analysis

The aim of this empirical part is to find the version of AIDS model that suits the data best and to estimate through this model the demand for food of Czech households. Firstly, the shape of the Engel curves will be analysed. It means the test on the significance of the parameter λ_i in the QUAIDS model will be conducted. The linear version of AIDS (LA/AIDS) is not considered as it is only an approximation of original AIDS model. Secondly, the validity of consumer theory will be evaluated on the basis of comparison of unrestricted and restricted model by likelihood ratio test. Thirdly, the basic (QU)AIDS model will be compared to models with incorporated censoring treatment techniques. The final model will be estimated using socio-demographic variables (included by translation in the α_i coefficient such as in the equation (2.32)) and the results will be explored.

3.2 Description of data

3.2.1 Source of data

The data used in this work comes from the Czech Statistical Office (CSO) and for the purposes of this work it is provided by the Environment Center of Charles University in Prague. There are two main data sources. The budget survey data for 2013 which provides the descriptive information about households and expenditures information. The second source are the average consumer prices in the consumer basket for 2013.

The budget survey of the Czech households is collected every year by the CSO.[1] The goal of this data is to gain the very precise information about the economy of the households. By the household, it is meant an individual or a group of individuals which are living together and manage the household together or share the main expenditures.

The large variety of information is obtained from the households, for example, about

[1] All the scripts for the analysis were written by the author, however the routines have been run by Dr. Milan Ščasný who has signed the contract about the data usage including the confidentiality promise with the Czech Statistical Office.

type of housing (house or flat, sources of energies), about equipment (number of different appliances), about individuals (age, gender, occupation) and mainly about all expenditures and incomes flows, including monetary and barter (material) flows. The expenditures are separated in the classes according to Czech equivalent of the Classification of Individual Consumption by Purposes (COICOP) which was established in 1999. The brief overview of the categories is given in the following Table 3.1.

Table 3.1: Expenditure categories - COICOP

Label	Category
1	Food and non-alcoholic beverages
2	Alcoholic beverages, tobacco and narcotics
3	Clothing and footwear
4	Housing, water, energies, fuel
5	Equipment of household
6	Health
7	Transport
8	Post and telecommunications
9	Recreation and culture
10	Education
11	Accommodation and eating out
12	Other goods and services

ČSÚ (1999)

The importance and the usefulness of the budget survey data is very broad. It is due to the fact that the information about the differences of consumption of households arranged according to diverse aspects or about the influence of the different factors on the structure of expenditures and consumption habits cannot be obtained from any other data source. By the factor, it is meant the changes in prices or situation on the market, for instance. The budget surveys are so called primarily data as they cannot be gained from another statistics but on the other hand, they are source for many other purposes. The CSO mentions that the budget surveys serve as the basis for the qualified decision-making during the realization of the social state politics, for the social and economic research, for the intern purposes of the CSO (forming of the consumer basket to rebalance the index of consumer prices, for instance) and finally for the international comparisons.

The choice of the households is given by the quotas, the number of households is given according to certain aspect. In the case of the Czech budget survey, the main aspect is the economic activity of the principal of the household. The households with the economically active person in the head are divided according to the profession of this person. The household with economically inactive person in the head are divided according to the professions of the other members of the household. The economically active person is a person older than 15 years and is employed or unemployed and actively is looking for a job. The secondary aspect is the differentiation of the households according to their incomes.

The data is collected every year and approximatively 3000 households are included. The sample is chosen according to the aspects above in the way the structure of the sample is analogous to the structure of households in the Czech Republic. It is important to take into account that the sample represents only one thousandth of Czech households. It means that taking results of the analysis on such data as a conclusion for the whole population can be misrepresenting mainly in the case when we are interested in different income classes as the income is the quota variable and is based on the Microcensus.

The budget survey data will provide us the socio-demographic variables, the expenditure information and the income information. The problem arises with the price information. For every good or service in the CZ-COICOP categories the expenditures are collected. For some of them the quantity is collected as well. Upon these two information, the unit values can be calculated. These unit values are used in a part of the demand studies but it is important to bear in mind that these values reflect the price and quality information in the same time and it is not possible to separate them. On the other hand, the reflection of quality in the unit price can be an advantage too. As the data is incomplete in terms of quantity information, the unit values cannot be calculated in this case. Thus, an alternative source of data is used.

The CSO collects every month the average prices of selected commodities from all 12 CZ-COICOP categories which compose the consumer basket. These prices are weighted to create overall price level of the given period that serves further to compute consumer price indexes (CPI). There are monthly average prices of consumer goods for 2013 at our disposition together with the weights assigned to these prices. To create category- and household- specific price levels, the following procedure will be applied.

i. The monthly prices can be converted to the yearly prices by the general average according to CSO.

ii. The general price levels of different goods in budget survey categories and which expenditures are collected will be determined as an average composed by different consumer prices according to CSO's weights.

iii. To create price level of demand system's categories the price average weighted by the expenditures of given household will be used to obtain household-specific information for each food category in demand system.

iv. This procedure would result in zero price level in case of households which do not consume anything from the given food category. These households cannot be omitted from the estimations, thus some price level has to be assigned to them.

One possibility, which is often used and will be used in this case as well, is to assign to these households the average price levels. This approach was, for example, used in Yen et al. (2002). Another possibility is to estimate these price levels on the basis of socio-demographic information and other prices such as in Heien and Pompelli (1988).

There are $i = 1, ..., K$ food categories in the demand system. The data are available for $h = 1, ..., N$ households and for each household in the budget survey, the expenditures ex on the good j in the category i are collected. The weights and average prices from the consumer basket are available for good k in the subcategory j. Hence, the (ii) is equivalent to $p_{ij} = \frac{\sum_k p_{ijk} w_k}{\sum_k w_k}$ and (iii) is equivalent to $p_{ih} = \frac{\sum_j p_{ij} ex_j}{\sum_j ex_j}$.

3.2.2 Descriptive statistics

The goal is to evaluate the food demand of Czech households. All the food and non-alcoholic beverages expenditures are summarized in the first CZ-COICOP category. This category is further divided in the following subcategories which will create the categories in our QU/AIDS model. These food groups are provided in the next Table 3.2 and their descriptive statistics are in the Table 3.3. There are $N = 2903$ observations in total.

Table 3.2: Food and non-alcoholic beverages subcategories

Label	Category	Examples
1	Bakery products and cereals	Bread, Cookies, Flour, Pasta
2	Meat	Pork, Beef, Meat cans, Sausages
3	Fish	Fish, Seafood, Fish products
4	Dairy and eggs	Milk, Egg products, Yoghurt
5	Oils and fats	Butter, Fat, Vegetable oils
6	Fruit	
7	Vegetables	
8	Sugar and confectionery	Jam, Honey, Chocolate, Ice Cream
9	Food products and other food	Soup, Sauce, Spices
10	Coffee, tea, and cocoa	
11	Water, soft drink, and juices	

ČSÚ (1999)

From the Table 3.3., it can be seen that on average the most important products in the diet of Czech households are meat and meat products, dairy and eggs, and bakery products with respect to food expenditures. There exist also households which are spending half of food expenditures on meat or on dairy. There are households which food basket is mainly composed by vegetables as well. The variability in the prices is more extended. It is not surprising as the prices are household-specific and each household can spend the money for different products in the given food subcategory.

Table 3.3: Descriptive statistics of shares w_i and prices p_i

	Share w_i [in %]		Price p_i [in CZK]			Zero observation	
Category	Mean	Max	Mean	Min	Max	Number	Percentage
1 Bakery products	16.62	47.63	62.13	25.69	102.99	0	0
2 Meat	24.41	62.03	130.05	100.80	201.36	1	0.03
3 Fish	2.58	26.49	229.18	225.11	231.39	**316**	**10.89**
4 Dairy and eggs	18.19	52.16	88.95	15.19	192.11	0	0
5 Oils and Fats	4.87	20.94	114.96	46.13	274.69	8	0.28
6 Fruit	6.15	31.49	51.34	31.59	85.75	14	0.48
7 Vegetables	8.28	40.37	37.42	17.00	49.37	10	0.34
8 Sugars	6.47	27.14	57.48	15.72	149.88	18	0.62
9 Other food	3.34	19.30	63.97	21.85	94.58	29	1.00
10 Coffee, tea, and cocoa	4.15	33.55	56.55	32.42	60.26	68	2.34
11 Non-alcoholic drinks	4.95	29.41	19.89	8.25	41.89	93	3.20

The observations have been checked for possible outliers and none of the prices is aside from the rest of the sample.

Concerning the censoring of the data, the fish subcategory is an example of the fact that even the data aggregated on higher level can contain the significant part of zero observations. In this case, there are 10.89 % of them. For other subcategories the percentage is rather small or zero. There are two food groups (bakery products and dairy and eggs) which are not censored. Basically, it means that even though we would use the censored techniques for these two equations it will result in the original form of the QU/AIDS model. For example, for Shonkwiler and Yen's estimator, the probit would yield in $\Phi(Z'_{ih}\hat{\eta}_i) = 1$ and $\phi(Z'_{ih}\hat{\eta}_i) = 0$ for every observation as the variable in the probit is not dichotomous. Next, for the second category of meat and meat products, there is only 1 zero observation. This small (basically zero) variation in probit dependent variable leads us to not to treat this food group as censored as well. It can be also seen from the Table 3.3. that there is not any household in this dataset, which would by strictly vegan and there is possibly one vegetarian household which do not buy neither any meat products nor the fish products but their expenditures are mainly spent on the dairy products and eggs (40 %).

The socio-demographic variables will be used in two cases. Once in the first-step of censored demand models' estimators and secondly in the final QU/AIDS model extended by these variables included by the development of coefficients α_i. The commonly used socio-demographic information in the literature for both these purposes are: household size, age and gender composition of the household, if the woman is responsible for the shopping, education or type of employment of the family head, geographical position of living (urban vs rural, region of the country), presence of children, or city size. The following Table 3.4. gives the description of the variables which will be included in the analysis.

Table 3.4: Description of socio-demographic variables

Variable	Description
$purchase_i$	dummy variable equal to 1 if the expenditures for the subcategory i are bigger than 0, equivalently when $w_i > 0$
$size$	number of members living in the household
age	age of the principal of the housholds which is taken to approximate the age category of the household
$woman$	dummy variable equal to 1 if there is a woman in the household
$d\_retired$	dummy variable equal to 1 if there is a retired person in the household
$d\_children$	dummy variable equal to 1 if there is at least one child in the household
$education$[1]	discrete variable which takes values between 0-9, it represents the maximal education level gained by the principal of the household or its partner 0 stands for no education and 9 stands for postgraduate or higher education level
$primary$	dummy variable equal to 1 if the $education$ is on the no education or primary level
$secondary$	dummy variable equal to 1 if the $education$ is on the secondary level
$tertiary$	dummy variable equal to 1 if the $education$ is on the tertiary level
$natu\_i$	dummy variable equal to 1 if the household consume something from the subcategory i which has not been purchased but obtained from the kept animals or as a gift, for instance
$income\_cap$	yearly income of the household in the Czech crowns divided by the number of household members
d_5	number of children less then 5 years old
$village$	dummy variable if the household live in the village

[1] It would be more appropriate to use dummy variables for different levels of education but for the purpose of the probit model this definition was chosen in order to be estimable without problems the large number of dummies can represent.

The probit model of the Heien and Wessels' estimator and Shonkwiler and Yen's estimator will take form of the following equation:

$$purchase_i = \eta_{0i} + \eta_{1i} size + \eta_{2i} age + \eta_{3i} woman + \eta_{i4} d\_children \tag{3.21}$$

$$+\eta_{5i} d\_retired + \eta_{6i} education + \eta_{7i} natu_i + \eta_{8i} income\_cap + u.$$

The same vector of variables will be also used in the first stage of the Cosslett's semi-parametric estimator. According to Newman et al. (2001) mainly the qualitative variables influence the decision about the purchase. Further, mainly the financial factors influence the decision about the quantities purchased and also about the chosen price of product partially. The previously named variables have been chosen because of the following potential effects on the purchase decision:

- $size$: the bigger the household and more different members of household, the more possibly different needs or preferences among food and beverages items
- age : during the lifetime, the different diet requirements or preferences appear
- $woman$: the presence of a woman in the household can also influence the diet requirements especially when having children, also the different preferences can

be possibly distinguished between women and men

- $d\_children, d\_retired$: the presence of children or retired members may lead to different diet requirements as well
- $education$: the education level can have divers effects, it is possible that more educated people think more about their diet and health, they can also be in better financial situation and thus be able to purchase different products, on the other hand they can be also so busy to not to think about their diet
- $natu\_i$: this variable means that the household receives and consumes the item from given subcategory without buying it which influences the further purchase of this item
- $income\_cap$: the income per one household member describes the financial situation which can influence the decision about purchase as well.

The information on the natural consumption is available only for limited number of products in 5 food subcategories: meat, dairy and eggs, oils and fats, fruit, and vegetables. As the probit model (neither the Cosslett's first stage) will not be estimated for forth equation (dairy and eggs), this variable will be used in 4 cases only.

In the second step of the censored estimators or in the QU/AIDS model in general, the financial factors are included in the form of a set of prices and total expenditures. Different socio-demographic variables will be included that may influence the quantity of purchased items in the given subcategory or its expenditures:

- $size$: more household members usually means bigger consumption of food
- age : also the quantity consumed changes during the lifetime, moreover the incomes can be associated with the age as well
- $primary$, $secondary$, $tertiary$: the education level can be also linked with incomes and influence the composition of the diet
- d_5 : the number of children beside the specific composition of the diet influences the quantity purchased as younger children consume less food
- $village$: the location out of the big cities can influence the availability of food, their price and also the consumer habits and the barter flows are more common in the small villages.

Hence, the coefficient α_i will be extended as follows:

$$\alpha_i = \alpha_{0i} + \alpha_{1i} size + \alpha_{2i} age + \alpha_{3i} primary + \alpha_{4i} tertiary + \alpha_{5i} d\_5 + \alpha_{6i} village. \quad (3.22)$$

The variable of $secondary$ education is omitted because of dummy variable trap.

3.3 Results

3.3.1 QU/AIDS model and consistency with economic theory

In this subsection, both AIDS and QUAIDS models are estimated under unrestricted and restricted conditions to see the consistency of the data with the economic theory. Next, the quadratic shape of Engel curves is tested upon QUAIDS model.

The economic theory imposes three desirable properties - additivity, homogeneity and symmetry (Slutsky condition). In the AIDS model the additivity is fulfilled by $\sum_i \alpha_i = 1$, $\sum_i \gamma_{ij} = 0$ for $\forall j$ and $\sum_i \beta_i = 0$. The additivity is in the QUAIDS model augmented by $\sum_i \lambda_i = 0$. The additivity conditions are automatically incorporated in the estimation constraints as the coefficients of excluded equation are estimated through these equalities. Contrary, the homogeneity condition $\sum_j \gamma_{ij} = 0$ for $\forall i$ and the Slutsky condition $\gamma_{ij} = \gamma_{ji}$ do not have to be necessarily imposed as an constraint of AIDS or QUAIDS estimates. In this case, the unrestricted model means that only the additivity condition is fulfilled, the restricted model supposes all three conditions fulfilled. In total, four models are estimated in this section, unrestricted and restricted AIDS and unrestricted and restricted QUAIDS.

The estimates of restricted and unrestricted AIDS are in the Appendix, in Table A.1 and A.2, respectively. The likelihood ratio (LR) test is conducted to compare them. The LR statistic $LR = 2(logL_r - logL_u)$ has the χ^2_h distribution where $logL_r$ is the value of log-likelihood of restricted model and $logL_u$ of unrestricted model, h is number of restrictions. The null hypothesis represents restricted model. The same test is also conducted on the QUAIDS model, its constrained and unconstrained estimate. The results of the test are given in the next Table 3.5.

Table 3.5: Likelihood ratio test for restricted vs. unrestricted QU/AIDS models

	AIDS	QUAIDS
H_0:	restricted	restricted
H_1:	unrestricted	unrestricted
LR	900.11	1032.1
χ^2_{55}	73.29	73.29

The null hypotheses H_0 are rejected in favour of alternative hypotheses in both cases. Thus, for the Czech data in case of food demand the homogeneity and symmetry conditions do not hold. The unrestricted estimates of model's coefficients are more appropriate.

The results of restricted QUAIDS model can be found in the Appendix Table A.3. In the next Table 3.6. the results of unrestricted QUAIDS model are given. The coefficient λ of quadratic term is significant in 10 out of 11 equations (except Fruit). Moreover,

Table 3.6: Estimates of unrestricted QUAIDS

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.0031 ***	0.0025***	0.0007***	0.0024***	0.0026***	0.0010***
	0.0002	0.0001	0.0002	0.0002	0.0002	0.0001
β	0.0093***	0.0091***	0.0029***	0.0086***	0.0081***	0.0035***
	0.0005	0.0007	0.0008	0.0008	0.0009	0.0005
γ_1	0.0135***	-0.0035**	-0.0088***	0.0005	-0.0264***	-0.0001
	0.0013	0.0015	0.0017	0.0008	0.0044	0.0013
γ_2	0.0099***	0.0077***	0.0003	0.0087***	0.0074***	0.0013
	0.0006	0.0007	0.0006	0.0006	0.0013	0.0008
γ_3	0.0151***	0.0122***	0.0003	0.0119***	0.0120***	0.0038***
	0.0007	0.0007	0.0011	0.0007	0.0011	0.0006
γ_4	0.0015***	-0.0007	0.0045	-0.0004	-0.0117***	-0.0012
	0.0016	0.0007	0.0028	0.0008	0.0013	0.0016
γ_5	-0.0034***	-0.0084***	0.0082***	0.0012	0.0178***	-0.0017
	0.0008	0.0012	0.0015	0.0012	0.0024	0.0022
γ_6	-0.0044***	-0.0006	-0.0063***	-0.0012	0.0016	0.0137***
	0.0013	0.0006	0.0012	0.0012	0.0014	0.0023
γ_7	0.0038***	-0.0019*	-0.0002	0.0046***	-0.0046***	0.0037***
	0.0010	0.0010	0.0010	0.0007	0.0014	0.0013
γ_8	-0.0129***	-0.0030***	0.0047***	-0.0061***	-0.0009	-0.0019
	0.0019	0.0011	0.0016	0.0012	0.0010	0.0012
γ_9	-0.0009	0.0001	-0.0014	0.0047***	0.0080***	-0.0022
	0.0013	0.0006	0.0014	0.0013	0.0015	0.0025
γ_{10}	0.0086***	0.0079***	-0.0009	0.0047***	0.0053***	0.0016
	0.0004	0.0007	0.0017	0.0007	0.0012	0.0013
γ_{11}	0.0049***	0.0076***	0.0013	0.0111***	-0.0078***	-0.0062**
	0.0012	0.0011	0.0014	0.0011	0.0022	0.0028
λ	-0.0014***	0.0017***	-0.0002**	-0.0010***	-0.0010***	-0.0003
	0.0002	0.0003	0.0001	0.0003	0.0001	0.0003

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.0021***	0.0011***	0.0012***	0.0022***	0.9810***
	0.0002	0.0002	0.0001	0.0002	0.0008
β	0.0066***	0.0041***	0.0045***	0.0071***	-0.0638***
	0.0005	0.0007	0.0004	0.0008	0.0046
γ_1	-0.0005	0.0049***	-0.0034**	-0.0018*	0.0256***
	0.0011	0.0009	0.0015	0.0019	0.0098
γ_2	0.0049***	0.0024***	0.0010	0.0066***	-0.0500***
	0.0009	0.0006	0.0011	0.0007	0.0029
γ_3	0.0095***	0.0044***	0.0052***	0.0105***	-0.0850***
	0.0010	0.0008	0.0005	0.0010	0.0036
γ_4	0.0077***	-0.0066***	0.0035*	-0.0086***	0.0120*
	0.0019	0.0014	0.0019	0.0017	0.0069
γ_5	0.0021	-0.0090***	-0.0070***	0.0009	-0.0008
	0.0020	0.0015	0.0021	0.0017	0.0048
γ_6	-0.0003	-0.0015	-0.0022	-0.0016	0.0027
	0.0013	0.0009	0.0016	0.0018	0.0051
γ_7	-0.0174***	-0.0041***	0.0050***	0.0008	0.0104
	0.0028	0.0010	0.0016	0.0008	0.0070
γ_8	0.0005	0.0255***	0.0021	-0.0019	-0.0061***
	0.0014	0.0025	0.0013	0.0012	0.0019
γ_9	-0.0039***	-0.0032*	-0.0028	-0.0035 **	0.0052
	0.0014	0.0018	0.0026	0.0017	0.0049
γ_{10}	0.0047***	-0.0015	0.0018***	0.0095***	-0.0418***
	0.0015	0.0010	0.0006	0.0013	0.0046
γ_{11}	0.0063**	0.0019*	0.0014	-0.0059***	-0.0144***
	0.0026	0.0011	0.0020	0.0019	0.0031
λ	-0.0008***	-0.0003**	-0.0004***	-0.0010***	0.0048***
	0.0003	0.0002	0.0001	0.0001	0.0003
logL	58872.28				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

the λ's are significant on 5 % or rather 1% level. It means that in the case of Czech data, the relationship between income and expenditures is not linear.

From the above tests it follows that the unrestricted QUAIDS is the suitable model to evaluate the food demand for Czech households. Further, we can see that most of the coefficients are significant. To explore the information the coefficients bring, the income, own-price, and cross-price elasticities are calculated and are given in the Table 3.7. and 3.8.

Concerning income elasticity, the mean and median values are close to each other, there are not any suspicious estimates. For most of the subcategories, the elasticity lies between 0 and 1 signifying necessary goods. Meat and fish products and non-alcoholic beverages have income elasticity higher than 1 and thus are considered as luxuries and their demand is elastic in income.

The values of own-price elasticities are more widely spread. For example, for group of sugars the mean and median are considerably different and the standard deviation is also high. The negative elasticity for each group of products means that there are no Giffen goods. The demand is elastic for dairy&eggs, vegetables, non-alcoholic beverages and others. For vegetables, this is not surprising as there are many different kinds of vegetables and whose prices differ during the year. On the other hand, the non-alcoholic beverages are not necessary to be purchased so it is expectable that the demand for them will be elastic.

Table 3.7: Income and own-price Marshallian elasticities upon unrestricted QUAIDS

	Income elasticities			Own price elasticities		
	Mean	Median	Std dev	Mean	Median	Std dev
Bakery	0.9316	0.9357	0.0266	-0.9096	-0.9153	0.0318
Meat	1.1735	1.1429	0.2039	-0.9781	-0.9824	0.0251
Fish	1.0332	1.0188	0.0448	-0.9759	-0.9864	0.0348
Dairy&eggs	0.9604	0.9631	0.0222	-1.0002	-1.0001	0.0004
Fats	0.8088	0.8467	0.1910	-0.4908	-0.5976	0.5083
Fruit	0.9703	0.9771	0.0290	-0.6711	-0.7489	0.3340
Vegetables	0.9137	0.9313	0.0835	-1.2808	-1.2188	0.2870
Sugars	0.9871	0.9904	0.0181	-0.3871	-0.5621	0.7702
Others	0.9566	0.9681	0.0618	-1.1303	-1.0900	0.2104
Coffee etc.	0.6991	0.7843	0.3222	-0.6128	-0.7229	0.4086
Non-alco	1.1758	1.1146	0.4941	-1.1072	-1.0697	0.2725

The cross-price elasticities differentiate the substitutes and complements between goods. Fish and meat products are substitutes for most of the goods except non-alcoholic beverages. It is important to say that the standard deviations for cross-price elasticities of non-alcoholic beverages are rather high which means that the attitudes across households are very various. The same is true for subcategory others which is not surprising as this group contains very different goods. The bakery products are

Table 3.8: Cross-price Marshallian elasticities upon unrestricted QUAIDS

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
Bakery	-	0.0752	0.1101	0.0132	-0.0171	-0.0284
	-	0.0269	0.0394	0.0046	0.0060	0.0100
Meat	-0.0117	-	0.0400	-0.0047	-0.0499	-0.0025
	0.0109	-	0.0449	0.0043	0.0622	0.0031
Fish	-0.6575	0.0239	-	0.3436	0.6230	-0.4706
	0.9277	0.0345	-	0.4852	0.8796	0.6640
Dairy&eggs	0.0039	0.0596	0.0790	-	0.0111	-0.0066
	0.0020	0.0319	0.0423	-	0.0059	0.0036
Fats	-0.7297	0.2328	0.3594	-0.3131	-	0.0468
	0.7288	0.2327	0.3595	0.3131	-	0.0469
Fruit	-0.0005	0.0362	0.0964	-0.0252	-0.0371	-
	0.0007	0.0362	0.0973	0.0258	0.0380	-
Vegetables	-0.0064	0.0935	0.1677	0.1315	0.0421	-0.0038
	0.0072	0.0953	0.1715	0.1338	0.0425	0.0042
Sugars	0.1212	0.0622	0.1086	-0.1556	-0.2113	-0.0336
	0.1523	0.0778	0.1361	0.1956	0.2657	0.0423
Others	-0.1540	0.0601	0.2537	0.1754	-0.3206	-0.1010
	0.2485	0.0948	0.4068	0.2817	0.5178	0.1630
Coffee etc.	-0.0668	0.2861	0.4335	-0.3089	0.0549	-0.0547
	0.0711	0.3027	0.4588	0.3251	0.0577	0.0576
Non-alco	0.8664	-1.9400	-3.1747	0.3618	-0.1092	0.0674
	1.4655	3.2985	5.3879	0.6077	0.1943	0.1130

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
Bakery	0.0280	-0.0833	-0.0040	0.0650	0.0701
	0.0099	0.0293	0.0015	0.0232	0.0256
Meat	-0.0118	-0.0184	-0.0020	0.0267	-0.1056
	0.0131	0.0217	0.0027	0.0307	0.1250
Fish	-0.0128	0.3596	-0.1065	-0.0617	0.0176
	0.0178	0.5077	0.1500	0.0865	0.0293
Dairy&eggs	0.0303	-0.0358	0.0302	0.0339	0.0862
	0.0158	0.0189	0.0159	0.0183	0.0456
Fats	-0.1173	-0.0174	0.2297	0.1687	-0.1231
	0.1177	0.0177	0.2291	0.1686	0.1273
Fruit	0.0911	-0.0441	-0.0499	0.0429	-0.1458
	0.0925	0.0450	0.0508	0.0431	0.1502
Vegetables	-	0.0121	-0.0610	0.0872	0.1433
	-	0.0124	0.0626	0.0889	0.1428
Sugars	-0.0967	-	-0.0744	-0.0323	0.0312
	0.1216	-	0.0936	0.0409	0.0385
Others	0.2429	0.1019	-	0.0928	0.0483
	0.3905	0.1636	-	0.1479	0.0710
Coffee etc.	0.0413	-0.0593	-0.1196	-	-0.0534
	0.0435	0.0622	0.1259	-	0.0622
Non-alco	0.3212	-0.2649	0.1499	-1.6139	-
	0.5399	0.4542	0.2503	2.7433	-

Mean values of cross-price elasticities with standard deviations below complements to almost all food groups except dairy&eggs and sugars. Fruit are complements to all food groups except fats and oils, it can be expected that there is not big relation between these two food categories. It can be also observed that most of the effects are not symmetric.

3.3.2 Censored estimators of the QUAIDS

In this section, the unrestricted QUAIDS is estimated with incorporated treatment techniques for censored data. Namely, the Heien and Wessels's, Shonkwiler and Yen's, and Cosslett's estimators are applied.

Probit model

The first step of Heien and Wessels's estimator and Shonkwiler and Yen's estimator is the estimation of probit model. For both estimators, the probit model is proposed as in equation (3.21). The results are given in the Table 3.9.

Table 3.9: Probit model of Heien and Wessels' and Shonkwiler and Yen's estimator

	Fish pur_3	Fats pur_5	Fruit pur_6	Veget pur_7	Sugars pur_8	Others pur_9	Coffee etc. pur_{10}	Non-al pur_{11}
η_0	0.0092	-0.1323	0.6603	-0.1272	1.4601***	0.3210	0.5132 *	1.0436***
	(0.2106)	(0.7505)	(0.7669)	(1.0782)	(0.5371)	(0.4248)	(0.2907)	(0.3343)
$size$	0.3208***	0.6209*	-0.02519	0.9110**	0.5163***	0.6688***	0.3391***	0.6541***
	(0.0464)	(0.3264)	(0.1260)	(0.3991)	(0.1652)	(0.1574)	(0.0859)	(0.1000)
age	0.0089***	0.0248**	-0.0119	-0.0005	0.0277***	0.0182***	0.0138***	-0.0092*
	(0.0031)	(0.0115)	(0.0110)	(0.0134)	(0.0087)	(0.0070)	(0.0048)	(0.0049)
$woman$	0.0501	0.4287	0.5202	0.6063*	-0.1328	0.4019**	0.4926***	0.0043
	(0.1197)	(0.3188)	(0.3350)	(0.3120)	(0.3091)	(0.1967)	(0.1572)	(0.1561)
$d\_children$	-0.2754**	-0.5490	-0.0404	-1.4151**	-0.4926	-0.7710***	-0.4321**	-0.3402*
	(0.1130)	(0.5273)	(0.3581)	(0.6582)	(0.3323)	(0.2781)	(0.1879)	(0.1968)
$d\_retired$	-0.0792	-0.1602	0.6628*	-0.9909**	-0.8831***	-0.2732	-0.2112	0.1182
	(0.1026)	(0.4928)	(0.3903)	(0.5024)	(0.3127)	(0.2601)	(0.1707)	(0.1568)
$education$	-0.0122	0.1110	0.3500**	0.3000*	-0.1476***	-0.0008	-0.0229	-0.0087
	(0.0198)	(0.1211)	(0.1506)	(0.1787)	(0.0486)	(0.0486)	(0.0319)	(0.0335)
$incomecap$	0.1420***	0.0589	0.6048*	0.6437	-0.0318	-0.0919	-0.0275	0.1272
	(0.0504)	(0.2434)	(0.3176)	(0.4753)	(0.0983)	(0.0639)	(0.0583)	(0.0904)
$natu5$		-0.0981						
		(0.3800)						
$natu6$			0.0798					
			(0.2085)					
$natu7$				omitted				
N	2903	2903	2903	2903	2903	2903	2903	2903
logL	962.53	44.15	76.41	46.13	93.76	136.39	297.16	359.37

Standard errors in parentheses
Significance of coefficients' estimates: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Generally, the purchase decision is dependent in the most of the cases on the size of household, on the age category approximated here by the age of the head of the household and on the presence of children. The other factors are relevant in few cases only, according to the probit's results.

The purchase of fishes and fish products is also influenced by the income of household per member. The prices of fishes are usually high comparing to other food, thus the income can be a relevant factor in this purchase decision. For subcategory of fats and oils, the presence of children is no longer significant. These products are not supposed to be the basic part of the children's diet, moreover they are mostly used to produce other meals (with the exception of butter). In the probit estimates for fruit, only the presence of retired people, education and income are significant, which is different from other probit models. The prices of fruit vary across different types and are seasonally influenced. The presence of retired people is also significant in the probit model for vegetables together with size, presence of woman, children and education. The significance of presence of retired people for fruit and vegetables can be linked to special diets when fats, oils and meat should be replaced by more light meals. Also, the diabetes is very common disease among elder people so the purchase of sugars can be influenced by their presence as well. The presence of woman is significant in probit for other food. Subcategory of others includes the spices, salt, soups, sauces, and utensils for cooking and baking. The explanation can be that it is usually the role of woman who is cooking in the household for what all these products are needed. In the case of coffee, tea and cacao, the variables of size, age, presence of woman and children are significant. For non-alcoholic beverages, only the size, age and presence of children is significant. The children are usually interested in different sweet soft drinks like Coke that are also included in this subcategory.

The natural purchase has been omitted in the seventh equation because of collinearity. In two remaining probits this variable is not significant. It is possible that these variables would be significant for concrete products. For example, for the group of fats and oils, only the information of donation of pork fat is available. The group itself is more broad so the natural income of pork fat does not strongly influence the purchase of other oils or butter as these commodities are not strict substitutes.

Heien and Wessels' estimator

The estimates of coefficients are given in the Appendix Table A.4. As in the previous non-treated QUAIDS model, all the α's, β's and λ's are significant. Hence, the QUAIDS model is still preferable for this estimator as coefficients at the quadratic terms are significant. Next, all the δ's, coefficients at inverse Mills ratio term, are significant as well. Regarding the γ's coefficients, they are mostly significant. Moreover most of all coefficients are significant at 1% level. The results of estimated elasticities are more important and are given in the next Tables 3.10. and 3.11.

Concerning the income elasticities, all of them are positive signifying that there are no inferior goods. Mean and median values are relatively close to each other except for coffee etc. and non-alcoholic beverages whose standard deviations are also high. According to this estimator meat and meat products and non-alcoholic beverages are luxury goods with elastic demand in income. Here the censoring is controlled. The data is censored mainly for subcategory of fish (10 % of zeros) and the income elasticity of fish has changed determining fish as necessary good instead of luxury good.

In the case of own-price elasticities the mean and median values differ more than in previous cases. For non-alcoholic beverages the sign between mean and median differs as well. Also the standard deviation of this elasticity is very high. Considering the median values, all elasticities are negative meaning that there are no Giffen goods. The demand is elastic in own price for subcategory of fish, dairy&eggs, and vegetables. This is different from previously estimated QUAIDS without control of censoring where demand was elastic in own-price for dairy&eggs, vegetables, others, and non-alcoholic beverages. The data for dairy&eggs is uncensored and for vegetables there are only 0.4 % of zeros. The change of elastic character is observed for fish (10 % of zeros), others (1 % of zeros), and non-alcoholic beverages (3.2 % of zeros). On the other hand, the variability of elasticities across households is higher than for the previous estimator.

Table 3.10: Income and own-price Marshallian elasticities upon unrestricted QUAIDS estimated by Heien and Wessel's estimator

	Income elasticities			Own price elasticities		
	Mean	Median	Std dev	Mean	Median	Std dev
Bakery	0.9172	0.9219	0.0324	-0.7067	-0.7253	0.1034
Meat	1.1905	1.1560	0.2267	-0.9651	-0.9717	0.0415
Fish	0.8056	0.8936	0.3164	-1.3196	-1.1837	0.4446
Dairy&eggs	0.9796	0.9811	0.0131	-1.0212	-1.0198	0.0112
Fats	0.7783	0.8217	0.2239	-0.5216	-0.6217	0.4769
Fruit	0.8288	0.8695	0.1678	-0.1301	-0.3357	0.8822
Vegetables	0.9343	0.9465	0.0634	-1.9322	-1.7268	0.9527
Sugars	0.8955	0.9230	0.1289	-0.4178	-0.5840	0.7301
Others	0.9839	0.9837	0.0632	-0.9791	-0.9852	0.0308
Coffee etc.	0.5640	0.6863	0.4644	-0.4680	-0.6184	0.5592
Non-alco	1.5855	1.3696	1.2530	0.1168	-0.3189	1.9622

Regarding cross-price elasticities, the standard deviations of estimates are also high. They are often of the same order as median values, which are reported in the table. Considering the censored groups, the fish group has became complementary to fruit, non-alcoholic beverages are complementary to all food categories except bakery products and dairy&eggs. In this case, the bakery products are no more complementary to the most of the food groups, the same is also true for fruit category. These are the most remarkable changes between these two different estimators.

Table 3.11: Cross-price Marshallian elasticities upon unrestricted QUAIDS estimated by Heien and Wessel's estimator

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
Bakery	-	0.0723	0.1435	-0.0043	-0.0663	-0.1253
	-	0.0274	0.0545	0.0023	0.0251	0.0472
Meat	-0.0481	-	0.0608	-0.0082	-0.1088	-0.0129
	0.0659	-	0.0857	0.0106	0.1636	0.0202
Fish	-0.4148	0.0307	-	0.2449	0.5059	-0.1318
	1.0232	0.0855	-	0.6093	1.2534	0.3257
Dairy&eggs	0.0039	0.0541	0.0757	-	0.0270	-0.0403
	0.0023	0.0302	0.0428	-	0.0150	0.0228
Fats	-1.0214	0.2895	0.5596	-0.1907	-	0.0811
	1.2920	0.3622	0.7052	0.2409	-	0.1030
Fruit	0.0715	-0.1695	-0.1340	0.0421	0.0685	-
	0.0937	0.2279	0.1830	0.0555	0.0902	-
Vegetables	0.0029	0.0342	0.1795	0.2984	0.1603	0.0630
	0.0035	0.0414	0.2301	0.3881	0.2059	0.0814
Sugars	0.3887	0.0120	0.0354	-0.1409	-0.1426	0.0565
	0.6825	0.0200	0.0598	0.2492	0.2532	0.0987
Others	0.0796	-0.1753	0.0408	0.0715	-0.0427	-0.0491
	0.1887	0.4209	0.0850	0.1622	0.1078	0.1164
Coffee etc.	-0.1209	0.1577	0.3589	-0.3045	0.1974	-0.0809
	0.1778	0.2307	0.5301	0.4493	0.2889	0.1185
Non-alco	0.0694	-0.8200	-2.1723	0.0168	-0.2525	-0.2407
	0.2210	2.1934	5.9128	0.1001	0.6534	0.6547

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
Bakery	0.0526	-0.0913	-0.0307	0.0483	0.0157
	0.0199	0.0345	0.0116	0.0182	0.0059
Meat	-0.0376	-0.0250	0.0003	0.0483	-0.0653
	0.0510	0.0355	0.0019	0.0727	0.0989
Fish	0.1471	0.1212	0.0879	-0.1769	-0.1634
	0.3688	0.3036	0.2219	0.4325	0.4062
Dairy&eggs	0.0729	-0.0658	0.0632	-0.0220	0.0605
	0.0405	0.0373	0.0353	0.0125	0.0337
Fats	-0.2092	0.0159	0.1633	0.0709	-0.1929
	0.2673	0.0200	0.2039	0.0883	0.2435
Fruit	0.1972	0.0616	-0.0291	-0.1475	-0.0626
	0.2629	0.0806	0.0396	0.1964	0.0823
Vegetables	-	0.0530	-0.0654	0.0277	-0.0155
	-	0.0673	0.0877	0.0346	0.0218
Sugars	-0.0718	-	0.0285	-0.2287	-0.0785
	0.1276	-	0.0483	0.4031	0.1374
Others	0.3844	0.0636	-	-0.1006	-0.0980
	0.8893	0.1449	-	0.2392	0.2297
Coffee etc.	0.0505	0.0265	-0.0579	-	-0.0725
	0.0746	0.0394	0.0863	-	0.1067
Non-alco	0.6916	-0.1418	-0.2022	-0.1508	-
	1.9544	0.3731	0.5277	0.3910	-

Median values of cross-price elasticities with standard deviations below

Shonkwiler and Yen's estimator

The estimates of coefficients of this censored technique are given in the Appendix Table A.5. All the α's coefficients are significant on 1 % level. Comparing to non-treated QUAIDS model, only six β's are significant. Further, only three λ's at quadratic terms are significant. On the other hand, all the δ's coefficients specific for this estimator are significant at 1 % level.

Table 3.12: Income and own-price Marshallian elasticities upon unrestricted QUAIDS estimated by Shonkwiler and Yen's estimator

	Income elasticities			Own price elasticities		
	Mean	Median	Std dev	Mean	Median	Std dev
Bakery	0.9374	0.9409	0.0228	-0.8109	-0.8230	0.0667
Meat	0.9977	0.9967	0.0112	-0.9542	-0.9625	0.0541
Fish	0.4375	0.6801	0.7980	-0.1984	-0.5491	1.1509
Dairy&eggs	0.9521	0.9554	0.0251	-0.9985	-0.9986	0.0009
Fats	0.3897	0.5129	0.6034	-0.5275	-0.6271	0.4705
Fruit	0.9629	0.9714	0.0366	-0.4446	-0.5750	0.5619
Vegetables	1.1099	1.0851	0.1158	-1.5593	-1.4365	0.5704
Sugars	1.1763	1.1269	0.2206	-0.5590	-0.6852	0.5547
Others	0.3151	0.5153	1.0112	-0.9148	-0.9390	0.1232
Coffee etc.	1.5325	1.3846	0.5611	-0.6844	-0.7746	0.3324
Non-alco	2.2156	1.7587	2.0463	-1.5288	-1.3282	0.9059

In the Table 3.12. the income elasticities together with own-price elasticities are given. The income elasticities are all positive. The mean and median values are considerably different for fish, oils and fats, others, coffee etc., and non-alcoholic beverages subcategories. The luxury goods with elastic demand are vegetables, sugars, coffee etc., and non-alcoholic beverages. This is different from previous estimators where demand was elastic for meat and non-alcoholic beverages in case of Heien and Wessels' estimator and it was elastic for meat, fish and non-alcoholic beverages in case of non-treated QUAIDS.

The own-price elasticities are all negative meaning that all these goods are normal. The standard deviations show that the attitudes towards some food groups are highly varying across households, mainly towards fish products and non-alcoholic beverages. The demand is elastic in own-price for vegetables and non-alcoholic beverages as well. This is also different form Heien and Wessels where the demand was elastic for fish, dairy&eggs, and vegetables. It differs also from non-treated QUAIDS where demand was elastic for dairy&eggs, vegetables, and others.

The median values of cross-price elasticities are given in the Table 3.13. As in the previous case, the magnitude of standard deviations shows the broad variety in the attitudes across households towards food products. Exploiting the signs of estimates, this estimator is closer to the Heien and Wessels' estimator than to the non-treated

Table 3.13: Cross-price Marshallian elasticities upon unrestricted QUAIDS estimated by Shonkwiler and Yen's estimator

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
Bakery	-	0.0504	0.1279	0.0045	-0.0398	-0.0636
	-	0.0191	0.0482	0.0017	0.0150	0.0239
Meat	-0.0116	-	0.0677	0.0149	-0.0448	0.0059
	0.0177	-	0.0984	0.0217	0.0683	0.0082
Fish	-0.2783	0.0857	-	0.1982	0.4156	-0.2140
	0.7146	0.2158	-	0.5051	1.0681	0.5509
Dairy&eggs	0.0186	0.0435	0.0819	-	0.0211	-0.0133
	0.0105	0.0243	0.0456	-	0.0118	0.0077
Fats	-0.6941	0.2092	0.5227	-0.1233	-	0.1406
	0.8735	0.2613	0.6568	0.1556	-	0.1786
Fruit	0.0415	-0.1305	-0.0455	-0.0197	-0.0173	-
	0.0547	0.1743	0.0621	0.0265	0.0232	-
Vegetables	0.0039	-0.0233	0.0639	0.1739	0.0658	0.0173
	0.0058	0.0318	0.0820	0.2264	0.0853	0.0224
Sugars	0.2246	-0.0645	-0.0351	-0.1469	-0.1594	-0.0029
	0.3957	0.1130	0.0610	0.2588	0.2811	0.0056
Others	-0.1141	-0.0110	0.2493	0.0561	-0.0663	-0.0550
	0.2742	0.0715	0.5456	0.1200	0.1603	0.1341
Coffee etc.	-0.1366	-0.0003	0.1623	-0.3925	-0.0487	-0.1498
	0.2006	0.0112	0.2421	0.5777	0.0708	0.2206
Non-alco	0.1231	-0.2141	-0.1964	0.0304	-0.1431	-0.1737
	0.3529	0.5823	0.5269	0.0919	0.3944	0.4776

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
Bakery	0.0442	-0.0774	-0.0008	0.0322	0.0508
	0.0167	0.0292	0.0010	0.0121	0.0192
Meat	-0.0031	0.0038	0.0234	0.0343	0.0151
	0.0057	0.0047	0.0332	0.0516	0.0219
Fish	0.0669	0.2788	-0.0770	-0.0100	-0.1110
	0.1718	0.7109	0.2000	0.0257	0.2861
Dairy&eggs	0.0570	-0.0447	0.0614	-0.0123	0.0688
	0.0318	0.0254	0.0343	0.0067	0.0387
Fats	-0.0492	0.0439	0.3034	0.0496	0.0492
	0.0654	0.0549	0.3800	0.0628	0.0617
Fruit	0.1248	0.0466	-0.0253	-0.1279	-0.0541
	0.1645	0.0610	0.0347	0.1690	0.0721
Vegetables	-	-0.0067	-0.0601	-0.0145	0.0102
	-	0.0094	0.0794	0.0193	0.0126
Sugars	-0.0691	-	-0.0133	-0.1814	-0.0188
	0.1219	-	0.0231	0.3196	0.0331
Others	0.2661	0.1584	-	-0.0878	0.0676
	0.6044	0.3547	-	0.2038	0.1484
Coffee etc.	-0.0136	-0.0350	-0.1856	-	-0.1842
	0.0216	0.0526	0.2728	-	0.2716
Non-alco	0.0006	-0.1612	-0.2760	-0.1665	-
	0.0200	0.4509	0.7569	0.4588	-

Median values of cross-price elasticities with standard deviations below

QUAIDS. The remarkable change is that meat and meat products are complements to fruit, vegetables, sugars, others, coffee etc., and non-alcoholic beverages. Generally, there are more complementary relations between products than in the previous cases.

Cosslett's estimator

The coefficients' estimates can be found in the Appendix Table A.6. All the α's coefficients except two and all the β's coefficients except one are significant. Moreover, most of them are significant on the 1 % level. On the other hand, only two λ's at quadratic term are significant. The coefficients of Cosslett's dummies are not reported as they do not have important interpretation for demand but most of them are significant on the 1 % level as well.

In the next Table 3.14. the income and own-price elasticities are given. Concerning the income elasticities, it can be observed that the mean values are more distant from the median values. It can be also seen from the high standard deviations that the elasticities' estimates are various across the households. Actually, the estimate for group of fats and oils is negative in the mean value but positive in the median value. Considering the median values, all the income elasticities are positive and most of them are lower than 1 determining inelastic demand and necessary goods. The luxury goods having elastic demand in income are in this case meat and meat products, fish and fish products, coffee, tea and cocoa, and non-alcoholic beverages. These results differ from the previous estimates. Meat was determined as luxury good in non-treated QUAIDS and Heien and Wessels' estimator. Fish was luxury good for non-treated QUAIDS, group of coffee etc. for Shonkwiler and Yen's estimator. Non-alcoholic beverages were denoted as luxury good in all four cases.

Table 3.14: Income and own-price Marshallian elasticities upon unrestricted QUAIDS estimated by Cosslett's estimator

	Income elasticities			Own price elasticities		
	Mean	Median	Std dev	Mean	Median	Std dev
Bakery	0.7051	0.7216	0.1074	-0.7737	-0.7875	0.0796
Meat	1.3046	1.2508	0.3534	-1.0580	-1.0480	0.0655
Fish	1.3124	1.1720	0.4977	-1.3217	-1.1797	0.4662
Dairy&eggs	0.6841	0.7053	0.1658	-0.9744	-0.9761	0.0134
Fats	-0.1450	0.0874	1.1290	-0.6189	-0.6986	0.3799
Fruit	0.0428	0.2700	0.9518	-0.3239	-0.4805	0.6768
Vegetables	0.4852	0.5936	0.4977	-1.2964	-1.2294	0.3131
Sugars	0.4659	0.6104	0.6606	-0.5240	-0.6575	0.5930
Others	0.7321	0.8082	0.3859	-0.8830	-0.9182	0.1784
Coffee etc.	1.2125	1.1520	0.2236	-0.6056	-0.7172	0.4149
Non-alco	2.4006	1.8730	2.3649	-1.8588	-1.5328	1.4738

Table 3.15: Cross-price Marshallian elasticities upon unrestricted QUAIDS estimated by Cosslett's estimator

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
Bakery	-	0.1595	0.1194	0.0251	-0.0369	-0.0132
	-	0.0607	0.0449	0.0095	0.0140	0.0056
Meat	-0.0473	-	-0.0133	-0.0182	-0.0698	-0.0313
	0.0625	-	0.0209	0.0251	0.1081	0.0459
Fish	-0.2635	-0.0968	-	0.1396	0.3425	-0.2124
	0.6836	0.2669	-	0.3537	0.8800	0.5503
Dairy&eggs	0.0731	0.1595	0.0923	-	0.0234	0.0340
	0.0407	0.0886	0.0516	-	0.0133	0.0182
Fats	-0.5186	0.4954	0.3947	-0.1150	-	0.2036
	0.6499	0.6178	0.4974	0.1452	-	0.2580
Fruit	0.1847	0.2030	0.0196	0.0369	-0.0139	-
	0.2364	0.2630	0.0262	0.0481	0.0202	-
Vegetables	0.1031	0.1949	0.1317	0.2007	0.0617	0.1010
	0.1279	0.2419	0.1681	0.2593	0.0820	0.1270
Sugars	0.3188	0.1646	0.0150	-0.0883	-0.1319	0.0955
	0.5499	0.2790	0.0256	0.1561	0.2286	0.1639
Others	0.0012	0.0233	-0.1270	0.0324	-0.1008	-0.0005
	0.0158	0.0440	0.2993	0.0696	0.2323	0.0142
Coffee etc.	0.0937	0.1167	0.1317	-0.3862	-0.0418	-0.1057
	0.1378	0.1713	0.1936	0.5680	0.0613	0.1553
Non-alco	0.0317	-0.2624	-0.1231	-0.0307	-0.1522	-0.1987
	0.1129	0.7095	0.3329	0.0809	0.4245	0.5462

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
Bakery	0.1350	-0.0656	0.0601	0.1430	0.1083
	0.0515	0.0245	0.0234	0.0542	0.0413
Meat	-0.0602	-0.0252	-0.0349	-0.0272	-0.0324
	0.0840	0.0363	0.0504	0.0372	0.0484
Fish	0.1335	0.1100	0.0053	-0.0638	-0.0145
	0.3259	0.2800	0.0233	0.1783	0.0503
Dairy&eggs	0.1494	-0.0280	0.1202	0.1088	0.1260
	0.0819	0.0163	0.0664	0.0609	0.0706
Fats	0.2464	0.0534	0.4235	0.4243	0.2789
	0.3042	0.0688	0.5293	0.5303	0.3498
Fruit	0.3831	0.0813	0.1578	0.2001	0.1319
	0.4999	0.1032	0.2043	0.2598	0.1697
Vegetables	-	0.0341	0.0562	0.1900	0.1291
	-	0.0439	0.0676	0.2382	0.1647
Sugars	0.1150	-	0.1164	0.0483	0.1014
	0.1961	-	0.1982	0.0807	0.1732
Others	0.4259	0.0239	-	0.0775	0.0784
	0.9662	0.0528	-	0.1631	0.1718
Coffee etc.	0.0332	-0.0150	-0.1385	-	-0.2204
	0.0484	0.0226	0.2033	-	0.3237
Non-alco	-0.1158	-0.1594	-0.3200	-0.2462	-
	0.3029	0.4527	0.8795	0.6683	-

Median values of cross-price elasticities with standard deviations below

In the case of own-price elasticities, the difference between mean and median values is evident for some food groups such as fruit or sugars. Considering the median values, all the elasticities are negative signifying the normal goods. The demand is elastic in own-price for meat and meat products, fish and fish products, vegetables, and non-alcoholic beverages. Once again, this is quite different from previous results. Meat is assigned to have elastic demand for the first time. Demand for fish was elastic in case of Heien and Wessels' estimator. Demand for vegetables was elastic in all four cases and demand for non-alcoholic beverages was elastic in non-treated QUAIDS model.

The cross-price elasticities are shown in the previous Table 3.15. Considering the signs of elasticities' estimates which determine the complementary or substitutionary relation between goods, there are less complementary relations than in the previous results. These results show that for meat and meat products, all other groups are complements. For non-alcoholic beverages all the food groups except bakery products are complements. Fats and oils are complements for every group except fish, dairy&eggs, and vegetables, which is the same behaviour as observed in the Shonkwiler and Yen's estimates. Finally, it can be also observed that as in previous cases the effects of cross-price elasticities are not symmetric.

Comparison of results

We have seen that the results vary throughout different estimators of unrestricted QUAIDS considering censoring of the data and non-treated unrestricted QUAIDS. Most of the coefficients are significant in all models and the values of estimated elasticities are valuable. To visualize the results, the following Figures 3.1 and 3.2 bring the comparison of median values of income and own-price elasticities with their 95 % confidence intervals for all different estimators. The summary of median values is given in the Table 3.16.

Table 3.16: Comparison of median values of income and own-price elasticities

	Income elasticities				Own price elasticities			
	QUAIDS	Heien &Wessels	Shonkwiler &Yen	Cosslett	QUAIDS	Heien &Wessels	Shonkwiler &Yen	Cosslett
Bakery	0.9357	0.9219	0.9409	0.7216	-0.9153	-0.7253	-0.8230	-0.7875
Meat	1.1429	1.1560	0.9967	1.2508	-0.9824	-0.9717	-0.9625	-1.0480
Fish	1.0188	0.8936	0.6801	1.1720	-0.9864	-1.1837	-0.5491	-1.1797
Dairy&eggs	0.9631	0.9811	0.9554	0.7053	-1.0001	-1.0198	-0.9986	-0.9761
Fats	0.8467	0.8217	0.5129	0.0874	-0.5976	-0.6217	-0.6271	-0.6986
Fruit	0.9771	0.8695	0.9714	0.2700	-0.7489	-0.3357	-0.5750	-0.4805
Vegetables	0.9313	0.9465	1.0851	0.5936	-1.2188	-1.7268	-1.4365	-1.2294
Sugars	0.9904	0.9230	1.1269	0.6104	-0.5621	-0.5840	-0.6852	-0.6575
Others	0.9681	0.9837	0.5153	0.8082	-1.0900	-0.9852	-0.9390	-0.9182
Coffee etc.	0.7843	0.6863	1.3846	1.1520	-0.7229	-0.6184	-0.7746	-0.7172
Non-alco	1.1146	1.3696	1.7587	1.8730	-1.0697	-0.3189	-1.3282	-1.5328

Generally, it can be observed in the figures that the confidence intervals are very

Figure 3.1: Comparison of median values with 95 % confidence interval bars of income elasticities

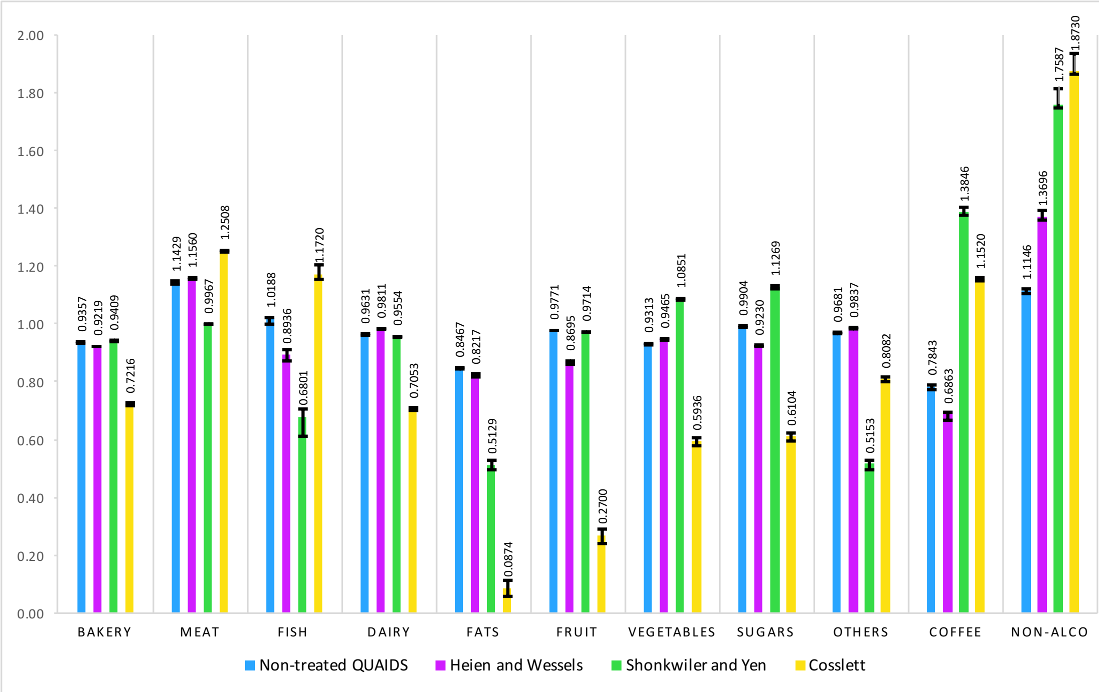
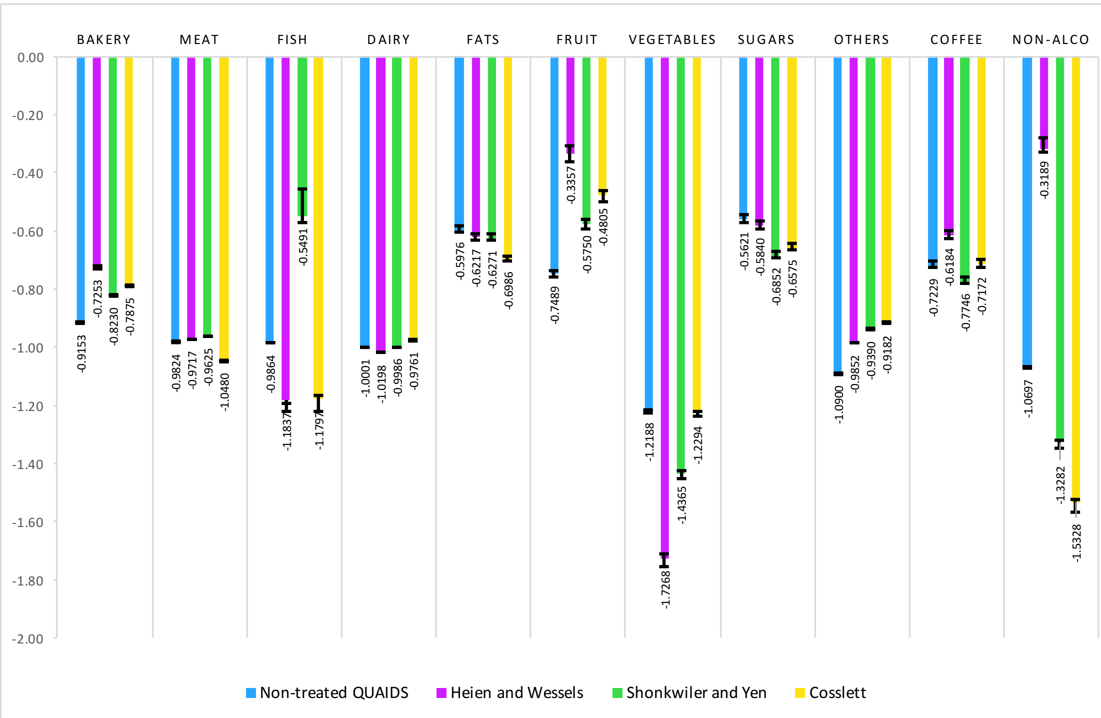


Figure 3.2: Comparison of median values with 95 % confidence interval bars of own-price elasticities



narrow and usually do not include values of the estimates of other estimators when the median values themselves are not close to each other. For example, for income elasticity for meat, the confidence intervals of non-treated QUAIDS and of Heien and Wessels' estimator do not overlap.

The income elasticities of non-treated QUAIDS and Heien and Wessels' estimator are very close to each other. The values from Shonkwiler and Yen's estimator are slightly different. The Cosslett's estimator's values stand apart from the other estimates in most of the cases. The attention is supposed to be paid on the values for fish category as the level of censoring is the highest one, 10 %. The values are different across all estimators in this case.

For the own-price elasticities, there are more similar values across some food subcategories but there is no clear similarity between two or more estimators across all estimated groups. The biggest differences between estimates are for subcategories of fish and non-alcoholic beverages and these are the categories with the highest levels of censoring as well. On the other hand, all the estimated own-price elasticities have identical sign and there are not any suspicious values that would be distant from other results.

The comparison of overall results does not lead to the definite decision about the final model. The data is censored especially for the fish and fish products equation, thus the censoring technique is necessary. From the theory, the Heien and Wessels' estimator is inconsistent and can provide misleading results especially in the case of high levels of censoring. In our case, its estimates are close to the estimates of non-treated QUAIDS (for fish subcategory). Secondly, the semi-parametric method does not rely on the distributional assumption, it can be then more sensible in case of small number of zeros in the data set, which is the case here. For example, the subcategories of fats and fruit have minimal levels of censoring and the estimated income elasticities are completely different form other estimators. Finally, we have decided to build up the final model upon the Shonkwiler and Yen's estimator as this technique treats the censoring and was proved to be consistent estimator.

3.3.3 Final censored model of QUAIDS

The final model is the unrestricted QUAIDS estimated by Shonkwiler and Yen's estimator in order to deal with selectivity problem in the dataset. The socio-demographic variables controlling size of the household, the age and education of the head of the household, number of children younger than 5 years and the location in the village, are incorporated in the model. The coefficients' estimates are given in the next Table 3.17.

The number of household members influences significantly the expenditures on all

Table 3.17: Estimates of final censored model of unrestricted QUAIDS

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.0118***	0.0114***	0.0103***	0.0121***	0.0118***	0.0091***
	0.0002	0.0003	0.0004	0.0003	0.0003	0.0002
$size$	0.0182***	0.0126***	0.0144***	0.0062***	0.0054***	-0.0010
	0.0017	0.0020	0.0047	0.0015	0.0018	0.0013
age	0.0028***	0.0060***	0.0045***	0.0017**	0.0026***	0.0029***
	0.0010	0.0016	0.0006	0.0007	0.0005	0.0006
$primary$	0.0126***	0.0113***	0.0072***	0.0079***	0.0117***	0.0064***
	0.0009	0.0006	0.0008	0.0008	0.0011	0.0021
$tertiary$	0.0025*	-0.0006	0.0132***	0.0111***	0.0024*	0.0086***
	0.0014	0.0013	0.0029	0.0008	0.0014	0.0017
d_5	-0.0008	0.0047***	0.0089***	0.0255***	0.0043***	0.0065***
	0.0014	0.0011	0.0022	0.0019	0.0014	0.0017
$village$	0.0153***	0.0104***	0.0066***	-0.0058***	0.0162***	-0.0020
	0.0024	0.0015	0.0013	0.0015	0.0019	0.0019
β	-0.0001	0.0353***	0.0073**	0.0083***	0.0000	0.0142***
	0.0023	0.0028	0.0032	0.0027	0.0030	0.0013
γ_1	0.0149***	-0.0086***	-0.0138***	0.0019**	-0.0294***	-0.0017*
	0.0011	0.0011	0.0016	0.0008	0.0016	0.0009
γ_2	-0.0011	-0.0061***	-0.0140***	0.0016	-0.0040**	-0.0114***
	0.0011	0.0011	0.0008	0.0011	0.0018	0.0009
γ_3	-0.0009	-0.0072***	-0.0183***	-0.0004	-0.0036***	-0.0146***
	0.0010	0.0011	0.0012	0.0007	0.0010	0.0008
γ_4	-0.0043***	-0.0075***	-0.0045**	-0.0024	-0.0138***	-0.0046***
	0.0012	0.0015	0.0021	0.0018	0.0026	0.0014
γ_5	-0.0120***	-0.0197***	-0.0035	-0.0026	0.0172***	-0.0080***
	0.0015	0.0021	0.0028	0.0021	0.0017	0.0017
γ_6	-0.0089***	-0.0068***	-0.0126***	-0.0017	-0.0028***	0.0135***
	0.0017	0.0010	0.0016	0.0014	0.0010	0.0012
γ_7	0.0024***	-0.0059***	-0.0042***	0.0061***	-0.0044***	0.0040***
	0.0009	0.0013	0.0009	0.0012	0.0016	0.0011
γ_8	-0.0207***	-0.0122***	-0.0035**	-0.0092***	-0.0084***	-0.0091***
	0.0018	0.0015	0.0016	0.0019	0.0018	0.0014
γ_9	-0.0061***	-0.0059***	-0.0080***	0.0040***	0.0068***	-0.0053*
	0.0010	0.0013	0.0017	0.0012	0.0009	0.0029
γ_{10}	0.0017***	-0.0017**	-0.0104***	0.0000	-0.0014	-0.0074***
	0.0006	0.0008	0.0018	0.0006	0.0013	0.0014
γ_{11}	0.0020*	0.0007	-0.0003	0.0041*	0.0011	-0.0039**
	0.0012	0.0014	0.0011	0.0021	0.0032	0.0019
λ	0.0022	-0.0002	0.0022*	0.0002	0.0012***	0.0007
	0.0022	0.0015	0.0012	0.0009	0.0004	0.0008
δ			0.0118***		0.0102***	0.0099***
			0.0004		0.0001	0.0003

Estimates of final censored model of unrestricted QUAIDS continued

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.0096***	0.0096***	0.0094***	0.0105***	0.0092***
	0.0002	0.0002	0.0003	0.0003	0.0002
$size$	-0.0005	-0.0011	0.0032*	0.0029	-0.0028
	0.0020	0.0024	0.0019	0.0021	0.0017
age	0.0038***	0.0014***	0.0027***	0.0041***	0.0021***
	0.0008	0.0004	0.0005	0.0006	0.0006
$primary$	0.0122***	0.0073***	0.0131***	0.0074***	0.0100***
	0.0011	0.0010	0.0013	0.0013	0.0011
$tertiary$	0.0160 ***	0.0081***	0.0082***	0.0130***	0.0008
	0.0014	0.0021	0.0022	0.0018	0.0014
d_5	0.0068***	0.0068***	0.0038	0.0071***	0.0012
	0.0017	0.0013	0.0023	0.0022	0.0012
$village$	-0.0058***	0.0142***	0.0080***	0.0082***	0.0092***
	0.0018	0.0019	0.0012	0.0018	0.0017
β	0.0191***	0.0062***	0.0127***	0.0112***	0.0200 ***
	0.0023	0.0024	0.0027	0.0022	0.0038
γ_1	-0.0046***	0.0083***	-0.0041**	-0.0063***	0.0032***
	0.0007	0.0009	0.0018	0.0012	0.0009
γ_2	-0.0090***	-0.0068***	-0.0108***	-0.0061***	-0.0060***
	0.0010	0.0008	0.0020	0.0015	0.0009
γ_3	-0.0100***	-0.0107***	-0.0133 ***	-0.0080***	-0.0077***
	0.0010	0.0008	0.0012	0.0011	0.0013
γ_4	0.0035***	-0.0081***	0.0010	-0.0180***	0.0009
	0.0013	0.0015	0.0021	0.0024	0.0012
γ_5	-0.0044***	-0.0124***	-0.0129***	-0.0074***	-0.0071***
	0.0011	0.0026	0.0031	0.0014	0.0015
γ_6	-0.0047***	-0.0019	-0.0057***	-0.0078***	-0.0026**
	0.0009	0.0014	0.0017	0.0016	0.0013
γ_7	-0.0246***	-0.0034***	0.0058***	-0.0018	0.0010
	0.0014	0.0009	0.0018	0.0015	0.0015
γ_8	-0.0048***	0.0294***	-0.0024	-0.0096***	-0.0056***
	0.0013	0.0017	0.0018	0.0014	0.0010
γ_9	-0.0101***	-0.0040**	-0.0048*	-0.0089***	-0.0067***
	0.0018	0.0016	0.0026	0.0023	0.0018
γ_{10}	-0.0052***	-0.0079***	-0.0064***	0.0011	-0.0051***
	0.0010	0.0007	0.0012	0.0014	0.0012
γ_{11}	0.0051**	0.0028	0.0022	-0.0015	-0.0197***
	0.0023	0.0018	0.0023	0.0020	0.0006
λ	0.0008	0.0005	0.0006	0.0013	0.0003
	0.0010	0.0006	0.0006	0.0010	0.0011
δ	0.0104***	0.0101***	0.0116***	0.0109***	0.0123 ***
	0.0002	0.0001	0.0002	0.0003	0.0003
logL	138019.38				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

analysed subcategories except fruit, vegetables, sugars and both subcategories of non-alcoholic beverages. The effect is positive in all cases except non-alcoholic beverages. The biggest influence the household size has on the consumption is for the bakery (coefficient equal to 0.0182), meat (0.0126) and fish products (0.0144). The bakery and meat products are creating the base of expenditure shares according to the descriptive statistics. The age category of the household approximated by the age of the household head has significant and positive effect in all cases, thus there is no food category which would be significantly excluded from the diet because of the age of consumers. The demand is larger especially for meat (0.0060), fish (0.0045) and coffee(0.0041), and increases with age with relatively lower growth for dairies (0.0017) and sugars (0.0014). The different levels of gained education by household head are positive and significant in most of the cases. The positive coefficient at the variable $primary$ indicates that these households spend higher levels of expenditures on given subcategory than the households where the secondary education was the highest reached level of education. The households with primary education spend more on the basic food categories such as bakery, meat, and fats. On the other hand, the households with tertiary education have higher expenditures on subcategories of fish, dairy&eggs, fruit, vegetables and coffee, tea and cocoa. The presence of small children is not significant for bakery products, others and non-alcoholic beverages, the effect is otherwise positive in all cases. Having children younger than 5 years increases particularly the expenditures on dairy&eggs (0.0255) that are usually creating a basis of their diet. The non-significance for others group is not surprising as the group consists mainly of spices and baking utensils. The influence of location in the villages is negative for dairy&eggs (-0.0058) and vegetables (-0.0058) as can be expected as these aliments are usually gained from the gardening or keeping own hens in the garden.

Next, in the second stage, all the α's coefficients are significant at the 1 % level that is true for eight β's as well. All the δ's, coefficients specific for this estimator, are also significant. Only two λ's, coefficients at quadratic terms, are significant, which is the same as for the unrestricted QUAIDS estimated by Shonkwiler and Yen estimator without control for socio-demographic variables. Finally, almost all γ's coefficients are significant as well.

In the next Table 3.18. the estimated income and own-price elasticities are shown. The estimates of income elasticities are to some extent surprising. The mean and median values are close to each other meaning that there are not any distant results. On the other hand, the standard deviations are quite high for some of the subcategories. Differently from the estimates of previous section, all the income elasticities are higher than 1, which would signify that the demand for all food groups is elastic and these goods can be denoted as luxuries. This is relatively unconventional result for food

Table 3.18: Income and own-price Marshallian elasticities upon final censored model of QUAIDS

	Income elasticities			Own price elasticities		
	Mean	Median	Std dev	Mean	Median	Std dev
Bakery	1.1704	1.1606	0.0749	-0.8251	-0.8369	0.0626
Meat	1.1661	1.1352	0.2020	-0.9777	-0.9819	0.0272
Fish	3.3161	2.2786	3.4145	-0.7133	-0.8392	0.4339
Dairy&eggs	1.0681	1.0636	0.0376	-0.9951	-0.9955	0.0026
Fats	1.4190	1.3233	0.4386	-0.2693	-0.4229	0.7289
Fruit	1.5203	1.3960	0.5352	-0.5879	-0.6845	0.4190
Vegetables	1.4578	1.3610	0.4528	-1.3906	-1.3058	0.3957
Sugars	1.2949	1.2118	0.3891	-0.2032	-0.4307	1.0056
Others	1.9237	1.6323	1.5259	-1.0536	-1.0369	0.0874
Coffee etc.	1.9856	1.7022	1.0475	-0.7853	-0.8462	0.2268
Non-alco	1.8085	1.5019	1.3441	-1.7404	-1.4598	1.2677

demand. For example, the Engel's law that is saying that with higher income the expenditures on food diminish implies that the income elasticity for food should be between 0 and 1. In the previous section, the results of income elasticities for Shonkwiler and Yen's estimator without control of socio-demographic variables were considerably lower for some categories (for fats, the elasticity was almost three times lower). It can be caused by the fact, that in the final model the part of the influence of the income can be covered by some coefficients of the socio-demographic variables such as $size$ or educations levels that are possibly correlated with income of households.

Concerning the own-price elasticities, the demand for vegetables, others and non-alcoholic beverages is elastic. Elasticities for other food groups are close to -1 except fats, fruit and sugars. Further, the mean and median values are almost identical for subcategories which are creating the basis of expenditures according to descriptive statistics: bakery, meat, dairy&eggs. Comparing to the results from previous section for Shonkwiler and Yen's estimator, the demand for others was inelastic, the rest of results are similar.

Next Table 3.19. brings the results of cross-price elasticities. Most of the calculated values are negative. For previous estimates, only the half of the relations was described by the negative value of elasticity. Hence, in this case, most of the analysed groups are complements for each other. For example, for meat and meat products, all other groups are complements. The same is true for the subcategory of coffee, tea and cocoa and on the other hand, this food group is also complement for all other groups. The subcategory of sugar, jams and chocolate is also complement for all other groups except fish and others. Fruit are complement to all other groups as well. It can be also easily observed that the values are not symmetric.

Table 3.19: Cross-price Marshallian elasticities upon final censored model of QUAIDS

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
Bakery	-	-0.0322	0.0094	-0.0087	-0.0521	-0.0570
	-	0.0131	0.0142	0.0088	0.0250	0.0230
Meat	-0.0353	-	-0.0132	-0.0118	-0.0643	-0.0225
	0.0416	-	0.0159	0.0140	0.0926	0.0313
Fish	-0.5198	-0.7081	-	-0.0183	0.0312	-0.4873
	1.3725	1.8347	-	0.1654	0.2638	1.2384
Dairy&eggs	0.0106	0.0042	0.0049	-	-0.0067	-0.0080
	0.0085	0.0067	0.0073	-	0.0052	0.0052
Fats	-0.6400	-0.1365	-0.0454	-0.2664	-	-0.0653
	0.8029	0.1703	0.0690	0.3347	-	0.0840
Fruit	-0.0281	-0.2361	-0.2150	-0.0303	-0.0895	-
	0.0444	0.3138	0.2880	0.0448	0.1246	-
Vegetables	-0.0561	-0.1412	-0.0843	0.0918	-0.0087	-0.0522
	0.0824	0.1957	0.1281	0.1198	0.0384	0.0756
Sugars	0.1424	-0.1358	-0.1571	-0.1106	-0.1819	-0.0294
	0.2614	0.2315	0.2698	0.1946	0.3147	0.0516
Others	-0.1295	-0.4061	-0.3609	0.1157	-0.3387	-0.1730
	0.2706	0.8849	0.7890	0.2930	0.7524	0.3790
Coffee etc.	-0.1603	-0.2373	-0.1312	-0.3821	-0.0972	-0.1983
	0.2471	0.3490	0.2242	0.5713	0.1717	0.2960
Non-alco	0.0686	-0.1575	-0.1070	0.0877	-0.0883	-0.0389
	0.1802	0.5114	0.3662	0.2270	0.2850	0.1617

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
Bakery	-0.0116	-0.1062	-0.0388	-0.0029	-0.0160
	0.0066	0.0432	0.0160	0.0045	0.0094
Meat	-0.0376	-0.0293	-0.0192	-0.0130	-0.0106
	0.0525	0.0402	0.0259	0.0183	0.0171
Fish	-0.3563	0.0431	-0.3033	-0.4965	-0.2204
	0.9125	0.2416	0.7856	1.2733	0.5765
Dairy&eggs	0.0274	-0.0430	0.0244	-0.0038	0.0156
	0.0166	0.0241	0.0149	0.0044	0.0087
Fats	-0.1491	-0.1384	0.1481	-0.0559	-0.0332
	0.1899	0.1762	0.1879	0.0717	0.0484
Fruit	0.0220	-0.1032	-0.0846	-0.1553	-0.1218
	0.0394	0.1398	0.1155	0.2091	0.1624
Vegetables	-	-0.0058	-0.1190	-0.0874	0.0177
	-	0.0314	0.1627	0.1239	0.0232
Sugars	-0.0857	-	-0.0642	-0.1475	0.0174
	0.1497	-	0.1093	0.2550	0.0327
Others	0.1080	0.0183	-	-0.2478	-0.0115
	0.2807	0.0895	-	0.5448	0.0353
Coffee etc.	-0.1441	-0.1470	-0.2234	-	-0.1429
	0.2136	0.2337	0.3334	-	0.2096
Non-alco	-0.0303	-0.0432	-0.1314	-0.1385	-
	0.1459	0.1592	0.4004	0.4362	-

Median values of cross-price elasticities with standard deviations below

Chapter 4

Discussion of results

The aim of the empirical part of this thesis was to find the most suitable version and estimator of AIDS model to evaluate food demand of Czech households. The preliminary tests have showed that the quadratic version fits the data better than its linear counterpart. It means that the relationship of income and expenditures is not linear in case of Czech households. Further, the conditions of homogeneity and symmetry from the consumer theory were tested and the evidence of data has shown that these conditions are violated. Hence the unrestricted and quadratic version of AIDS model is the most suitable for Czech household budget survey data from 2013.

Although the demand for food is divided in 11 subcategories such as bakery products, meat and meat products, fruit, and other categories, the significant portion of zeros occurs in the data. Concretely for fish and fish products 10.89 % of observed expenditure shares are equal to 0. This is the case of highest percentage of zeros in our dataset. The data for some subcategories are censored and thus needs the special attention when estimating the model. Three different techniques have been evaluated and Shonkwiler and Yen's estimator has been determined as theoretically plausible and empirically giving the valuable results.

The final unrestricted QUAIDS model has been estimated involving the socio-demographic variables and the results are given in the previous section. To compare these results with previous studies held on Czech data, the next Table 4.1. brings the estimated income and own-price elasticities from the relevant studies known to the author.

The dataset from reported studies are almost 20 years older than the data used in this thesis. On the other hand, these are the most relevant data to be compared because of the same origin and approximatively same food subcategories. The elasticities for bakery products (income 1.1606 and own-price -0.8369) can be compared to some extent to starches, carbohydrates, and cereals. The values are different between the studies as well but the income elasticity is similar to carbohydrates in Janda et al. (2000) and the own-price to starches in Crawford et al. (2003). The meat and meat products were analysed in all cited studies but the category contained fishes as well. Our estimate of income elasticity 1.1352 is similar to both Janda et al. (2000) and Crawford et al. (2003) as well, estimate of own-price elasticity -0.9819 is more closer to Crawford et al. (2003).

Table 4.1: Elasticities estimated in other studies

Author	Data	Category	Income el.	Own price el.
Crawford et al. (2003)	1991-1992	meat	1.195	-0.968
		dairy	0.596	-0.062
		starches	0.864	-0.721
		vege/fruit	0.356	1.067
		sweet	0.860	-1.078
Janda et al. (2000)	1993-1997	meat	1.15	-1.14
		carbohydrates	1.13	-1.23
		vege/fruit	0.82	-0.46
		dairy&sweets	1.14	-0.86
Brosig (1998)	1991-1996	carbohydrates	0.52	-0.60
		fruit/vege	0.83	-0.80
		proteins	0.83	-0.68
		oils/fats	0.70	-0.63
		dairy	0.44	-0.32
		other	0.55	-0.68
Janda and Rausser (1998)	1993-1995	meat	0.59	-0.74
		milk	0.26	-0.85
		cereals	0.93	-1.56
		fruit/vege	0.62	-0.82
		other	1.23	-0.48

The estimates for subcategory of dairy&eggs (income 1.0636 and own-price -0.9955) are quite different from the estimates of given studies. Fats and oils were analysed in Brosig (1998) only and the values are once again distant. The groups of fruit and vegetables were analysed in one group in given studies and the estimates are not really comparable. The subcategory of sugars (income 1.2118 and own-price -0.4307) can be compared to sweet in Crawford et al. (2003) but the values are distant as well.

In sum, the estimates of elasticities given in previous section are valuable and most of them are comparable with previous studies. The exact values are not the same but it is rather expectable as there is almost 20 years difference in expenditures data and the Czech society has evolved and faced the transition period during these times.

Chapter 5

Conclusion

The theoretical part of this thesis aims to discuss models proposed in the literature to analyse food demand. Linear expenditures system (LES), Quadratic expenditure system (QES), Rotterdam demand system, Translog demand system, and Almost ideal demand system (AIDS) are discussed. The last named is the most popular among researchers. The model is built up with the possibility (not necessity) to satisfy the conditions from consumer theory - additivity, homogeneity and symmetry, the interpretation of coefficients is straightforward, it allows for quadratic shape of Engel's curve, it can be easily extended to control for socio-demographic characteristics and it is simple to be applied.

The demand for food is divided in eleven subcategories to be analysed: bakery, meat, fish, dairy and eggs, oils and fats, fruit, vegetables, sugar, jams and chocolates, others, coffee, tea and cocoa, and non-alcoholic beverages. This division corresponds to the first COICOP category of expenditures. The demand for food is analysed on the budget survey data of Czech households from 2013. The problem of frequent zeros in the dataset arises for some food subcategories. The zero observation means that the household does not buy any item from given subcategory. Such dataset is called censored and the researcher faces the selectivity problem that has to be treated to receive unbiased results of the analysis. Various estimators, which deal with selectivity problem, are described and empirically examined in this thesis. Concretely, the Heien and Wessels's estimator, Shonkwiler and Yen's estimator and semi-parametric estimator of Cosslett are studied.

The unrestricted QUAIDS model that does not incorporate directly the homogeneity and symmetry conditions is estimated by the Shonkwiler and Yen's estimator dealing with selectivity as this framework was evaluated to be the most suitable for the data from Czech households. The effect of socio-demographic variables is controlled as well. The income and Marshallian price elasticities are estimated for all food subcategories in the demand system as elasticities are the most important tool in demand analysis. The results show that all income elasticities are higher than 1 that is not expected as according to Engel's law the income elasticity for food is supposed to be between 0 and 1. All the own-price elasticities are negative and the demand is elastic for subcategory of vegetables, others, non-alcoholic beverages. Finally, most of the cross-price elasticities are negative signifying that the relation between the groups is complementary.

This thesis brings the empirical comparison of different estimators, which deal with selectivity problem. Such comparison has not been found in the literature. Moreover, for the first time the food demand is estimate incorporating such technique on the data from Czech households. Usually, the zero observations are simply not treated or even deleted. The interesting extension of this thesis would be a deeper comparison of different selectivity techniques by Monte Carlo experiments. Some of the estimators rely on the distributional assumptions and their behaviour was not examined for the case of violation of these assumption. Their behaviour can be influenced by different levels of censoring as well. Hence, the variety of Monte Carlo experiments to test general behaviour of different estimators remains the open question.

Bibliography

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24(1–2):3 – 61.

Andrews, D. W. and Schafgans, M. M. (1998). Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies*, 65(3):497–517.

Anwar, A., Aziz, B., and Ali, S. (2012). The rotterdam demand model and its application to major food items in pakistan. *Journal of Basic and Applied Scientific Research*, 2(5):5081–5087.

Banks, J., Blundell, R., and Lewbel, A. (1997). Quadratic engel curves and consumer demand. *Review of Economics and statistics*, 79(4):527–539.

Barten, A. P. (1964). Consumer demand functions under conditions of almost additive preferences. *Econometrica: Journal of the Econometric Society*, pages 1–38.

Barten, A. P. (1969). Maximum likelihood estimation of a complete system of demand equations. *European economic review*, 1(1):7–73.

Berges, M. and Casellas, K. (2002). a demand system analysis of food for poor and non poor households. the case of argentina. *Paper prepared for presentation at the Xth EAAE Congress 'Exploring Diversity in the European Agri -Food System', Paper prepared for presentation at the Xth EAAE Congress 'Exploring Diversity in the European Agri - Food System', Zaragoza (Spain), 28-31 August 2002.*

Blanciforti, L. and Green, R. (1983). An almost ideal demand system incorporating habits: an analysis of expenditures on food and aggregate commodity groups. *The Review of Economics and Statistics*, pages 511–515.

Brosig, S. (1998). A model of food consumption in czech private households 1991-96. Technical report, Inst. für Agrarökonomie.

Capps Jr, O. and Schmitz, J. D. (1991). A recognition of health and nutrition factors in food demand analysis. *Western Journal of Agricultural Economics*, pages 21–35.

Chern, W. S. and Wang, G. (1994). The engel function and complete food demand system for chinese urban households. *China Economic Review*, 5(1):35–57.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1975). Transcendental logarithmic utility functions. *The American Economic Review*, 65(3):367–383.

Cosslett, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica*, 51(3):765–82.

Cosslett, S. R. (1991). Semiparametric estimation of a regression model with sample selectivity. *Nonparametric and semiparametric methods in econometrics and statistics*, pages 175–97.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5):pp. 829–844.

Crawford, I., Laisney, F., and Preston, I. (2003). Estimation of household demand systems with theoretically compatible engel curves and unit value specifications. *Journal of Econometrics*, 114(2):221 – 241.

ČSÚ (1999). *Klasifikace individuální spotřeby (CZ-COICOP) - III. Systematická část.* Czech Statistical Office, dostupné z [https://www.czso.cz/csu/czso/klasifikace_individualni_spotreby_-cz_coicop-].

Davis, C., Blayney, D., Cooper, J., Yen, S., et al. (2009). An analysis of demand elasticities for fluid milk products in the us. In *International Association of Agricultural Economists Meeting, August*, pages 16–22.

Deaton, A. S. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3):312–26.

Ecker, O. and Qaim, M. (2011). Analyzing nutritional impacts of policies: an empirical study for malawi. *World Development*, 39(3):412–428.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, pages 363–390.

Green, R. and Alston, J. M. (1990). Elasticities in aids models. *American Journal of Agricultural Economics*, 72(2):442–445.

Green, R. and Alston, J. M. (1991). Elasticities in aids models: a clarification and extension. *American Journal of Agricultural Economics*, 73(3):874–875.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, pages 313–318.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 475–492. NBER.

Heien, D. and Pompelli, G. (1988). The demand for beef products: cross-section estimation of demographic and economic effects. *Western Journal of Agricultural Economics*, pages 37–44.

Heien, D. and Wessels, C. R. (1990). Demand Systems Estimation with Microdata: A Censored Regression Approach. *Journal of Business & Economic Statistics*, 8(3):365–71.

Howe, H., Pollak, R. A., and Wales, T. J. (1979). Theory and time series estimation of the quadratic expenditure system. *Econometrica: Journal of the Econometric Society*, pages 1231–1247.

Hussinger, K. (2008). R&d and subsidies at the firm level: an application of parametric and semiparametric two-step selection models. *Journal of Applied Econometrics*, 23(6):729–747.

Janda, K., McCluskey, J. J., and Rausser, G. C. (2000). Food import demand in the czech republic. *Journal of Agricultural Economics*, 51(1):22–44.

Janda, K. and Rausser, G. (1998). The estimation of hicksian and expenditure elasticities of conditional demand for food in the transition economy 1993| 1995. *University of California, Berkeley.*

Janský, P. (2014). Consumer demand system estimation and value added tax reforms in the czech republic. *Czech Journal of Economics and Finance*, 64(3):246–273.

Klein, L. R. and Rubin, H. (1947). A constant-utility index of the cost of living. *The Review of Economic Studies*, 15(2):84–87.

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica*, 61(2):pp. 387–421.

Kumar, P., Kumar, A., Parappurathu, S., Raju, S., et al. (2011). Estimation of demand elasticity for food commodities in india. *Agricultural Economics Research Review*, 24(1):1–14.

Mittal, S. et al. (2010). Application of the quaids model to the food sector in india. *Journal of Quantitative Economics*, 8(1):42–54.

Newman, C. F., Henchion, M., and Matthews, A. (2001). A double-hurdle model of irish household expenditure on prepared meals. Economic papers, Trinity College Dublin, Economics Department.

Nicol, C. J. (1989). Estimating a third-order translog using canadian cross-sectional micro-data. *Canadian Journal of Economics*, pages 543–560.

Poi, B. P. (2002). From the help desk: Demand system estimation. *Stata Journal*, 2(4):403–410.

Pollak, R. A. and Wales, T. J. (1978). Estimation of complete demand systems from household budget data: the linear and quadratic expenditure systems. *The American Economic Review*, 68(3):348–359.

Pollak, R. A. and Wales, T. J. (1980). Comparison of the quadratic expenditure system and translog demand systems with alternative specifications of demographic effects. *Econometrica: Journal of the Econometric Society*, pages 595–612.

Raper, K. C., Wanzala, M. N., and Nayga Jr, R. M. (2002). Food expenditures and household demographic composition in the us: a demand systems approach. *Applied Economics*, 34(8):981–992.

Selvanathan, E. and Selvanathan*, S. (2004). Economic and demographic factors in australian alcohol demand. *Applied Economics*, 36(21):2405–2417.

Shonkwiler, J. S. and Yen, S. T. (1999). Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics*, 81(4):972–982.

Stone, R. (1954). Linear expenditure systems and demand analysis: an application to the pattern of british demand. *The Economic Journal*, pages 511–527.

Theil, H. (1965). The information approach to demand analysis. *Econometrica: Journal of the Econometric Society*, pages 67–87.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1):pp. 24–36.

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

Appendix A

Complementary tables

Table A.1: Estimates of restricted AIDS

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.1404***	0.1156***	0.0909***	0.1292***	0.0908***	0.1256***
	0.0041	0.0028	0.0041	0.0031	0.0058	0.0035
β	0.0086***	0.0150***	-0.0009	0.0087***	-0.0088***	-0.0054***
	0.0013	0.0015	0.0010	0.0009	0.0010	0.0008
γ_1	0.0686***	-0.0129*	0.0010	-0.0064**	-0.0281***	-0.0153***
	0.0063	0.0066	0.0019	0.0032	0.0018	0.0030
γ_2	-0.0129*	0.0266***	0.0092**	0.0085	-0.0093***	-0.0233***
	0.0066	0.0048	0.0047	0.0055	0.0028	0.0034
γ_3	0.0010	0.0092**	-0.0530***	0.0074***	0.0129***	-0.0048***
	0.0019	0.0047	0.0038	0.0016	0.0021	0.0015
γ_4	-0.0064**	0.0085	0.0074***	-0.0077**	-0.0068***	0.0002
	0.0032	0.0055	0.0016	0.0035	0.0015	0.0025
γ_5	-0.0281***	-0.0093***	0.0129***	-0.0068***	0.0164***	0.0051**
	0.0018	0.0028	0.0021	0.0015	0.0021	0.0020
γ_6	-0.0153***	-0.0233***	-0.0048***	0.0002	0.0051**	0.0300***
	0.0030	0.0034	0.0015	0.0025	0.0020	0.0031
γ_7	0.0080**	-0.0060	0.0104***	0.0236***	0.0054**	0.0169***
	0.0036	0.0049	0.0026	0.0019	0.0026	0.0022
γ_8	-0.0050**	-0.0079**	0.0025*	-0.0111***	-0.0024***	0.0014
	0.0020	0.0033	0.0014	0.0019	0.0009	0.0014
γ_9	0.0021	-0.0137***	0.0017	0.0019	0.0004	0.0010
	0.0031	0.0031	0.0014	0.0016	0.0015	0.0024
γ_{10}	-0.0160***	0.0252***	0.0122***	-0.0141***	0.0080***	-0.0059**
	0.0019	0.0053	0.0039	0.0025	0.0024	0.0027
γ_{11}	0.0042**	0.0037	0.0005	0.0045***	-0.0016	-0.0051***
	0.0020	0.0040	0.0010	0.0013	0.0012	0.0017

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.1034***	0.1109***	0.0777***	0.0086**	0.0067
	0.0044	0.0028	0.0039	0.0042	0.0118
β	-0.0103***	-0.0054***	-0.0049***	-0.0015	0.0048***
	0.0014	0.0007	0.0007	0.0011	0.0016
γ_1	0.0080**	-0.0050**	0.0021	-0.0160***	0.0042**
	0.0036	0.0020	0.0031	0.0019	0.0020
γ_2	-0.0060	-0.0079**	-0.0137***	0.0252***	0.0037
	0.0049	0.0033	0.0031	0.0053	0.0040
γ_3	0.0104***	0.0025*	0.0017	0.0122***	0.0005
	0.0026	0.0014	0.0014	0.0039	0.0010
γ_4	0.0236***	-0.0111***	0.0019	-0.0141***	0.0045***
	0.0019	0.0019	0.0016	0.0025	0.0013
γ_5	0.0054**	-0.0024***	0.0004	0.0080***	-0.0016
	0.0026	0.0009	0.0015	0.0024	0.0012
γ_6	0.0169***	0.0014	0.0010	-0.0059**	-0.0051***
	0.0022	0.0014	0.0024	0.0027	0.0017
γ_7	-0.0764***	0.0016	0.0066***	0.0028	0.0070***
	0.0068	0.0018	0.0025	0.0024	0.0022
γ_8	0.0016	0.0236***	0.0023**	-0.0033***	-0.0015***
	0.0018	0.0012	0.0010	0.0010	0.0006
γ_9	0.0066***	0.0023**	0.0017	-0.0057**	0.0016*
	0.0025	0.0010	0.0025	0.0023	0.0010
γ_{10}	0.0028	-0.0033***	-0.0057**	0.0029	-0.0060***
	0.0024	0.0010	0.0023	0.0045	0.0012
γ_{11}	0.0070***	-0.0015***	0.0016*	-0.0060***	-0.0073***
	0.0022	0.0006	0.0010	0.0012	0.0015
logL	58938.28				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Estimates of unrestricted AIDS

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.0074***	0.0072***	0.0022**	0.0041***	0.0091***	-0.0021**
	0.0009	0.0012	0.0009	0.0007	0.0010	0.0010
β	-0.0150***	0.0426***	0.0000	-0.0071***	-0.0061***	-0.0054***
	0.0026	0.0037	0.0010	0.0020	0.0013	0.0011
γ_1	0.0694***	-0.0388***	-0.0105***	-0.0045	-0.0404***	0.0045
	0.0082	0.0088	0.0023	0.0039	0.0049	0.0038
γ_2	0.0064*	0.0116***	-0.0029	0.0073**	0.0091**	-0.0212***
	0.0034	0.0032	0.0039	0.0032	0.0044	0.0053
γ_3	0.0359***	0.0384***	-0.0078***	0.0207***	0.0454***	-0.0141**
	0.0035	0.0062	0.0030	0.0034	0.0049	0.0060
γ_4	-0.0079	-0.0031	0.0068***	-0.0053	-0.0119***	0.0052
	0.0052	0.0044	0.0013	0.0037	0.0022	0.0032
γ_5	-0.0144***	-0.0418***	0.0100***	0.0126***	0.0145***	0.0107**
	0.0039	0.0063	0.0022	0.0047	0.0016	0.0047
γ_6	-0.0381***	-0.0091***	0.0002	-0.0112**	-0.0003	0.0446***
	0.0068	0.0024	0.0021	0.0044	0.0024	0.0033
γ_7	0.0189***	-0.0166***	0.0102***	0.0280***	-0.0117***	0.0179***
	0.0056	0.0048	0.0035	0.0055	0.0035	0.0052
γ_8	-0.0129***	-0.0023	0.0032**	-0.0056**	-0.0002	-0.0001
	0.0020	0.0045	0.0014	0.0026	0.0008	0.0022
γ_9	-0.0090*	0.0000	0.0008	0.0132***	0.0040**	0.0024
	0.0054	0.0038	0.0025	0.0040	0.0017	0.0047
γ_{10}	0.0089***	0.0244***	-0.0022	-0.0117***	-0.0023	-0.0112***
	0.0025	0.0063	0.0041	0.0036	0.0061	0.0030
γ_{11}	-0.0023	0.0135***	0.0004	0.0078**	-0.0068***	-0.0044*
	0.0024	0.0048	0.0013	0.0037	0.0021	0.0025

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.0064***	0.0012	0.0018***	0.0036***	0.9589***
	0.0012	0.0008	0.0007	0.0008	0.0045
β	0.0000	-0.0041***	-0.0002	-0.0088***	0.0041***
	0.0024	0.0013	0.0011	0.0013	0.0014
γ_1	-0.0068	0.0298***	-0.0040	-0.0129***	0.0143**
	0.0054	0.0049	0.0030	0.0033	0.0057
γ_2	-0.0016	-0.0042	-0.0072**	-0.0013	0.0042
	0.0033	0.0032	0.0036	0.0050	0.0161
γ_3	0.0320***	0.0048	0.0099***	0.0197***	-0.1847***
	0.0063	0.0043	0.0034	0.0038	0.0168
γ_4	0.0259***	-0.0091***	0.0031*	-0.0083**	0.0046
	0.0026	0.0032	0.0018	0.0033	0.0029
γ_5	0.0168***	-0.0060*	-0.0051**	0.0033	-0.0006
	0.0037	0.0031	0.0022	0.0024	0.0024
γ_6	0.0158**	0.0067	-0.0020	0.0002	-0.0068
	0.0063	0.0047	0.0019	0.0041	0.0049
γ_7	-0.0937***	0.0028	0.0103**	0.0073***	0.0266***
	0.0135	0.0039	0.0042	0.0028	0.0057
γ_8	-0.0015	0.0222***	0.0012	0.0000	-0.0040**
	0.0034	0.0016	0.0013	0.0010	0.0018
γ_9	-0.0022	0.0008	-0.0024	-0.0030	-0.0045
	0.0031	0.0037	0.0033	0.0026	0.0055
γ_{10}	0.0051	-0.0221***	0.0039	0.0200***	-0.0127**
	0.0044	0.0046	0.0046	0.0040	0.0065
γ_{11}	0.0050*	-0.0014	0.0002	-0.0068***	-0.0054***
	0.0028	0.0016	0.0016	0.0015	0.0014
logL	59388.33				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Estimates of restricted QUAIDS

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.0151***	0.0108***	0.0017***	0.0136***	0.0048***	0.0087***
	0.0047	0.0031	0.0004	0.0034	0.0007	0.0033
β	0.0506***	0.0370***	0.0087***	0.0458***	0.0155***	0.0277***
	0.0022	0.0025	0.0015	0.0006	0.0015	0.0017
γ_1	0.0111***	-0.0006	0.0038***	-0.0012***	-0.0191**	-0.0058
	0.0012	0.0006	0.0006	0.0003	0.0097	0.0103
γ_2	-0.0006	0.0006***	-0.0031***	-0.0004	-0.0058***	-0.0065
	0.0006	0.0001	0.0005	0.0071	0.0004	0.0072
γ_3	0.0038***	-0.0031***	-0.0123*	0.0080	0.0070	-0.0049***
	0.0006	0.0005	0.0064	0.0121	0.0049	0.0014
γ_4	-0.0012***	-0.0004	0.0080	-0.0025***	-0.0073	-0.0047***
	0.0003	0.0071	0.0121	0.0006	0.0059	0.0009
γ_5	-0.0191**	-0.0058***	0.0070	-0.0073	0.0155***	0.0037
	0.0097	0.0004	0.0049	0.0059	0.0037	0.0069
γ_6	-0.0058	-0.0065	-0.0049***	-0.0047***	0.0037	0.0140***
	0.0103	0.0072	0.0014	0.0009	0.0069	0.0030
γ_7	-0.0033***	-0.0089***	0.0023	0.0047	0.0019***	0.0020***
	0.0012	0.0032	0.0067	0.0113	0.0007	0.0003
γ_8	-0.0062	-0.0075***	0.0017***	-0.0106**	-0.0008***	-0.0002*
	0.0048	0.0011	0.0007	0.0051	0.0001	0.0001
γ_9	-0.0017*	-0.0076***	-0.0032	0.0029	0.0005	0.0003***
	0.0009	0.0022	0.0082	0.0081	0.0005	0.0001
γ_{10}	0.0021	0.0015	-0.0026**	-0.0092***	0.0043***	-0.0005*
	0.0066	0.0031	0.0012	0.0028	0.0006	0.0003
γ_{11}	0.0209**	0.0383***	0.0033***	0.0204**	0.0001	0.0025
	0.0094	0.0074	0.0006	0.0088	0.0057	0.0033
λ	-0.0039***	-0.0002**	-0.0005***	-0.0031***	-0.0016	-0.0026***
	0.0013	0.0001	0.0001	0.0010	0.0013	0.0010

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.0092***	0.0081***	0.0031**	0.0064***	0.9187***
	0.0031	0.0031	0.0015	0.0024	0.0114
β	0.0301***	0.0258***	0.0101***	0.0203***	-0.2715***
	0.0015	0.0007	0.0003	0.0015	0.0179
γ_1	-0.0033***	-0.0062	-0.0017*	0.0021	0.0209**
	0.0012	0.0048	0.0009	0.0066	0.0094
γ_2	-0.0089***	-0.0075***	-0.0076***	0.0015	0.0383***
	0.0032	0.0011	0.0022	0.0031	0.0074
γ_3	0.0023	0.0017***	-0.0032	-0.0026**	0.0033***
	0.0067	0.0007	0.0082	0.0012	0.0006
γ_4	0.0047	-0.0106**	0.0029	-0.0092***	0.0204**
	0.0113	0.0051	0.0081	0.0028	0.0088
γ_5	0.0019***	-0.0008***	0.0005	0.0043***	0.0001
	0.0007	0.0001	0.0005	0.0006	0.0057
γ_6	0.0020***	-0.0002*	0.0003***	-0.0005*	0.0025
	0.0003	0.0001	0.0001	0.0003	0.0033
γ_7	-0.0199***	-0.0008	0.0034	0.0019	0.0167***
	0.0057	0.0022	0.0087	0.0096	0.0011
γ_8	-0.0008	0.0227*	0.0003	-0.0009***	0.0023**
	0.0022	0.0130	0.0154	0.0002	0.0011
γ_9	0.0034	0.0003	-0.0005	-0.0004	0.0061***
	0.0087	0.0154	0.0207	0.0203	0.0022
γ_{10}	0.0019	-0.0009***	-0.0004	0.0070***	-0.0030
	0.0096	0.0002	0.0203	0.0025	0.0229
γ_{11}	0.0167***	0.0023**	0.0061***	-0.0030	-0.1075
	0.0011	0.0011	0.0022	0.0229	0.0034
λ	-0.0027***	-0.0024***	-0.0006***	-0.0022*	0.0197***
	0.0003	0.0009	0.0002	0.0013	0.0014
logL	58776.02				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Estimates of unrestricted QUAIDS estimated by Heien and Wessels' estimator

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.0121***	0.0123***	0.0095***	0.0104***	0.0129***	0.0066***
	0.0005	0.0003	0.0004	0.0004	0.0005	0.0004
β	0.0137***	0.0194***	0.0100***	0.0138***	0.0175***	0.0039***
	0.0014	0.0013	0.0022	0.0015	0.0026	0.0014
γ_1	0.0433***	-0.0126***	-0.0102***	-0.0001	-0.0476***	0.0037
	0.0038	0.0026	0.0027	0.0021	0.0030	0.0029
γ_2	0.0086***	0.0103***	-0.0005	0.0078***	0.0105***	-0.0107***
	0.0024	0.0018	0.0027	0.0021	0.0038	0.0032
γ_3	0.0200***	0.0207***	-0.0051**	0.0119***	0.0232***	-0.0089***
	0.0021	0.0015	0.0020	0.0021	0.0026	0.0020
γ_4	-0.0021	-0.0014	0.0050***	-0.0043	-0.0100***	0.0017
	0.0026	0.0032	0.0013	0.0036	0.0022	0.0027
γ_5	-0.0128***	-0.0243***	0.0108***	0.0034	0.0155***	0.0028
	0.0026	0.0035	0.0022	0.0026	0.0018	0.0036
γ_6	-0.0205***	-0.0028	-0.0033	-0.0073*	0.0033*	0.0363***
	0.0033	0.0019	0.0026	0.0039	0.0019	0.0030
γ_7	0.0070**	-0.0085***	0.0027	0.0117***	-0.0110***	0.0102***
	0.0027	0.0019	0.0030	0.0029	0.0040	0.0029
γ_8	-0.0158***	-0.0053**	0.0023***	-0.0122***	-0.0004	0.0029*
	0.0023	0.0026	0.0009	0.0034	0.0009	0.0017
γ_9	-0.0065*	0.0019	0.0014	0.0101***	0.0060**	-0.0023
	0.0034	0.0025	0.0020	0.0034	0.0026	0.0041
γ_{10}	0.0066***	0.0129***	-0.0046	-0.0046**	0.0021	-0.0086***
	0.0023	0.0011	0.0044	0.0020	0.0038	0.0026
γ_{11}	0.0030	0.0109***	-0.0003	0.0153***	-0.0045***	-0.0050***
	0.0021	0.0028	0.0013	0.0022	0.0014	0.0018
λ	-0.0018***	0.0013***	-0.0009***	-0.0012***	-0.0018***	-0.0008***
	0.0003	0.0003	0.0002	0.0003	0.0002	0.0002
δ			-0.0137***		-0.0256***	-0.0117***
			0.0011		0.0054	0.0031

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.0104***	0.0082***	0.0082***	0.0099***	0.8997***
	0.0005	0.0003	0.0003	0.0003	0.0021
β	0.0146***	0.0104***	0.0108***	0.0087***	-0.1229***
	0.0018	0.0019	0.0019	0.0023	0.0119
γ_1	-0.0006	0.0222***	-0.0030	-0.0050	0.0099***
	0.0033	0.0033	0.0031	0.0036	0.0036
γ_2	0.0008	-0.0008	-0.0064**	0.0035	-0.0232**
	0.0023	0.0020	0.0031	0.0029	0.0096
γ_3	0.0125***	0.0007	0.0006	0.0109***	-0.0865***
	0.0022	0.0017	0.0014	0.0017	0.0083
γ_4	0.0223***	-0.0091***	0.0015	-0.0125***	0.0087**
	0.0026	0.0028	0.0016	0.0034	0.0034
γ_5	0.0110***	-0.0096***	-0.0022	0.0057**	-0.0004
	0.0022	0.0035	0.0021	0.0023	0.0031
γ_6	0.0046**	0.0030	-0.0017	-0.0034	-0.0083**
	0.0021	0.0039	0.0022	0.0029	0.0035
γ_7	-0.0579***	-0.0051**	0.0110**	0.0008	0.0390***
	0.0038	0.0022	0.0045	0.0042	0.0065
γ_8	0.0033**	0.0238***	0.0015	0.0001	-0.0002
	0.0015	0.0018	0.0011	0.0010	0.0018
γ_9	-0.0061	0.0008	0.0000	-0.0034	-0.0019
	0.0038	0.0041	0.0024	0.0029	0.0055
γ_{10}	0.0014	-0.0140***	-0.0035	0.0134***	-0.0011
	0.0028	0.0028	0.0028	0.0046	0.0103
γ_{11}	0.0036	-0.0019	0.0018	-0.0042**	-0.0187***
	0.0023	0.0019	0.0017	0.0018	0.0044
λ	-0.0013***	-0.0010***	-0.0008***	-0.0014***	0.0098***
	0.0002	0.0002	0.0001	0.0002	0.0008
δ	-0.0090	-0.0106***	-0.0071***	-0.0117***	
	0.0071	0.0025	0.0014	0.0009	
logL	59520.50				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Estimates of unrestricted QUAIDS estimated by Shonkwiler and Yen's estimator

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.0275***	0.0308***	0.0320***	0.0276***	0.0292***	0.0231***
	0.0019	0.0024	0.0025	0.0021	0.0022	0.0016
β	0.0028	0.0185***	-0.0164***	0.0101***	0.0025	0.0049
	0.0019	0.0021	0.0056	0.0016	0.0041	0.0030
γ_1	0.0283***	-0.0030	-0.0072**	0.0031***	-0.0320***	0.0022
	0.0027	0.0026	0.0032	0.0009	0.0024	0.0019
γ_2	0.0063**	0.0075***	0.0022	0.0055***	0.0057	-0.0079***
	0.0029	0.0022	0.0022	0.0016	0.0044	0.0021
γ_3	0.0170***	0.0173***	0.0075**	0.0119***	0.0156***	-0.0028
	0.0027	0.0046	0.0032	0.0033	0.0028	0.0041
γ_4	0.0002	0.0028	0.0055**	-0.0005	-0.0066***	-0.0014
	0.0024	0.0031	0.0024	0.0023	0.0026	0.0019
γ_5	-0.0069***	-0.0110***	0.0107***	0.0032*	0.0159***	-0.0011
	0.0016	0.0031	0.0023	0.0018	0.0026	0.0029
γ_6	-0.0106***	0.0010	-0.0056***	-0.0028	0.0057**	0.0233***
	0.0028	0.0024	0.0021	0.0025	0.0027	0.0029
γ_7	0.0065***	-0.0017	0.0022	0.0091***	-0.0034	0.0066**
	0.0020	0.0025	0.0017	0.0013	0.0028	0.0026
γ_8	-0.0132***	0.0004	0.0070***	-0.0087***	0.0002	0.0023*
	0.0026	0.0022	0.0026	0.0019	0.0020	0.0013
γ_9	-0.0015	0.0041	-0.0017	0.0090***	0.0110 ***	-0.0020
	0.0022	0.0027	0.0026	0.0020	0.0014	0.0047
γ_{10}	0.0051**	0.0084***	-0.0005	-0.0021	0.0021	-0.0070***
	0.0022	0.0016	0.0024	0.0013	0.0026	0.0026
γ_{11}	0.0076***	0.0032	-0.0030	0.0114***	0.0010	-0.0032
	0.0022	0.0028	0.0033	0.0015	0.0023	0.0025
λ	-0.0008	-0.0012	0.0005	-0.0012	-0.0016***	-0.0004
	0.0009	0.0021	0.0008	0.0007	0.0005	0.0007
δ			0.0175***		0.0088***	0.0099***
			0.0022		0.0009	0.0010

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.0239***	0.0234***	0.0251***	0.0262***	0.0200***
	0.0015	0.0017	0.0019	0.0020	0.0014
β	0.0116***	0.0060*	0.0184***	-0.0016	0.0049
	0.0013	0.0034	0.0025	0.0045	0.0042
γ_1	0.0002	0.0132***	-0.0039	-0.0051***	0.0059***
	0.0026	0.0033	0.0028	0.0017	0.0017
γ_2	-0.0016	-0.0030	-0.0043	0.0025	-0.0044**
	0.0023	0.0021	0.0026	0.0036	0.0021
γ_3	0.0084***	0.0012	0.0033	0.0117***	0.0046
	0.0019	0.0028	0.0037	0.0026	0.0028
γ_4	0.0134***	-0.0086***	0.0003	-0.0144***	0.0026
	0.0020	0.0023	0.0014	0.0024	0.0022
γ_5	0.0054*	-0.0091**	-0.0030	-0.0011	-0.0048**
	0.0029	0.0040	0.0026	0.0026	0.0021
γ_6	0.0013	-0.0001	-0.0026	-0.0052**	-0.0068***
	0.0020	0.0026	0.0020	0.0026	0.0019
γ_7	-0.0346***	-0.0040*	0.0066	0.0002	0.0013
	0.0027	0.0023	0.0043	0.0027	0.0029
γ_8	-0.0003	0.0189***	0.0032*	-0.0001	-0.0047***
	0.0013	0.0031	0.0017	0.0020	0.0016
γ_9	-0.0048	-0.0004	-0.0013	-0.0052*	-0.0088***
	0.0033	0.0039	0.0018	0.0027	0.0033
γ_{10}	-0.0010	-0.0106***	-0.0027	0.0088***	-0.0072***
	0.0026	0.0017	0.0021	0.0028	0.0017
γ_{11}	0.0009	-0.0008	0.0009	-0.0062**	-0.0131***
	0.0022	0.0037	0.0026	0.0032	0.0036
λ	-0.0003	0.0001	-0.0022***	0.0010	0.0019***
	0.0009	0.0011	0.0006	0.0008	0.0005
δ	0.0112***	0.0101***	0.0136***	0.0146***	0.0087***
	0.0009	0.0007	0.0012	0.0013	0.0028
logL	146036.76				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Estimates of unrestricted QUAIDS estimated by Cosslett's estimator

	Bakery	Meat	Fish	Dairy&eggs	Fats	Fruit
α	0.0206***	0.0213***	0.0193**	0.0204***	0.0204***	0.0179
	0.0033	0.0037	0.0095	0.0041	0.0040	0.0125
β	0.0075***	0.0179***	-0.0141***	0.0131***	0.0011	0.0054***
	0.0018	0.0013	0.0034	0.0015	0.0023	0.0018
γ_1	0.0289***	-0.0054***	-0.0060***	0.0068***	-0.0282***	0.0055***
	0.0034	0.0012	0.0019	0.0010	0.0026	0.0018
γ_2	0.0117***	0.0053***	-0.0002	0.0115***	0.0101***	-0.0012
	0.0025	0.0010	0.0025	0.0014	0.0023	0.0016
γ_3	0.0136***	0.0058***	-0.0049**	0.0099***	0.0126***	-0.0040***
	0.0019	0.0014	0.0020	0.0022	0.0020	0.0013
γ_4	0.0021	-0.0033**	0.0045*	0.0018	-0.0067***	0.0004
	0.0021	0.0013	0.0024	0.0027	0.0018	0.0020
γ_5	-0.0058**	-0.0171***	0.0090**	0.0042**	0.0140***	-0.0006
	0.0028	0.0028	0.0040	0.0017	0.0018	0.0024
γ_6	-0.0066***	-0.0021	-0.0048*	0.0007	0.0053***	0.0246***
	0.0025	0.0021	0.0029	0.0034	0.0015	0.0023
γ_7	0.0113***	-0.0018*	0.0051***	0.0139***	0.0019	0.0118***
	0.0015	0.0011	0.0024	0.0016	0.0031	0.0025
γ_8	-0.0120***	-0.0047*	0.0033	-0.0067**	0.0012	0.0031***
	0.0028	0.0027	0.0031	0.0028	0.0016	0.0009
γ_9	0.0021	0.0006	0.0015	0.0120***	0.0126***	0.0019
	0.0014	0.0022	0.0018	0.0026	0.0009	0.0030
γ_{10}	0.0115***	0.0082***	-0.0001	0.0056***	0.0088***	0.0007
	0.0017	0.0019	0.0023	0.0013	0.0026	0.0029
γ_{11}	0.0108***	0.0002	0.0007	0.0143***	0.0068***	0.0014
	0.0017	0.0021	0.0031	0.0017	0.0015	0.0025
λ	-0.0036	0.0030***	0.0013	-0.0045*	-0.0030	-0.0032
	0.0022	0.0011	0.0059	0.0025	0.0020	0.0033

	Vegetables	Sugars	Others	Coffee etc.	Non-alco
α	0.0166	0.0123***	0.0104***	0.0108***	0.0125***
	0.0132	0.0035	0.0033	0.0039	0.0021
β	0.0115***	0.0054**	0.0109***	0.0070***	0.0076***
	0.0015	0.0022	0.0019	0.0018	0.0008
γ_1	0.0043*	0.0160***	-0.0009	-0.0032	0.0057***
	0.0025	0.0018	0.0028	0.0025	0.0016
γ_2	0.0049***	0.0024*	-0.0018	0.0057***	-0.0005
	0.0015	0.0013	0.0021	0.0017	0.0027
γ_3	0.0067***	-0.0019	-0.0041*	0.0062***	0.0002
	0.0016	0.0016	0.0024	0.0022	0.0027
γ_4	0.0141***	-0.0063***	0.0002	-0.0150***	-0.0005
	0.0012	0.0024	0.0018	0.0012	0.0014
γ_5	0.0049*	-0.0077***	-0.0031*	-0.0017	-0.0071***
	0.0027	0.0023	0.0017	0.0022	0.0022
γ_6	0.0046***	0.0033	-0.0008	-0.0036	-0.0054*
	0.0015	0.0026	0.0015	0.0026	0.0027
γ_7	-0.0257***	0.0014	0.0113***	0.0022	0.0033***
	0.0036	0.0014	0.0023	0.0021	0.0011
γ_8	0.0015	0.0193***	0.0003	-0.0006	-0.0063**
	0.0017	0.0016	0.0013	0.0024	0.0026
γ_9	-0.0013	0.0029	0.0011	-0.0047**	-0.0084***
	0.0025	0.0029	0.0011	0.0022	0.0027
γ_{10}	0.0065**	-0.0031**	0.0006	0.0121***	-0.0014
	0.0033	0.0014	0.0011	0.0027	0.0020
γ_{11}	0.0053***	0.0026	0.0013	-0.0078***	-0.0188***
	0.0020	0.0018	0.0014	0.0017	0.0015
λ	-0.0030	-0.0020	-0.0012	-0.0001	0.0022
	0.0026	0.0026	0.0021	0.0017	0.0016
logL	138565.62				
N	2903				

Standard errors under the estimates, calculated by 500 bootstrap replications
Significance of coefficients' estimates: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Cosslett's dummies are not reported, most of their coefficients are significant on 1 % level.