Ph.D. Thesis

Empirical studies on

"Health, Information and Consumer Behaviour"

by

Sinne Smed

AKF, Danish Institute of Governmental Research

and

University of Copenhagen, Institute of Economics

May 2008

Foreword and acknowledgements

This thesis consists of four papers about the effect of prices and information on the demand for healthy food plus an additional paper on data documentation. Each of these papers is independent and can be read separately; however, thematically and/or theoretically they are closely related and they are all based on the same panel dataset. One paper has been written together with Laura M. Andersen (Ph.D. student at AKF) and another paper with Martin Browning (Oxford) and Lars G. Hansen (AKF). The introduction provides a brief view into the topics of obesity, health, food demand and regulation of consumer behaviour. A short summary of each chapter is also provided in the introduction as well as differences and similarities between the papers.

I am thankful to my supervisor Martin Browning for valuable inspiration and guidance. I also especially want to thank Laura M. Andersen and Lars G. Hansen at AKF for great help and frequent discussions and contributions to my work. Derek Baker also deserves thanks for reading and commenting on this thesis. Furthermore, I would like to thank a bunch of current and former colleagues at AKF and FØI for help, comments and support. A special thanks to Ida, Sofie, Lisbeth, Søren and Markus for their patience, help and humour.

The majority of the work has been undertaken while I was employed at AKF (Danish Institute of Governmental Research), part of it while being employed at the Institute of Food and Resource Economics, Faculty of Life Sciences, University of Copenhagen.

Table of Content

Introduction

Today more than one billion adults globally are defined to be overweight, at least 300 million obese and the increase in the prevalence of obesity the last decades has been so large that it has been termed an epidemic (WHO, 2003). Furthermore, the prevalence of obesity and diet related illnesses have a social gradient as there is an inverse relation between income and education on one side and obesity and diet related illnesses on the other (Brønnum-Hansen et al., 2004; Lynch et al., 2006). The same social gradient is found in dietary patterns (Worsley, 2003, Groth et al., 2001). Diet composition is found to influence the prevalence of obesity (Swinburn et al., 2004; Binkley et al., 2000; Prentice and Jebb, 2003; WHO, 2007) as well as the link between obesity and chronic conditions such as cardiovascular diseases, some types of cancer, diabetes and osteoporosis are well established (see e.g. US Department of Health and Human services, 2003; WHO, 2007). The increase in obesity and diet related illnesses poses a considerable challenge to governments since as much as 2-6 per cent of the total direct health care costs in several developed countries are caused by obesity and 10-15 per cent of lost years of life in Europe can be attributed to poor nutrition (WHO, 2003). In countries with publicly financed health services or with privately financed health services, but with equally distributed rates of payment for everybody, this means that the unhealthy eating or lack of exercise of one citizen may impose economic burdens, externalities, on other citizens due to increased health care costs. Obesity is important from a health and welfare economic perspective, but also from a social perspective. For the individual, multiple studies have shown that obesity affects personal and working relations, earnings and wages in a negative and statistically significant way (see e.g. Harper, 2000; Cawley, 2004).

Basically the cause of obesity and diet style related illnesses seems rather simple – too much energy taken in through foods and drinks, too little energy spent in exercise and the excess stored as adipose tissue. However, the underlying determinants of this energy imbalance are very complex and are the subject of research in many areas ranging from genetics over psychology and sociology to economics. The increase in calorie consumption has been addressed to among other things an increasing consumption of convenience and fast food (Chou et al., 2004; French et al., 2000; Schwartz and Brownell, 2005), increasing work hours by women (Scholder, 2007; Bowers D.E., 2000; Anderson, 2003), increasing portion sizes in both pre-packaged food and in restaurants and fast food outlets (French et al., 2001; Wansink B., 2004) and massive advertising for unhealthy foods (French et al., 2001; Haddad, 2003). At the same time large shifts towards less physically demanding work, increasing use of automated transport and technology at home and more passive leisure pursuits resulting in a more sedentary lifestyle have been observed worldwide (WHO, 2003; Haddad, 2003; Cutler et al., 2003). Most of these changes can indirectly be related to technological developments that have reduced the economic incentive for a healthy balance between food intake and physical activity by lowering the costs of acquiring calories and increasing the costs of expending these calories (Philipson et al., 2004; Lakdawalla et al., 2005). Furthermore, the industrial production of food has eased a change in diets from vegetables and fibres towards more convenience foods and prepared fast food meals with a higher content of saturated fats and sugar. (Haddad, 2003; WHO, 2003; Cutler et al., 2003).

The most widely used instrument to counteract bad nutrition and obesity has been information campaigns and labelling and the effects of these on food demand are well established (see e.g. Kim and Chern, 1999; Rickertsen et al., 2003; Teisl et al., 2001; Guthrie et al., 1995), even though low educated and low income consumers react less and differently to these information campaigns. The reason for this social gradient is less well explained. Other suggested, but less widely used measures include for example tighter rules for advertising of unhealthy foods, promotion of healthier eating at schools and exercise on prescription (Finkelstein et al., 2004; WHO, 2007). A range of incentive-based economic instruments to adjust dietary and exercise patterns has been discussed. Direct instruments such as a BMI tax or BMI graduated health care costs have been suggested as effective (see for example Bhattacharya and Sood, 2007), but are not considered politically and socially acceptable. Prices seem to be a determinant for the choice of healthy foods, especially among low income and low educated groups (Kearney and McElhone, 1999; Lennernas et al., 1997; Smed et al., 2007). Indirect instruments such as modified food taxes or subsidies have therefore been considered as a way of reducing the negative effect of low income on poor diet quality which has been suggested in several studies (Glanz, 1998; Darmon et al., 2002; Finkelstein et al., 2004; WHO, 2007).

Despite this large body of literature concerning the effect of information, budget and prices on food consumption there is a lack of studies based on long-term individual panel data including both economic and nutritional variables. Studies based on macro-data describe average changes in food consumption, but not variability and cannot address the social and demographic differences to the same extent as micro data. The effect of changes in economic variables on food consumption cannot be addressed properly through studies based on cross-section data. The papers in this thesis are based on a dataset that comprises nutritional and economic data together with social and demographic variables. This allows us to infer more directly the effects of information, prices and budgets on nutrient composition and to reveal consumer preferences for nutrients in food. Furthermore, the panel consists of a large number of individuals followed over a long time, so changes in dietary patterns can be followed and the economic variables can be used to explain some of these differences and developments. The following section provides a summary of each of the papers. Paper 1 and 2 are mainly descriptive while paper 3, 4 and 5 draw on economic theory. Joint for the papers is that they consider demand for food as demand for the characteristics inherent in foods, not demand for the food itself, i.e. they draw on the characteristics model originally developed by Gorman (1980) and Lancaster (1966). Some of the articles focus on the composition of the whole diet, while others take out subsets to describe the effect of prices and information in more detail.

Paper 1: *Describing dietary patterns from purchase data – a data description* justifies how purchase data can be used to describe food and nutrient consumption. The use of purchase data implies that variables on nutrition are measured together with economic variables like prices and expenditure. This differs from dietary recall and other record methods which only collect data on consumption, not on prices. Purchase data for approximately 2500 Danish households in terms of quantity and value are collected from 1997 to 2004 at a very detailed level and these are concatenated with detailed nutrition matrices. The purchased foods are registered by exact weight measures diminishing problems of inaccuracy incurred by using reference categories. Furthermore, many households stay in the panel for several years giving a unique possibility to follow changes in dietary patterns over time together with changes in prices and expenditure. The purchase data include a wide range of background information about social and demographic characteristics of the individual households including height, weight and exercise questions for each individual in the household, as well as the households' media habits. The raw data can be weighted to represent the Danish population. These are used to describe average changes in food and nutrition composition in Danish households.

In **Paper 2**: *Measuring the health performance of diets* a Healthy Eating Index (HEI) is developed, that measures the health performance of diets in terms of compliance to multiple health recommendations simultaneously. Consumers' compliance to each of the dietary recommendations is measured individually and then weighted together to one measure using a Euclidian distance measure. By using the panel dataset constructed in paper 1 the HEI gives us the possibility of following the development in the health performance of diets in relation to the official Danish dietary recommendations from 1999 to 2004. The results, mostly based on descriptive statistics, show that only a minor part of the population fulfils all the diet recommendations. The relationship between dietary quality and BMI is confirmed, and the results suggest that observed differences in BMI between men and women might be caused by healthier dietary patterns among women than among men. A social gradient is identified, revealing that shorter educated eat less healthy than longer educated, and older households are found to be closer to fulfil the diet recommendations than younger households. The effect of age is so dominating that the best educated of the younger households are at the same level of health performance as the shortest educated of the older. Low educated younger households also seem to have smaller improvements in the health performance in the studied period than other types of households, suggesting that the social bias in dietary patterns persists. From a political perspective the identification of the differences in dietary patterns between age groups is important since the consequences of an inexpedient lifestyle in a young age are revealed later in life. To counteract the social bias in obesity and diet related illnesses the young and especially the low educated young have to be targeted.

In **Paper 3**: *Valuation of health* the differences in dietary patterns described in paper 2 is explained by estimating differences in valuation of health characteristics for various types of households. This is done in a hedonic price framework. Due to the fact that our data are panel data, it is possible to remove individual heterogeneity from the estimates of the valuation of health, contrary to most other studies estimating hedonic models. The richness of the data furthermore allows us to control for the influence of preferences for non-nutritional characteristics as e.g. taste. Consumers' valuation of health is estimated in six dimensions each representing one of the official Danish diet recommendations. Consumers are found to have preferences for energy dense foods. A positive correlation between the valuation of health and the valuation of the non-nutritional characteristics are found indicating that consumers either have a general high valuation or a general low valuation of all the characteristics in food. Under certain assumptions, the implicit price of the characteristics estimated in the hedonic model can be interpreted as exogenous. Using this, a positive correlation between healthiness of diets and expenditure is found, suggesting that cost might be a barrier, for some households, to change towards a healthier diet. Cost might especially be a barrier for increasing the consumption of fruit and vegetables. The identification of the relation between expenditure and a healthy diet might add to the discussion of regulation of consumer behaviour through the relative prices of healthy and unhealthy foods.

In **Paper 4**: *A censored structural characteristics model for milk* we turn away from the whole diet, and concentrate on one good from the dataset, fluent milk. In the paper we investigate preferences for fat in milk through a structural characteristics model. Since, on average, 5.7% of the total consumption of saturated fat comes from milk, changed consumption patterns for milk might be important in terms of health. The derivation of a structural model for the individual household brings us closer to separating preferences and changes in these due to e.g. information from reactions to prices and budget constraints. Contrary to the usual hedonic models, consumers' preferences over characteristics are here allowed to vary non-systematically through an error term placed on the structural parameters in the utility function. The functional form used is the quadratic form allowing the marginal utility of characteristics to become negative. In the empirical estimations we use the dataset constructed in paper 1 spanning the period from 1997 to 2004 and includes newspaper information about the link between fat consumption and health. The panel structure of the data is exploited fully since the suggested models are estimated household by household allowing for the maximum degree of individual heterogeneity. We find that a model with measurement errors performs better than a model with random parameters on the structural parameters and this allows us to formulate the final model as a two-sided censored Tobit model. We find that there has been a significant decrease in the consumption of fat from milk without any essential decrease in the consumption of milk. This decrease is generated by systematic changes in preferences due to a general trend and information. Higher educated households are found to prefer milk with a lower fat content than lower educated, but for older households this difference seems to disappear over time. This supports the findings from paper 3, which showed that especially younger low educated households should be targeted to improve average dietary patterns. In the discussion of whether to use price policy or information as an instrument to decrease the consumption of fat from milk, the price policy seems the most effective. Consumers who prefer milk with a very high fat content can be reached both by information and prices, while consumers who prefer milk with a moderate to high fat share are not influenced by information. However, they are rather price sensitive. This is of great importance since households that drink a lot of milk prefer milk with a moderate to high fat share.

The effects of information are further explored in **Paper 5**: *Information processing strategies and health information* that addresses how information is processed and transferred into changed consumption patterns. To our knowledge this is the only micro-data study analysing differences in consumers' information processing strategies. Again a subset of goods, fish, is taken out of the dataset from paper 1 to focus on the effects of positive and negative information on demand for fish. Information about the positive nutritional effects of fish applies generally to both fatty and lean fish. Fatty fish are substantially more susceptible to dioxin poisoning than lean fish since dioxin accumulates in fatty tissue. Some consumers may therefore know that it is possible to avoid the risk of dioxin poisoning by substituting away from fatty fish and toward lean fish without cutting back on total fish consumption (a sophisticated reaction). Other consumers may not be able to make the distinction between fatty and lean fish and may instead substitute away from all types of fish (unsophisticated reaction). A third way of reaction may be to ignore the information (ignorance). We assume that consumers, prior to our data period, choose a strategy about how to react to information concerning the health characteristics in fish. This assumption is based on the theoretical literature suggesting that consumers in daily routine purchasing situations choose a simple "rule of thumb" strategy as opposed to a more systematic information processing strategy used in situations of great relevance to the consumers. Consumers' choice of strategy is determined by estimating a two-stage demand system (AIDS) for each household, with new information treated as an adjustment to the prices. We find that approximately half of the consumers ignore the negative information, while two thirds choose to ignore the positive information. Conditional on reacting to the negative information, half of the consumers choose a sophisticated strategy. Based on the initial estimations a Probit is estimated to identify consumer characteristics that seem important for the choice of strategy. We find that especially age and education are positively correlated with the probability of choosing a sophisticated strategy while the volume share for fish heavily influences the probability of reacting to both negative and positive information. This way of categorising consumers according to information strategies should be of general interest to policy makers and marketing strategists, when designing and implementing health campaigns or other attempts to regulate or change food consumption behaviour. The need to focus on the comprehensibility frame of the target groups is important since consumers only react to information which is relevant to them and might choose a reaction strategy which is inexpedient in relation to general health.

References

Anderson P.M., K.F. Butcher and P. B. Levine (2003): Maternal employment and overweight children. *Journal of Health Economics*, 22(3): 477 - 504

Bhattacharya, J. and N. Sood (2007): Health insurance and the obesity externality. In J. Cawley and K. Bolin (eds.): The economics of obesity. Pp. 279-381 Elsevier ldt. Oxford

Binkley J.K., J. Eales and M. Jekanowski (2000): The relation between dietary change and rising US obesity. *International Journal of Obesity*, 24(8): 1032-1039

Bowers D. E. (2000): Cooking trends echo changing roles of women. *Food Review*, 23(1): 23-29

Brønnum-Hansen H., O. Andersen, M. Kjøller and N. Rasmussen (2004): Social gradient in life expectancy and health expectancy in Denmark. *Social and Preventive Medicin* 49(1): 36-41

Cawley J. (2004): The Impact of obesity on wages. *The Journal of Human Resources* 39(2): 451-474.

Chou S.Y., M. Grossmann and H. Saffer (2004): An economic analysis of adult obesity: Results from the behavioral factor surveillance system. *Journal of Health Economics*, 23(3): 565-587

Cutler D.M., E.L. Glaeser and J.M. Shapiro (2003): Why have Americans become more obese. *Journal of Economic Perspectives*, 17: 95-118

Darmon N., E.L. Ferguson and A. Briend (2002): A cost constraint alone has adverse effects on food selection and nutrient density: An analysis of human diets by linear programming. *Journal of Nutrition,* 132:3764 –3771

Finkelstein E., S. French, J.N. Variyam and P.S. Haines PS (2004): Pros and cons of proposed interventions to promote healthy eating. *American Journal of Preventive Medicine*, 27(3): 163-171

French S.A., L. Harnack and R.W. Jeffery (2000): Fast food restaurant use among woman in the Pound of Prevention study: dietary, behavioural and demographic correlates *International Journal of Obesity*, 24: 1353-1359

French S.A., M. Story and R.W. Jeffery (2001): Environmetal influences on eating and physical activity. *Annual Review of Public Health,* 22: 309-335

Glanz K. (1998): Why Americans eat what they do: taste, nutrition, cost, convenience and weight control concerns as influences on food consumption. *Journal of American Dietetic Association*, 98(10):1118-1126.

Gorman W.M. (1980): A possible procedure for analysing quality differentials in the egg market. *Review of Economic Studies*, 47(5): 843-856

Groth, M. V., S. Fagt, og L. Brøndsted (2001): Social determinants of dietary habits in Denmark. *European Journal of Clinical Nutrition*, 55(11): 959-966

Guthrie J.F., J.J. Fox, L.E. Cleveland and S. Welch (1995): Who uses nutrition information labelling and what effect does label use have on diet quality? *Journal of Nutrition education*, 27: 163-172

Haddad L. (2003): Redirecting the diet transition: what can food policy do? *Development Poliy Review*, 21(5-6): 599-614

Harper B. (2000): Beauty, stature and the labour market: a British Cohort Study.*Oxford Bulleting of Economics and Statistics*, 62(s1):771-800.

Kearney J.M. and S. McElhone (1999): Perceived barriers in trying to eat healthier: Results of a pan-EU consumer attitudinal survey. *British Journal of Nutrition*, 81: S133-S137

Kim S. and W.S. Chern (1999): Alternative measures of health information and demands for fats and oils in Japan. *Journal of Consumer Affairs*, 33(1): 92-109

Lakdawalla, D., T. Philipson and J. Bhattacharya (2005): Welfare-enhancing technological change and the growth of obesity. *American Economic Review* 95(2): 253-257

Lancaster K.J. (1966): A new approach to consumer theory. *Journal of Political Economy*, 74(2): 132-157

Lennernäs M., C. Fjellström, W. Becker, I. Giachetti, A. Schmitt, A.M. Renaut de Winter and M. Kearney (1997): Influences on food choice perceived to be important by nationally representative samples of adults in European Union. *European Journal of Clinical Nutrition*, 51(suppl. 2): S8-S15

Lynch J., G. D. Smith, S. Harper and K. Bainbridge (2006): Explaining the social gradient in coronary heart disease: comparing relative and absolute risk approaches. *Journal of Epidemiology and Community Health*, 60: 436-441

Philipson T., C. Dai and L. Helmchen and J. Variyam (2004): *The economics of obesity: A report on the workshop held at USDA's Economic Research Service.* Electronic publication from the food assistance and nutrition Research program Economic research service USDA, Wasington DC

Prentice A.M. and S.A. Jebb (2003): Fast food, energy density and obesity, A possible mechanistic Link. *Obesity Reviews*, 4(4): 187- 94

Rickertsen, K., D. Kristofersson, and S. Lothe (2003): Effects of Health Information on Nordic Meat and Fish Demand. *Empirical Economics*, 28(2): 249-273

Schwartz M. and K.D. Brownell (2005): *The need for courageous action to prevent obesity*. In Crawford D. and R.W. Jeffery (eds): Obesity prevention and public health. Oxford University Press, Oxford.

Scholder S.H.K. (2007): *Maternal employment and overweight children: does timing matter?* HEDG Workingpaper no. 07/12, The University of York, UK

Smed, S., J. D. Jensen and S. Denver (2007): Socio-economic characteristics and the effect of taxation as a health policy instrument. *Food Policy*, 32(5-6): 624-639.

Swinburn B.A., I. Caterson, J.C. Seidell and W.P.T. James (2004): Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutrition*, 7:123-46

Teisl M.F., N. E. Bockstael and A. Levy (2001): Measuring the welfare effects of nutrition information. *American Journal of Agricultural Economics*, 83(1): 133-149

U.S. Department of Health and Human Services (2003): *Statistics related to overweight and obesity*. NIH Publication, No. 03-4158, Washington DC

Wansink B. (2004): Environmental factors that increase the food intake and consumption volume of unknowing consumers. *Annual Review of Nutrition*, 24: 455-479

WHO (2003): Obesity and overweight, Global Strategy on Diet. *Physical Activity and Health* March [http://www.who.int/hpr/NPH/docs/gs_obesity.pdf]

WHO (2007): The challenge of obesity in the WHO European Region and the strategies for response, WHO, Copenhagen

Worsley A., R. Blasche, K. Ball and D. Crawford (2003): Income differences in food consumption in the 1995 Australian National Nutritional Survey. *European Journal of Clinical Nutrition*, 57:1198-1211

Chapter 1

Describing dietary patterns from purchase data

- A data description

Sinne Smed

AKF
(Danish Institute of Governmental Research)
Copenhagen

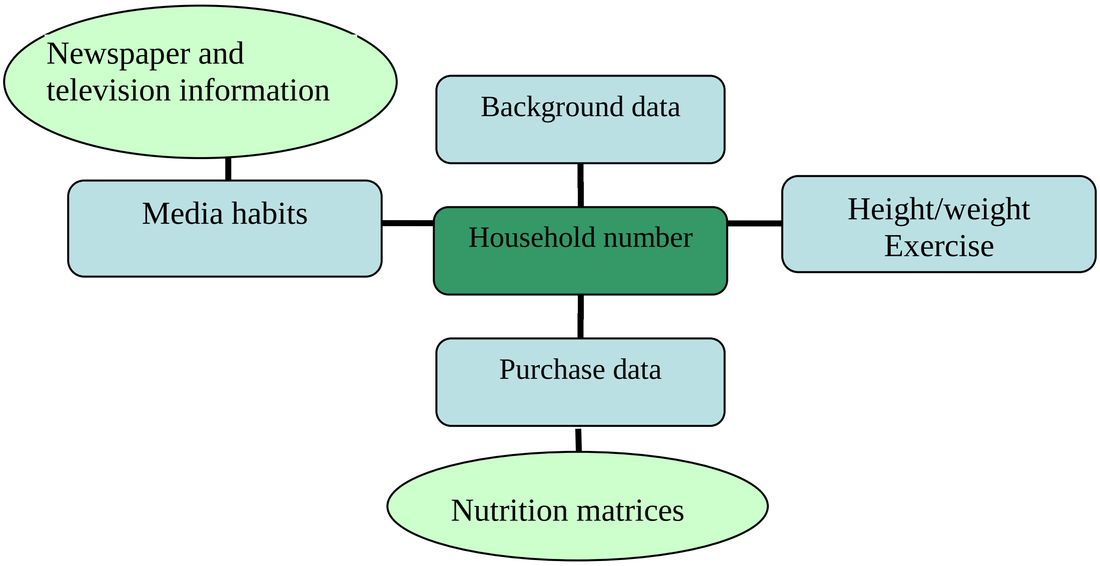
Abstract

This paper, which is purely descriptive, suggests an alternative method to dietary recall and record methods to collect data on nutrient consumption patterns. Detailed weekly purchase data for on average 2500 Danish households in terms of quantity and value is collected from 1997 to 2004 and combined with extensive nutrition matrices. This approach results in a dataset at household level where the nutritional composition of purchases are measured together with prices and expenditure. Social and demographic characteristics including height, weight and exercise questions are included in the dataset. The purchased foods are registered by exact weight measures diminishing problems of inaccuracy incurred by using references categories, as is standard in most dietary recall or record studies. Furthermore many households stay in the panel for several years yielding a unique possibility to follow changes in dietary patterns over time, together with changes in prices and expenditure. The data, weighted to represent the Danish population, show an increasing share of total energy intake from carbohydrates and a declining share from fat over time. Seasonality is seen for the consumption of all nutrients except fibres and protein. An educational gradient in Body Mass Index is found together with a positive correlation between Body Mass Index and the degree of inactivity.

Introduction

Whenever food and nutrient consumption patterns are described the most frequent method used for data collection is one or more days of dietary records or recall followed by social and demographic questions. These methods have the advantage that the individual's food consumption is described in detail. There are, however, several disadvantages involved in this form of data collection. One disadvantage is that often the respondents base their type-of-food and quantity-of-food measurements on comparison with reference categories, which will impose measurement errors in the data. Another disadvantage is that it is inconvenient for the respondents to register their food intake in such detail, which limits the length of the time in which the individuals can be followed. Furthermore these data lack economic variables (like e.g. prices), which is a necessity to describe the determinants of food demand. The data constructed in this study provides an alternative method to approximate food and nutrient consumption patterns, which overcome some of the problems in the methods above. In our approach household purchase data in terms of quantity and value are collected at a very detailed level which allows us to concatenate it with likewise detailed nutrition matrices. This implies that economic variables are measured together with detailed nutritional observations. The registration of purchase data may in some aspects appear less burdensome for the individual households than food registration, which implies that more households are willing to register for a longer period of time. This gives the possibility to explore the panel dimension in the data. Furthermore, exact weight measures of the purchased data are registered. The drawbacks of our approach are the impossibility to infer who consumes what in the household and how much of the purchased food that the households discard which introduces other sources of inaccuracy in the data than the dietary record or recall methods. The purchase data are combined with a wide range of background information about social and demographic characteristics of the individuals in the households, as well as attitudes and media habits. Included in the background data are also height and weight for the individual household members as well as exercise data. Figure 1.1 gives a summary of the data and the ways in which they can be combined. The background data, the media habits, the height and weight and the purchase data can all be combined through a household number. The nutrition data can be combined with purchase data through the amount of each type of food purchased, implying that the specific amount of nutrients purchased is known. Furthermore, through the information on the households' media habits it is possible to combine e.g. information from newspapers and television concerning health and nutritional issues with purchase data. This gives the possibility to exploit the influence of prices and information on household dietary choices simultaneously.

Figure 1. 1: Survey of purchase, nutrition and information data



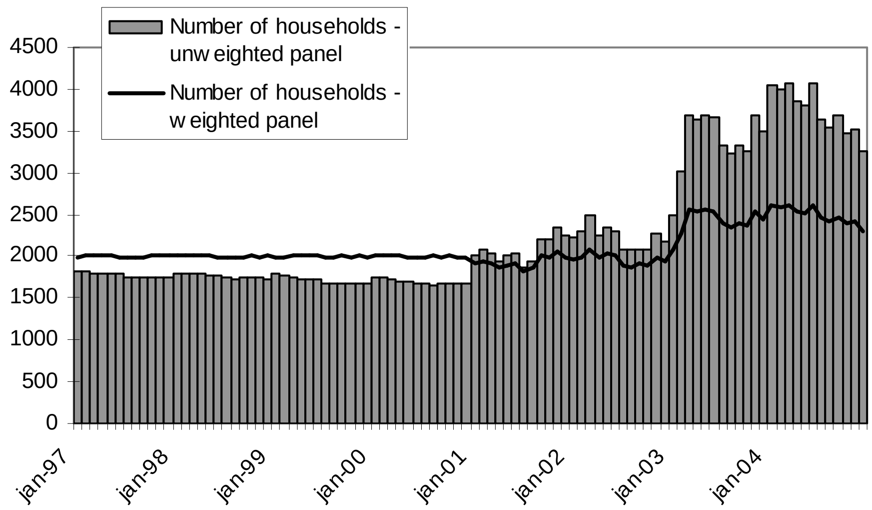
The rest of this paper is structured as follows: Section 1.1 gives a description of the purchase data, how they are collected and which background data are included. Section 1.2 gives a description of how the data are concatenated with nutrition matrices to reflect household purchases of nutrients. Section 1.3 gives a brief overview of how the data can be used to describe changes in nutrient and food consumption in Denmark. Section 1.4 is a description of the height, weight and exercise data. For likewise detailed, but differently focused descriptions of the dataset, see Smed (2002) or Andersen (2006).

1.1 Purchase data

Technical details

The purchase data are provided by GfK Denmark, which maintains, among other activities, a consumer panel. Households in the panel report purchases of foods and other staples in terms of quantity, price and other product characteristics. The diary is filled in by the diary keeper and is sent to GfK on a weekly basis. In principle the diary is filled in immediately after each shopping trip. The diaries were filled in by hand. GfK controls that the diaries are correctly completed and the diaries are controlled for consistency. Additionally, the households complete an annual questionnaire on their background, including social and demographic characteristics (family size, age, number of children, level of education, region, income etc.), weight, height [1] and media habits (e.g. preferred newspapers and magazines and frequency of reading) and several attitude questions. The purchase data and the background data are described in more detail below. The columns in Figure 1.2 show the monthly number of households handing in diaries. In 2003 the number of households increased considerably. For market analyse purposes GfK weigh the households[2] to make sure that the panel are in agreement with the social and demographic composition of the Danish population. From 1997 to 2000 the panel is weighted by doubling the performance of some of the underrepresented types of households. From 2001 and onwards more households than needed are recruited to the panel and each household is multiplied with a number around 1 to obtain the right composition of the panel. One household may e.g. weigh 0.98 in one month and 1.09 in the next. The curve in Figure 1.2 shows the number of "full" households in the weighted panel. This study uses the unweighted panel unless otherwise stated.

Figure 1. 2: Monthly number of household in the weighted and un-weighted panel



The households hand in a diary each week, but as some households miss a week and some households leave after a short while in the panel, the average number of households per week is lower than the average number per month and per year as shown in Table1.1.

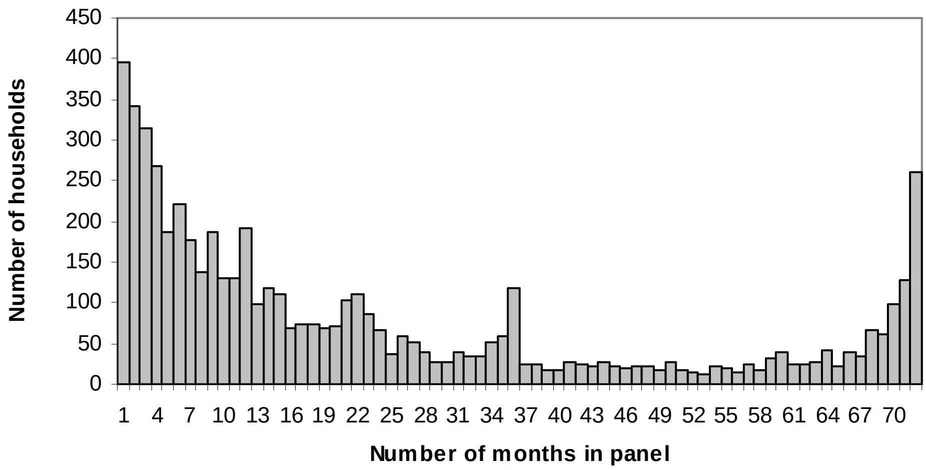
Table 1. 1: Average numbers of households in the panel per week, month and year

	1999	2000	2001	2002	2003	2004
Yearly average	2169	2082	2604	2779	4343	4635
Monthly average	1713	1691	2078	2224	3374	3746
Weekly average	1568	1550	1904	2045	3089	3408

[1] Weight, height and exercise are only reported from 2004 and onwards.
[2] To see the composition of the weighted panel according to age groups, family types, geography, social classes and income compared with numbers from Statistics Denmark see Smed (2002)

About 20 per cent of the households leave the panel each year; these are replaced by a similar type of household. Figure 1.3 gives an overview of how many months the households in general stay in the panel.

Figure 1. 3: Duration of households' panel membership



A smaller number of households (approximately 10 per cent) stay in the panel during the whole period from 1999 – 2004 (equal to 72 month). The panel consisting of these 260 households is referred to as the "balanced panel" while the panel consisting of all households is referred to as the "unbalanced panel". The differences between the balanced and the unbalanced panel are described in the subsections below.

Purchase data

For each shopping trip the diary keeper reports:

- the day of the week and time of the day
- the name of the store
- who participated in the trip
- total expenditure on the trip

The level of precision in the data varies from period to period and between various types of foods. Some years the level of detail can be very precise (e.g. the purchased good is registered by brand and detailed category) and then a couple of years later only the basics are recorded. For almost all goods in all periods the value and volume of the product is registered together with discounts and whether the food is organic or conventional produced. GfK compiles the data in the 52 basic food categories presented in Table 1.2. below. In this study we classify these 52 food groups into 8 main food categories. These are likewise presented in Table 1.2.

In Smed (2002) the purchase data from GfK is aggregated and compared to data from statistics Denmark to validate the data. It is found that total expenditure on each basic food category is lower in the GfK data than in data from Statistics Denmark, but if the budget shares for each basic food category is calculated and compared, the data are in accordance with the data from Statistics Denmark. This indicates that even though the households do not report everything there are no major biases in favour of specific basic food categories.

Table 1. 2: Food categories in GfK data

Main food group category	Basic food categories in GfK data
Processed food	Bouillon and soups
	Dishes with rice and pasta
	Salad dressing etc.
	Sauce
	Pizza
	Backing mixture
Juice and soft drinks	Fruit syrup
	Ice tea
	Juice
	Coffee
	Tea
	Fizzy drinks
Foods containing carbohydrates	Cereals
	Potatoes
	White bread
	Brown bread
	Flour
	Crisp bread
	Rice
	Pasta
Fruit and vegetables	Fruit
	Vegetables
	Frozen fruit and vegetables
Fats	Butter
	Oil
	Margarine
Meat	Other meat
	Bacon
	Fish
	Processed fish
	Poultry
	Rissole
	Processed meat for bread
	Liver pâté
	Beef
	Sausages
	Pork
	Brawn and pâté
Dairy	Speciality cheese
	Ordinary cheese
	Dairy snack
	Milk
	Yoghurt
	Eggs

Sugar	Chocolate (for bread)
	Ice-cream
	Biscuits
	Macaroons
	Marmalade
	Cake
	Sugar
	Small cakes

The varying precision in the data are caused by that GfK is a commercial firm and as such responds to the type of customers they have. This also have the implication that a few food groups; "backing mixture, processed fish, fish, rice, ketchup and edible oil" are missing in one or more years (See Table 1.3). As the consumption of edible oil, fresh and processed fish and rice are assumed to be important in order to give a satisfactory picture of the health status of Danish diets, fitted values for the purchased quantities of these foods are constructed in the missing time periods. The details of the estimation of fitted purchase quantities are described in appendix 1.A.

Table 1. 3: Overview of missing data

Basic food group	Period missing
Fish	January 2003 – December 2004
Processed fish	July 2001 – December 2002
Edible oil	January to July 1999
Rice	January 2003 - January 2004

Background data

The annual questionnaire provides information on a wide variety of background characteristics on each household. Most questions are posed to the households each year; a few are posed only in some years. Table 1.4 shows the most important and most interesting regularly posed questions.

Table 1. 4: Questionnaire content

Individual level: Social and demographic questions (posed to each individual in the household)	
Age and gender	Continuous, male/female
Level of primary school	7 years, 9 years, grammar school
Education level	No education, vocational (e.g. carpenter, nursing aide), short education (e.g. policeman, technical education), medium education (nurse, primary school teacher), long education (e.g. university degree)
Employment status	Full time (> 29 hour/week), part time (16 – 29 hours/week), less than part time (< 16 hours/week), self-employed, housewife/spouse, retired, early retired, pensioner, on leave, student, unemployed
Income	Intervals

Household level: Social and demographic questions	
Geographical region	Capital, Urban East, Rural East, Urban West, Rural West
Number of women above 21 years	Continuous
Number of men above 21 years	Continuous
Number of children 0 - 6 years	Continuous
Number of children 7 - 14 years	Continuous
Number of children 15 - 20 years	Continuous
Social class	Social-class 1 and 2: E.g. firm owners, people in leading positions, people with higher education. Social-class 3: E.g. owners of small firms, white-collar workers with few sub-ordinates or specialist skills. Social-class 4: E.g. white collar workers without subordinates and skilled blue collar workers. Social class 5: E.g. unskilled blue collar workers, unemployed and pensioners
Gender of main buyer	Female, male, both shop equally often
Electronic equipment in the household	TV set, video, dish washer, washing machine, tumble dryer, freezer, micro wave, laptop, PC, modem, scanner, CD, internet, DVD, printer, fax, answering machine, mobile phone
Type of housing	One-family house, farmhouse, two-family house, block of flats, room/studio
Ownership of house	Renter, lodger in room, free residence, co-operative flat, owner

Attitude questions	
Purchase frequency of selected convenience products	Frozen pizza, sauces, soups, ready-prepared dishes, bouillon, spice mixtures
Reason for use of convenience	Easy, quick, taste good, short term solution
Average time spent on daily cooking	Less than ½ hour, ½ hour, 45 min., 1 hour, 1½ hour, more than 1½ hour, varying, do not know.
Attitude towards cooking	Do not like, no opinion, like it, like it a lot, like it very much
Baking frequency	Once a week, once every second week, once a month, once every second month, once every half year, never
Frequency of home cooking	Never/not often, 1-2 times a week, 3-4 times a week, on a daily basis, only in the weekends.
Frequency of joint family breakfast and dinner	Never/seldom, 1-2 times a week, 3-4 times a week, on a daily basis, only in weekends

Questions on shopping behaviour	
Distribution of budget on store type	Per cent of budget used in; discount stores, supermarkets, kiosks and gas-stations, delicacy or speciality shops.
Price and brand awareness	The importance of; brand label versus cheaper products, special offers and price level on store choice.
Favourite and nearest shopping place	Names
Questions on media habit behaviour	
Advertisement flyers	Number received
	Number read
	Time used reading flyers
	Importance
	Reason for reading
Newspapers	Subscription
	Reading frequency
Magazines	Subscription
	Reading frequency
TV channels watched	Various channels to chose from

A membership of the panel and in particular a long term membership, may apply more to some people than to other. This might imply that the social and demographic profile of the balanced and the unbalanced panel may differ. Compared to the unbalanced panel the balanced panel consists of more households located in rural areas and more households living in houses. Furthermore there are more couples and more households where the woman is the main purchaser. A larger share of the households belongs to social class 4 and 5 in the balanced than in the unbalanced panel and they use a larger share of their food budget in discount stores and in ordinary supermarkets. There are more pensioners and people with no education, but fewer students and the income are in general lower in the balanced than the unbalanced panel. The average age is higher in the balanced panel and they have a smaller amount of electronic equipment.

1.2 Combination of household data and nutrition data

The purchase data from the GfK data are concatenated with nutrition matrices from the Food Composition Databank provided by the Danish Institute for Food and Veterinary research (http://www.foodcomp.dk/fcdb_default.asp). The nutrition data base provide detailed information about the content of 10 macronutrients (as e.g. protein, fats, carbohydrates and fibres), 18 vitamins (as e.g. vitamin A and vitamin C) and 13 minerals (as e.g. calcium and sodium) in 1032 different foods[3]. As all values are given per 100 g edible part in the nutrient matrices, it is possible to calculate the total amount of various macronutrients, vitamins and minerals purchased by the households by concatenating the nutrition matrices with the purchase data. For each type of food the match is done on a level beyond the level of detail in the basic food groups in the purchase data from Table 1.2. It is for example possible to separate the purchased quantity of milk into different types of milk (e.g. butter milk, whole milk, semi skimmed milk, skimmed milk and flavoured milk) and to match each type with a nutrition matrix describing the exact content of nutrients in this particular type of milk. Table 1.5 shows the concatenation level for each food group.

[3] The database covered 1032 different foods in 2005, but is continuously improved

Table 1. 5: Linking of nutrition and purchase data

Basic food group	Concatenation level
Butter and blends	Type (e.g. ordinary butter, butter mixtures, with or without salt)
Margarine	Type and fat content (e.g. for baking, for cooking)
Cheese	Type and fat content (e.g. emmentaler, gouda, havarti)
Eggs	Only one type of egg
Milk	Type (e.g. buttermilk, chocolate milk, skimmed milk)
Soured milk	Type and fat content (e.g. flavoured yoghurt, natural yoghurt, low fat yoghurt)
Brown bread	Type (e.g. whole grain, dark, less dark).
White bread	Type (e.g. toast, flutes, pita)
Pizzas	There is only one type in the nutrition database
Dairy snack	There is only one type in the nutritional data base
Crisp bread	Type (e.g. wheat, rice, whole grain)
Sugar	Type (e.g. cane sugar, brown sugar, white sugar)
Flour	Type (e.g. wheat, wholegrain, durum)
Breakfast cereals	Type (e.g. cornflakes, fibre products, muesli)
Ice-cream	Size and form (e.g. cake, litre, cornet).
Biscuits	Type (e.g. crackers, digestives)
Processed fish	Species (e.g. herring, tuna, codfish)
Canned foods and pickles	Type (e.g. tomatoes, sauerkraut, olives)
Bouillon and soups	There are only a few soups in the nutrition data. . The excluded are approximated by the average of raw products
Sweet biscuits and cookies	Type (e.g. macaroon, Christmas cakes)
Baking mixture	Baking mixture is approximated with an average of raw products
Jams	Flavour (e.g. strawberry, orange, apricot)
Poultry	Type (e.g. chicken, turkey, duck)
Frozen and canned vegetables	Type (e.g. onions, carrots, leeks). The packages with vegetable mix are approximated with an average
Cakes	Type (e.g. chocolate, lemon). There are a limited number of different cakes in the nutrient database. For those excluded the average is used
Pasta	Type and flavour of pasta (e.g. macaroni, green, white). Fresh pasta is approximated using nutrient data for boiled pasta
Rice	Type (e.g. brown, parboiled, pudding rice)
Sauces and spices	There are only a few sauces in the nutrition data. For those excluded the average is used
Spread-able chocolate	Type (e.g. dark, light)
Mustard and ketchup	Type (e.g. chilli, Mexican)
Fruit	Type (e.g. apple, pear, orange)
Vegetables	Type (e.g. carrot, onion, leeks)
Processed meat for bread	Type of meet (e.g. poultry, pork, beef) and to a limited extent also type of cut (e.g. roast beef, salami)
Liver pâté	Type (e.g. with mushrooms, garlic, cream)
Pâté and brawn	Type (e.g. with mushrooms, garlic, cream)
Bacon	Size and form (e.g. whole pieces, sliced)
Meat balls	Only one type
Mayonnaise	Only one type
Edible oils	Type (e.g. olive, sunflower seed)
Salad dressing	Type (e.g. mayonnaise, oil-vinegar)
Beef	Cut (e.g. fillet, minced, roast beef)
Pork	Cut (e.g. chop, minced, ham)
Other meats	Type (e.g. lamb, crocodile)
Sausages	Only one kind of sausage in the nutrition data
Fish	Species (e.g. tuna, herring, codfish)
Dishes with pasta and rice	Approximated with an average of raw ingredients
Soft drinks	Type (juice, ice tea)

An example of the concatenation of the GfK purchase data with the nutrition data is shown in Table 1.6 for milk.

Table 1. 6: An example of concatenation of purchase and nutrient data

	Content pr 100 g semi skimmed milk	Content in purchase of 2 litres (it is assumed that one litre weighs 1000 gram)
Total energy	202 kJ	4040 kJ
Total protein	3,5 g	70 g
Total fat	1,6 g	32 g
Saturated fat	1,11 g	22,2 g
Poly unsaturated fatty acids	0,037 g	0.74 g
Mono unsaturated fatty aids	0,38 g	7,6 g
Carbohydrates	4,9 g	98 g
Available carbohydrates	4,9 g	98 g
Added sugar	0 g	0 g
Fibres	0 g	0 g
Vitamin D	0,09 µg	1,8 µg
Vitamin C	1,3 mg	26 mg
Folacin acid	11 µg	220 µg
Cholesterol	6 mg	120 mg
Iron	0,031 mg	0,62 mg

1.3 Food consumption in Denmark

The concatenation of purchase data with nutrition data makes it possible to give a picture of the development in the composition of the Danish diets both in terms of energy shares, budget shares and in terms of nutritional composition. The following figures are based on the unbalanced weighted panel so in principle the figures represent the diet of the Danish population. Figure 1.4 and 1.5 below show the share of total energy intake from each of the main food groups in Table 1.2.[4] With time the energy shares from especially meat and processed foods, but also fruit and vegetables and carbohydrate containing foods have increased while the shares from dairy and especially from fats have declined. Seasonal patterns are found for almost all food groups. For meats peaks are seen around Christmas and in summertime while the opposite pattern is found for processed foods. The energy share from fruit and vegetables drops through summer. This is assumed to arise from the fact that different types of fruit and vegetables are eaten during the year. The energy shares from dairy and sugar-products show the opposite pattern – it increases during the summer. The large energy share from sugar products during summer might come from an increased consumption

[4] The results of a linear regression through the datapoints are shown in appendix 1.B

of ice-cream. Generally it seems like diets change considerable during summer (June, July and August), in January and around Christmas compared to the rest of the year

Figure 1. 4: Share of total energy intake from selected foods

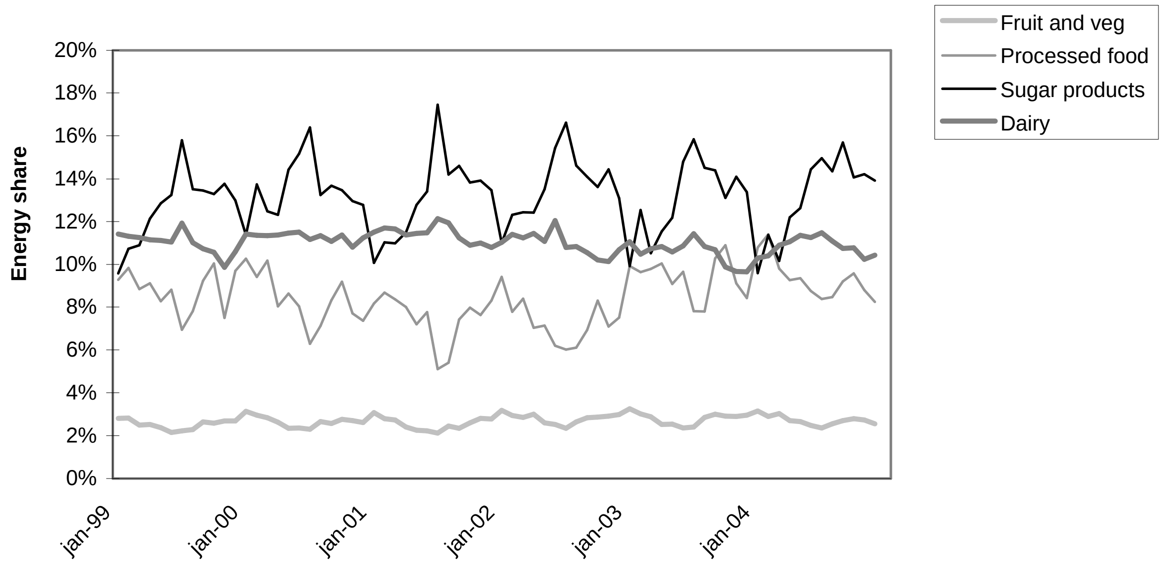


Figure 1. 5: Share of total energy intake from selected foods

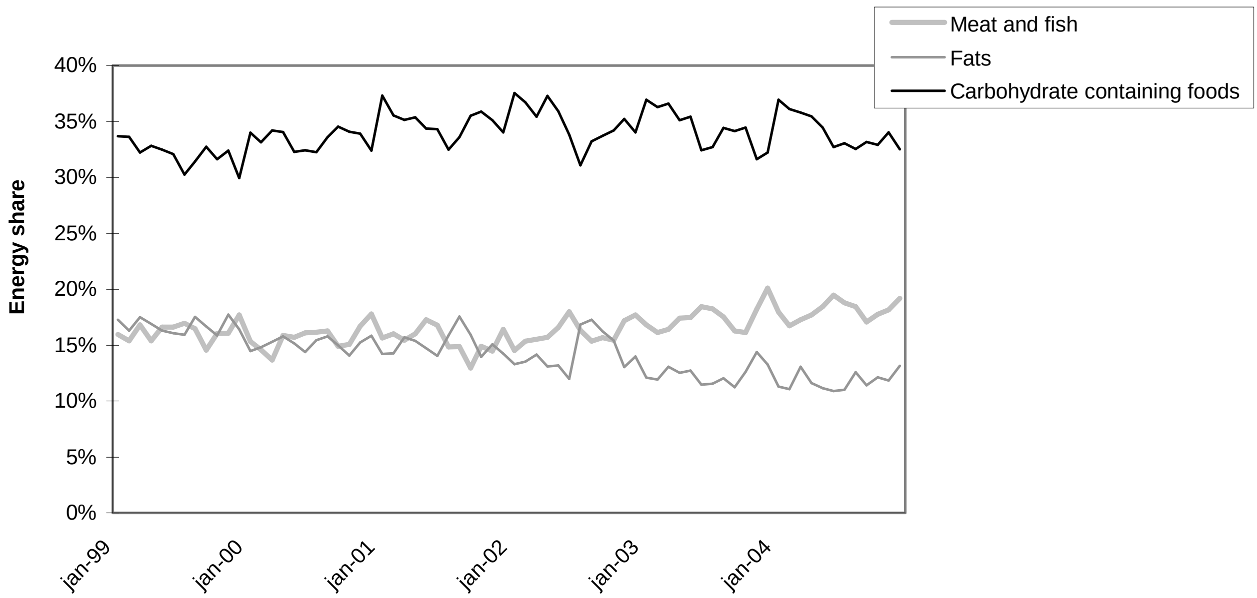


Figure 1.6 shows the share of total energy consumption from selected macronutrients.[5] There is a small increase in the energy share from carbohydrates over time, but with small repeated declines during December (Christmas) and summer. The opposite is seen for fat. The energy share from saturated fat and total fat decline over time, but there is a repeated increase in the

[5] The results of a linear regression through the datapoints are shown in appendix 1.B

energy shares in December[6] followed by a decline in January and February. Also July and August are months with a larger intake of fat than other months. The energy share from sugar shows a huge seasonal variation with a decrease in the beginning of the year and a huge increase during summer, but no trend. The general pattern shows an increase in the healthy nutrients (fibres, protein and carbohydrates) during January and February and an increase in the more unhealthy nutrients (total fat, saturated fat and sugar) during Christmas and summer.

Figure 1. 6: Energy shares from protein, fat, saturated fat, carbohydrates, and sugar

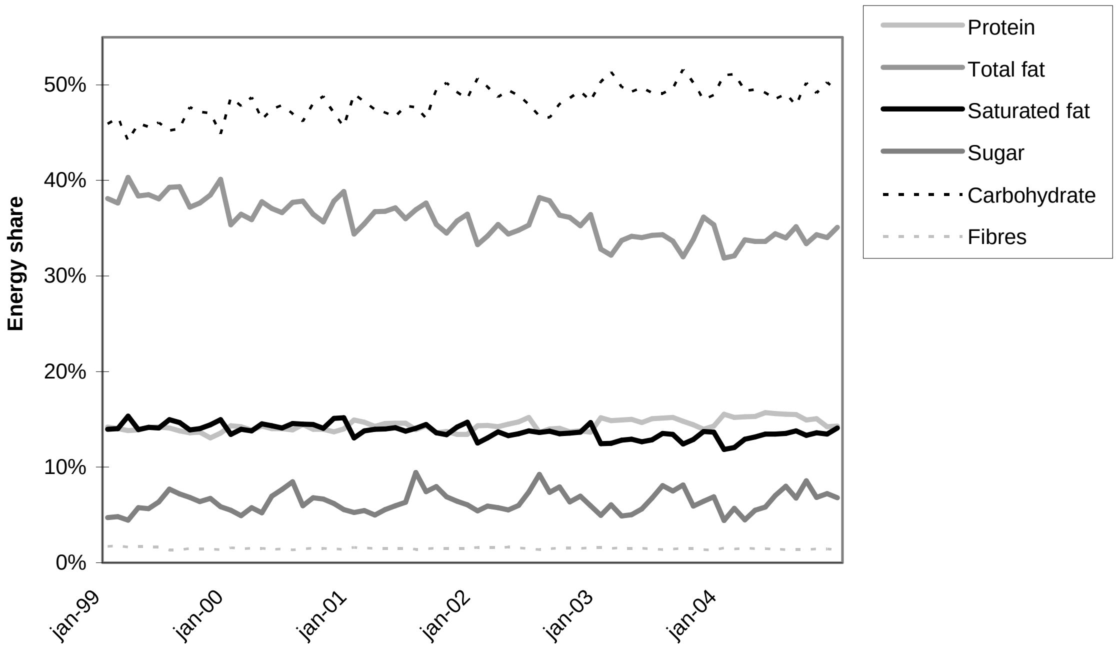


Figure 1.7. and 1.8 show that there is a close correlation between the energy share from fruit and vegetables and the energy shares of saturated fat and sugar respectively. Around 28 per cent of those who get less than 4 per cent of their energy from fruit and vegetables get between 10 – 15 per cent of their energy from saturated fat and 60 per cent gets between 15 – 20 per cent. The same numbers is 44 and 42 per cent respectively for those who get more than 8 per cent of their energy from fruit and vegetables. Likewise for sugar there is a tendency to get a small share of energy from added sugar for those who get a large share of their energy from fruit and vegetables.

[6] If the trend is pictured on a weekly basis it is easy to point to the week before Christmas and before New Year's Eve as the large contributor to the increased share of fat in December

Figure 1. 7: Correlation of energy shares from saturated fat and from fruit and vegetables

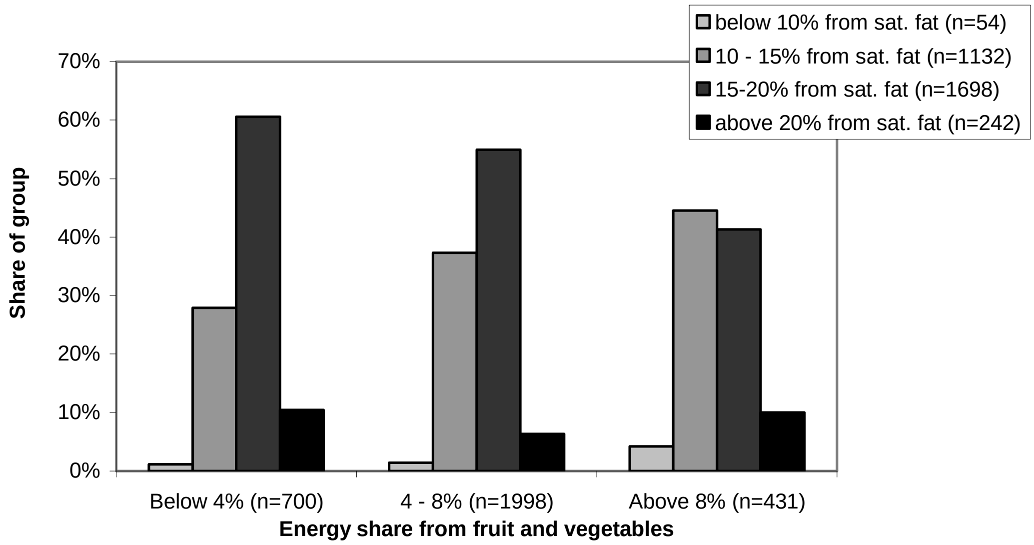
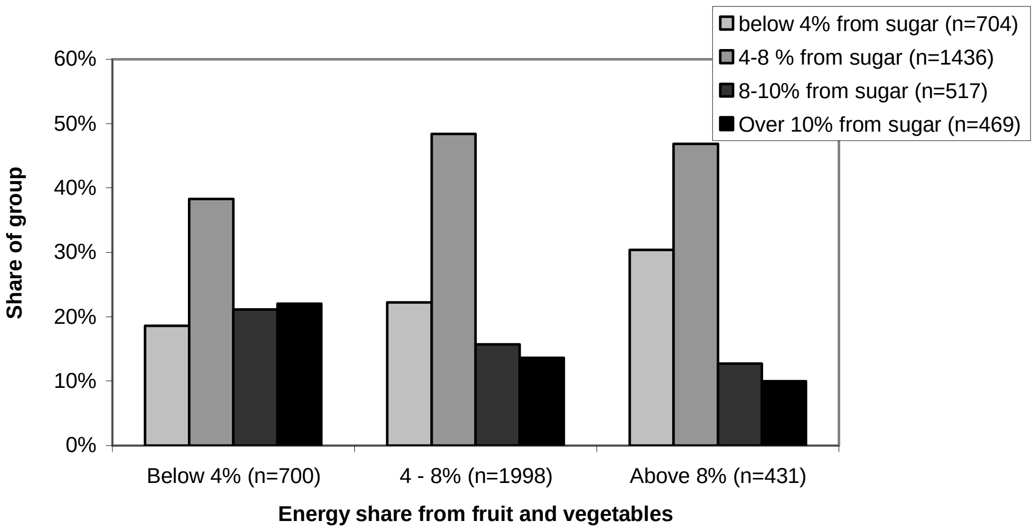


Figure 1. 8: Correlation of energy shares from sugar and from fruit and vegetables



1.4 Data on consumer health

In addressing obesity and overweight the GfK panel have in 2004 filled in a questionnaire concerning height, weight and exercise data for each individual in the household. The height and weight data can be combined into a measure of Body Mass Index (BMI).[7] BMI can to some extent describe the health status of the respondents, due to the increased occurrence of lifestyle related illnesses as e.g. diabetes, cardiovascular disease or cancer with overweight

[7] BMI is calculated as: $BMI = \frac{weight(kg)}{height(m) * height(m)}$. Overweight is then defined as a BMI above 25, but below 30, while obesity is defined as having a BMI above 30.

status. In addition to the exercise, height and weight questions the respondents answered questions concerning their dieting "habits", attitude towards their own weight, and active effort to change weight. In total there are 3567 respondents, 2930 adults (1214 males and 1716 females) and 637 children. Figure 1.9 shows the mean BMI with age (only adults, children below 18 years have been delete from the data since they follow another scale than adults) separated by gender. There is an increasing BMI with age until the age of 56 years for woman and 60 years for men where after BMI declines slightly. The gap between the average BMI for men and women widen over time.

Figure 1. 9: Mean BMI separated by age and gender

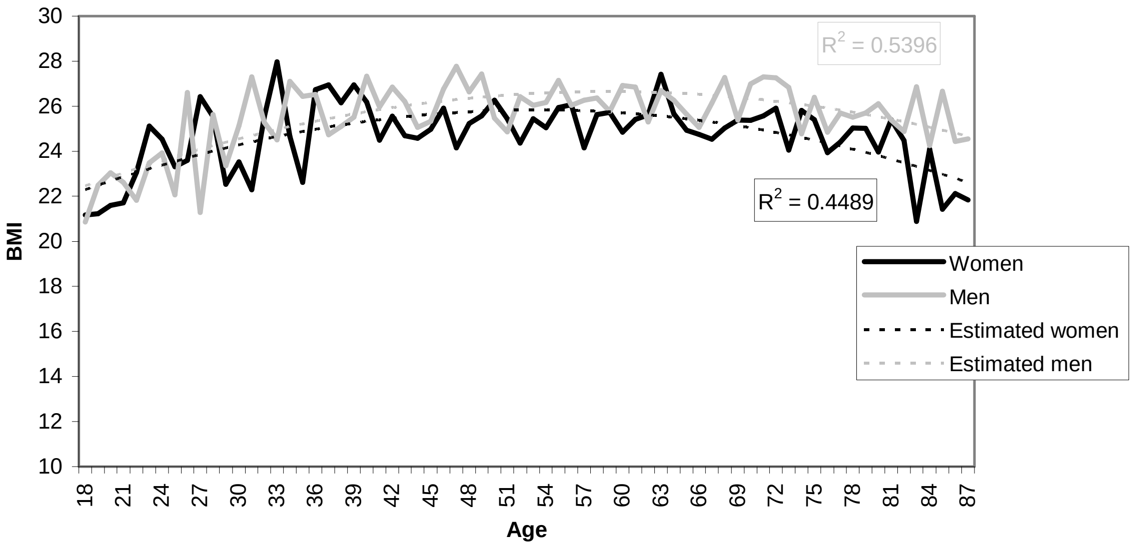


Figure 1.10 indicates that mean BMI declines with the level of education[8]. The mean BMI is significantly smaller for all educational groups except vocational educated compared to households where the head has no further education. Figure 1.11 illustrates the share of the panel being either overweight (BMI between 25 and 30) or obese (BMI equal to 30 or above) for various social and demographic groups. The figure clearly shows that the prevalence of overweight and obesity is larger among lower educated (no, vocational or short) than higher educated (medium or long) especially for younger. The prevalence of obesity is larger in the Capital than in other regions for younger households, while the opposite apply to older households. Overweight is generally more common among households in other regions than in the Capital. No specific pattern is found between age groups.

[8] For a definition of the educational groups see Table 1.4

Figure 1. 10: Mean BMI separated by education

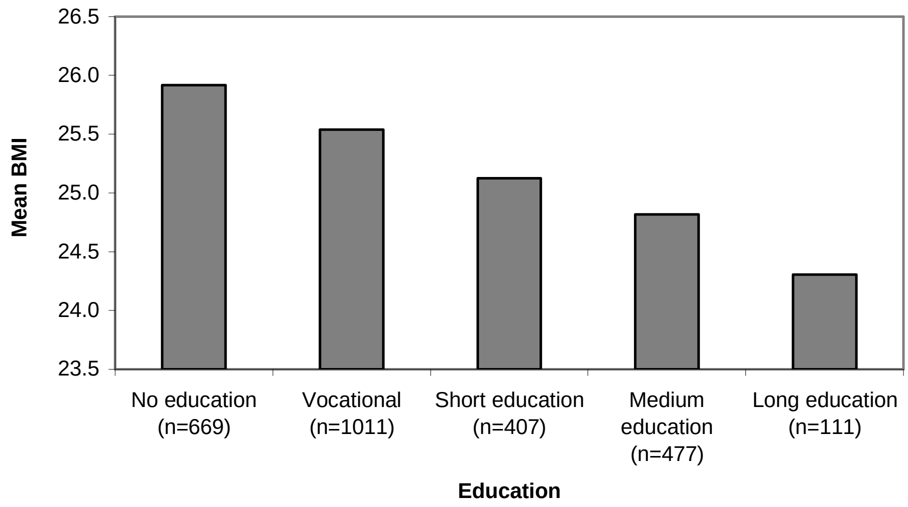
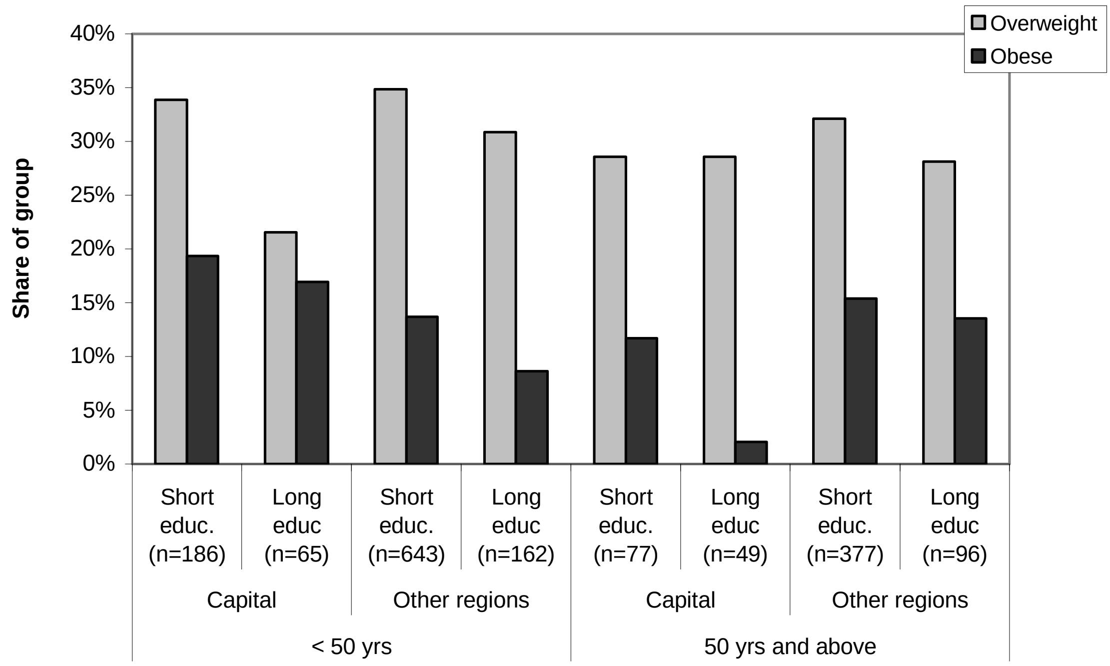


Figure 1. 11: Prevalence of overweight and obesity for social and demographic groups



Exercise patterns might be another indicator of health, due to the close relationship between daily exercise and health. The Danish health authorities recommend at least 30 minutes of exercise a day. Figure 1.12 shows exercise patterns for the same social and demographic groups as above. The figure indicates that there are more households among the longer educated who exercise more than 60 minutes a day. This applies in particular to Capital located households. Older households in the Capital exercise more than younger households in the Capital while no difference is seen between age groups for households located in other regions.

Figure 1. 12: Minutes of daily exercise for social and demographic groups

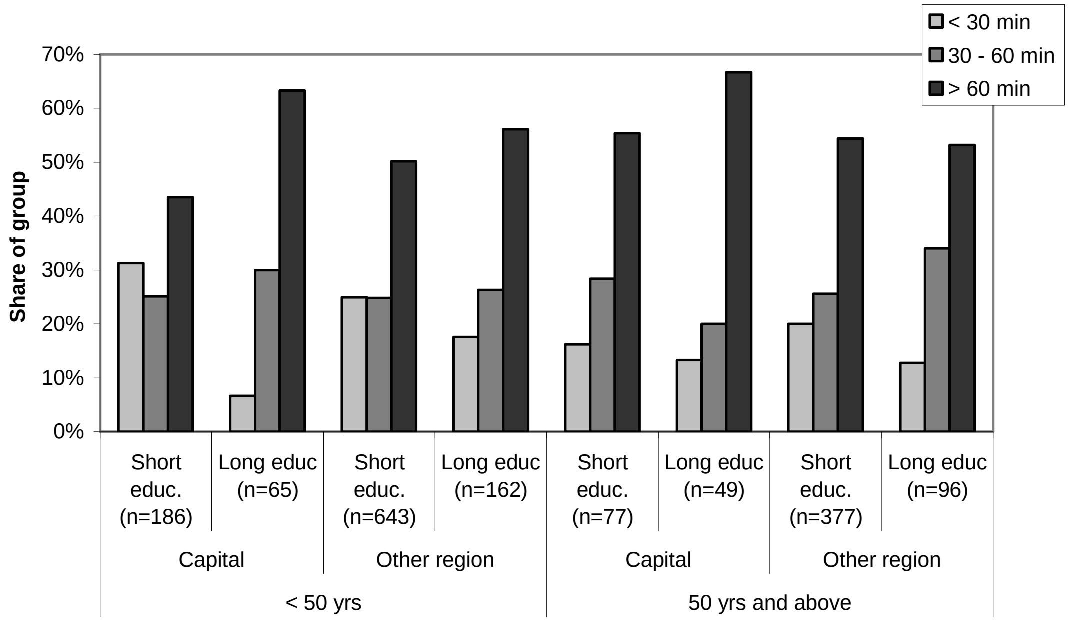
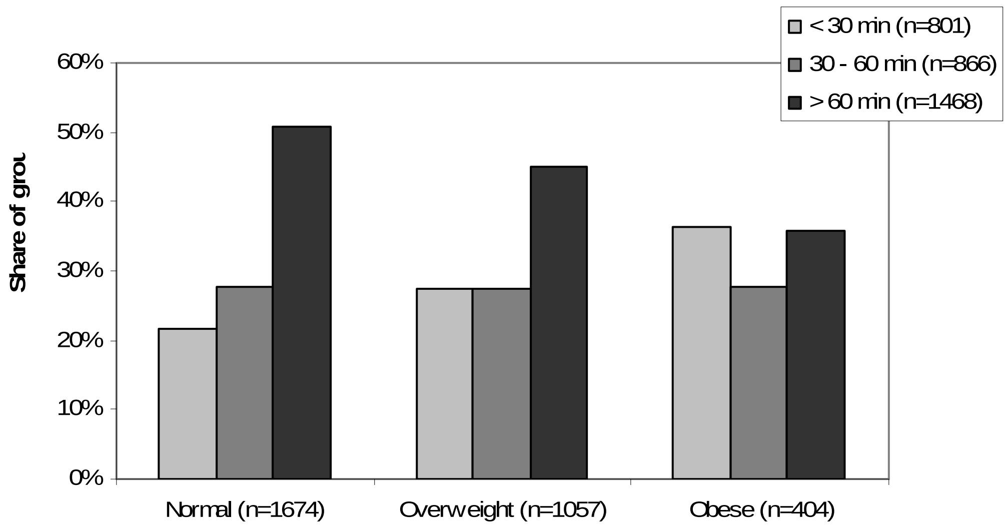


Figure 1.13 shows the relation between the daily number of minutes of exercise for normal weight, overweight and obese individuals. The figure indicates that there are more among the obese exercising less than 30 minutes a day and less exercising more than 60 minutes a day when compared to the normal weight.

Figure 1. 13: Minutes of daily exercise and overweight status



References

Andersen L.M. (2006): *Consumer Evaluation of Environmental and Animal Welfare Labelling*. AKF Forlaget, Copenhagen

Smed S. (2002): *En sociodemografiske analyse af den danske fødevareefterspørgsel* (A sociodemographic analysis of the danish demand for foods). Rapport nr. 146. Fødevareøkonomisk Institut, KVL, Copenhagen

Appendix 1.A. Estimation of predicted data

As described there are a couple of food groups; "processed fish, fresh fish, rice and edible oil" that are missing for one or several years. It is assumed that most of these foods are important in order to give a complete picture of diet composition. Therefore, it is chosen to predict the purchased quantities of these foods and the purchased amount of nutrients. The latter is done in order to reflect changed composition within individual food groups, e.g. the consumed amount of milk can be stable while at the same time the amount of saturated fat is decreasing due to a change from one type of milk to another. This gives seven equations to estimate for each food. The predictions are based on parameters estimated in a time-series analysis, with the monthly average purchased quantity of the food group in question as the explained variable and a time trend and seasonality as the explanatory variables.

$$y_{it} = \alpha_i + \beta_{trend} \cdot t + \beta_{trend2} \cdot t^2 + \sum_{s=1}^{12} \beta_s \cdot S_s \qquad \text{(A.1)}$$

The panel is divided into two age groups (< 50 or 50 yrs and above), three educational groups (no, vocationally or short, medium or longer further education), three family status (no kids, kids below 6 or kids between 7-14 yrs) and finally two regional groups (Capital, other regions). This leads to 36 different groups. For each food the purchased quantity, total energy, amount of fat, amount of saturated fat, amount of fibres, amount of carbohydrates and amount of added sugar is predicted individually i.e. seven versions of equation A.1 is estimated for each food and for each social and demographic group. The model is estimated as a pooled estimation in first differences. Finally, the households' consumption of the amount of the particular food and the nutrients from this particular food are predicted using the estimated parameters.

For those households who are only in the panel in the period where the data are missing α_i is approximated with the average consumption for the groups that the household belongs to in the first period.

Figures A.1-A.4 illustrate the estimated quantities together with actual quantities for the unbalanced unweighted panel (i.e. the figures reflected both change in panel composition and change in panel size).

Figure A. 1: Actual and estimated consumption of edible oil

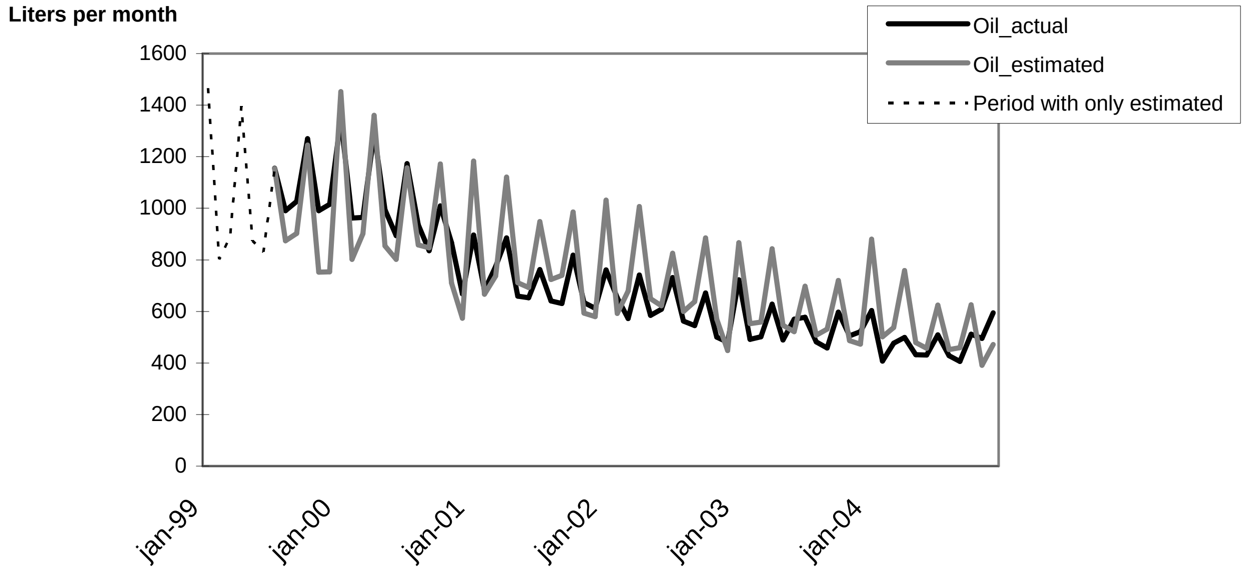


Figure A. 2: Actual and estimated consumption of rice

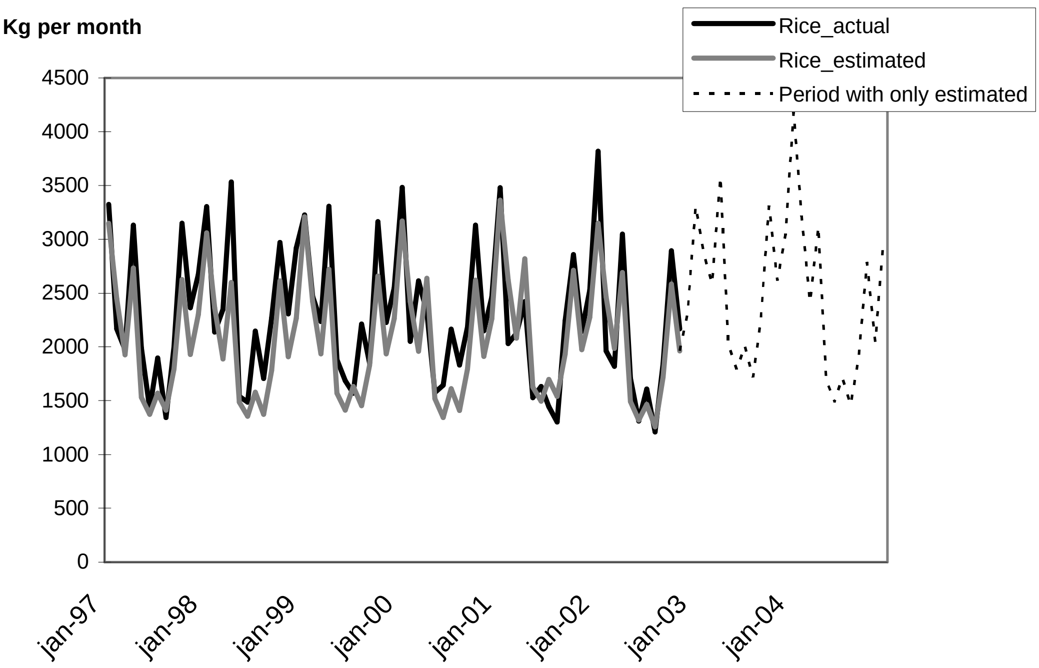


Figure A. 3: Actual and estimated consumption of fresh fish

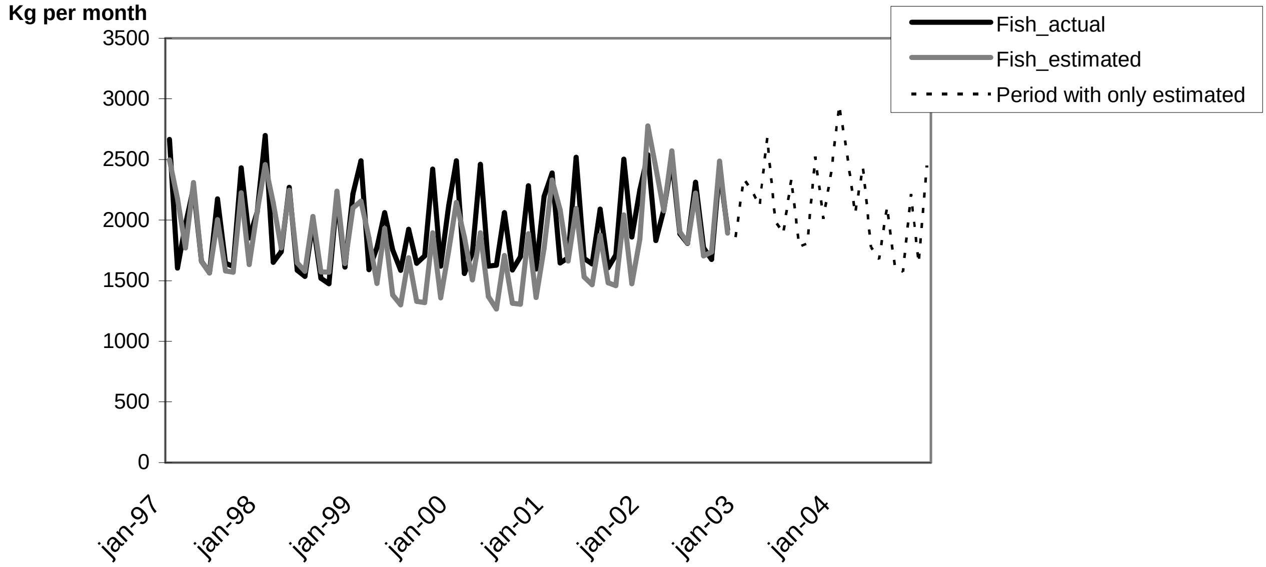
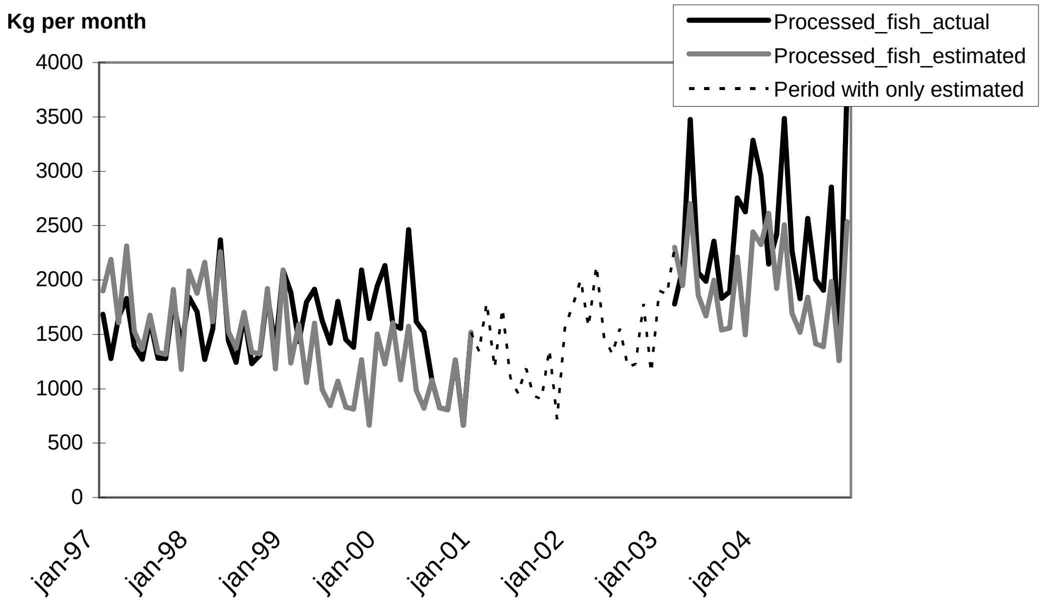


Figure A. 4: Actual and estimated consumption of processed fish



Appendix 1.B. Results from estimation through data points

Figure B. 1: Results from estimation through data points in figure 1.4 and 1.5

	Fruit and vegetables	Processed foods	Sugar products	Dairy	Meat_fish	Fats	Carbo
Constant	0.02580	0.08706	0.13545	0.10916	0.14744	0.17259	0.32206
Slope	0.00004	0.00012		-0.00008	0.00036	-0.00078	0.00031
January	0.00394		-0.03402	0.00461		-0.01065	0.02893
February	0.00196		-0.01714	0.00459		-0.01116	0.02026
Marts			-0.02425	0.00545			0.01649
April			-0.01661	0.00575			0.01744
May	-0.00261			0.00549	0.01004		
Jun	-0.00376	-0.00953	0.00732	0.00752	0.01528		
July	-0.00439	-0.02405	0.02493	0.00880	0.00803		
August	-0.00164	-0.02057		0.00579			
September							
November		-0.01239					
December		-0.00966			0.01894	0.00503	-0.00997
R-squared	0.7384	0.3690	0.7288	0.4173	0.5314	0.6889	0.4540

*Only parameters that are significant different from 0 (5%) are shown

Figure B. 2 Results from estimation through data points in figure 1.6

	Total fat	Saturated fat	Carbohydrate	Sugar	Fibres	Protein
Constant	0.38579	0.14542	0.46103	0.06471	0.01590	0.13414
Slope	-0.00080	-0.00024	0.00061		-0.00001	0.00021
January	-0.01764	-0.00889	0.01276	-0.01416	0.00081	0.00706
February	-0.01309	-0.00510	0.01090	-0.00972		0.00495
Marts				-0.01396		0.00311
April				-0.01028		0.00484
May						0.00518
Jun						0.00628
July	0.01147	0.00413	-0.00901	0.02026	-0.00147	
August	0.01403	0.00496	-0.01380	0.00576	-0.00086	
September				0.01235		
November	0.00983	0.00560			-0.00070	-0.0056
December	0.01858	0.01036	-0.01067		-0.00097	-0.0040
R-squared	0.8256	0.7768	0.7059	0.7577	0.5144	07493

*Only parameters that are significant different from 0 (5%) level are shown

Chapter 2

Measuring the Health Performance of Diets

Sinne Smed

AKF
(Danish Institute of Governmental Research)
Copenhagen

Abstract

In this paper a Healthy Eating Index is developed, that measures the health performance of diets in terms of compliance to multiple health recommendations simultaneously. Consumers' compliance to each of several dietary recommendations are measured individually and then weighted together to one measure using an Euclidian distance measure. Based on a panel dataset, that includes data on food and nutrition consumption together with food expenditure for approximately 2500 Danish households per week, changes in the health performance of diets in relation to the official dietary recommendations are described from 1999 to 2004 using this Healthy Eating Index. The results show that only 1.5 percent of the population fulfils all the diet recommendations, but minor improvements in the diet are found in the studied period. A social gradient is identified, revealing that shorter educated eat less healthy than longer educated, and older households are found to be closer to the diet recommendations than younger households. The effect of age is so dominating that the best educated of the younger households are on the same level of health performance as the shortest educated of the older. Low educated younger also seem to have smaller improvements in the health performance in the studied period than other types of households, suggesting that the social bias in dietary patterns persists. The relationship between dietary quality and BMI is confirmed, and the results suggest that observed differences in BMI between men and women might be caused by healthier dietary patterns among women than among men. Furthermore we find a positive correlation between food expenditure and the health performance of diets.

Introduction

Health problems related to poor diets are a major problem in most industrialised countries. WHO (2002) has estimated that at least 10 to 15 per cent of lost years of life in Europe can be attributed to poor nutrition. Poor nutrition contributes to the prevalence of a number of chronic conditions such as cardiovascular diseases, some types of cancer, diabetes, osteoporosis as well as obesity. During the last two decades the Danish health authorities have conducted massive campaigns aimed at the whole population to encourage a decreased intake of saturated fat and an increased intake of fish, fruit and vegetables. There is evidence that consumers have reacted positively to these messages (Holm et al., 2002; Haraldsdöttir et al., 2005; Astrup et al., 2005). Despite this positive development, statistics show that the share of the population that is severely overweight is still growing (Ekholm et al., 2006), especially for the lower educated and low income groups (Richelsen et al., 2003). The prevalence of lifestyle-related illnesses is also more prevalent in lower educated and low income groups than in other parts of the population (Brønnum-Hansen et al., 2004; Lynch et al., 2006). Part of this problem might be related to divergence in dietary patterns.

Due to the complexity of human diets and the many nutrient-to-nutrient interactions that occur in the body, conclusions about the health outcome effect of the consumption level of single nutrients may be misleading. This is supported by recent epidemiological studies of diet and health outcomes (Hu et al., 2000; Fung et al., 2001). For these reasons it is useful to examine diet quality and dietary patterns through indices of food and nutrient intake that express several related aspects of dietary intake concurrently. The overall idea of the Healthy Eating Index (HEI) developed in this paper is to measure how well Danish diets conform to recommended eating patterns, to record differences in dietary patterns among various household types and to examine changes over time. The data used in this paper distinguish themselves by having repeated measures of the same households over time. This means that we have the possibility of describing and modelling the development over time of dietary patterns to a larger extent than analysis based on dietary recalls or shorter time periods. Furthermore, the data give a unique possibility of analysing the effect of expenditure on healthy eating patterns since prices are reported along with amount of nutrients purchased.

The remainder of this paper is organised as follows: Section 2.1 briefly summarise the important literature in the area. Section 2.2 describes the theoretical construction of the Healthy Eating Index (HEI), the data used and the empirical construction of the HEI. Section

2.3 examines changes in the HEI over time. Section 2.4 contains a detailed description of the variation in how well different social and demographic groups perform according to the dietary recommendations, while the cross-section estimations in Section 2.5 only look at average effects, but control for the effect from other social and demographic characteristics. Finally, Section 2.6 is discussion and conclusion.

2.1 Methods to measure health performance of diets

Two methods are generally applied to study the quality of diets and dietary patterns. The first is the use of statistical techniques such as factor and cluster analysis to identify common dietary patterns, and relate these to health outcomes. The general results from these studies are that a diet with a large share of red meat, processed meat, and refined products is correlated with a higher risk of cardiovascular diseases and a larger BMI. Diets with a high intake of fruit, vegetables and whole grains have the opposite influence (Newby et al., 2003; Maskarinec et al., 2000; Hu et al., 2000; Fung et al., 2001). The second approach, which is followed in this paper, is to compare diets to some specific goals and score them according to a set of criteria to produce a composite index of diet quality. Within the latter approach three major ways to construct diet quality indices have been followed in the literature: indices derived from nutrients only; indices based on foods or food groups; and those based on a combination of the two (for a review, see Kant, 1996). Examples of indices which are based on a combination of nutrients and food groups together are the diet quality index (Patterson et al., 1994; Haines et al., 1999) and the Healthy Eating Index (here named US-HEI) constructed by the U.S. Department of Agriculture (USDA). In particular the US-HEI has similarities to the index developed here. The USDA introduced the US-HEI in late 1995 and the principles of this index have been applied to various datasets (see e.g. Guo et al., 2004; Basiostis et al., 2002; Variyam et al., 1998). The US-HEI is based on selected nutrient and food intake recommendations of the US food and nutrition board and provides a health measure of the diet based on 10 dietary components. The first five components measure the extent to which a diet conforms to the USDA Food Guide Pyramid serving recommendations for grain, vegetable, fruit, milk and meat groups. The next four components assign scores according to the amount of selected nutrients (total fat, saturated fat, cholesterol and sodium) evaluated against the dietary guidelines for maximum daily intake of these nutrients. The last component assesses the variety in the diet. The objective of this paper is to construct a similar measure which is in accordance with the official Danish dietary recommendations and to look at changes in this measure across households and over time. What distinguish our measure are the inclusion of the economic variables in the data and the use of the Euclidian distance measure to measure composite healthiness of the diet.

2.2 Construction of a Healthy Eating Index

How to define a healthy diet

It is a matter of definition to decide what a healthy diet is. Here, a healthy diet is defined as a diet being in accordance with the Danish official diet recommendations published by the Danish Ministry of Family and Consumer Affairs. The latest revision of the recommendations[9] is shown in Box 2.1 below. The text in parentheses is the more advanced version of the recommendations. The recommendations are made in cooperation with the Danish Nutrition Council and the Danish Institute for Food and Veterinary Research and are in accordance with the Nordic Nutrition Recommendations given by the Nordic Council of Ministers.

Box 2.1: The official Danish diet recommendations

1) Eat 6 pieces of fruit and vegetables a day (excl. potatoes) – (adults 600 g and children 300-400 g/daily)
2) Eat fish several times a week (200-300 g. a week)
3) Eat potatoes, rice, pasta or brown bread every day (50-60 per cent of total energy intake from carbohydrates, 22 g. of fibres per 10 MJ of food consumed)
4) Reduce the consumption of sugar – especially from soft drinks, sweets and cakes (max. 10 per cent of total energy intake from added sugar)
5) Reduce the amount of fats – especially from dairy products and meat (max. 30 per cent of total energy intake from fat and max. 10 per cent of total energy intake from saturated fat)
6) Eat a variety of foods and keep your normal weight
7) Drink water
8) Exercise – at least 30 minutes a day

The correspondence between these recommendations and a healthy diet might be debated. In particular the recent revision of the food guide pyramid, the so-called healthy eating pyramid (Willett and Stampfer, 2003; Willett et al., 2001) gave rise to a comprehensive discussion. Each of the above recommendations is discussed in relation to the healthy eating pyramid before inclusion in the HEI. There is general consensus about the recommendations for fruit and vegetables, fish, sugar, saturated fat and fibres. The controversial points are carbohydrates and unsaturated fat. According to the official diet recommendations in box 2.1 a diet with 50-

[9] The first Danish diet recommendations were formulated in 1970.

60 per cent of total daily energy intake from carbohydrates will protect against cardiovascular diseases by lowering the intake of fat, which is also the reason why a carbohydrate containing diet is said to prevent obesity. In contrast to that, the healthy eating pyramid by Walter Willett recommends a low intake of "fast" carbohydrates. These are rice, potatoes, pasta and white bread. As carbohydrates are recommended due to that it substitute for fat not because it is healthy in itself we choose to disregard the carbohydrate recommendation in the HEI. Consumption of unsaturated fats is another controversial point. Despite the healthy aspects of consumption of unsaturated fat, it is still richer in energy than nutrients such as sugar and carbohydrates and the consumption of large amounts of fat might have implications in relation to obesity. Therefore we choose to include the official recommendations of maximum 30 per cent of total energy consumption from fat in the HEI. The recommendations 6, 7 and 8 are not taken into account. The use of tap water is excluded since most datasets do not comprise the use of tap water. Exercise is not directly considered as part of a healthy diet. Furthermore, we choose not to focus on variability here due to problems of defining what a healthy variability in diets is. This might be a route for further research. In the empirical implementation it is also necessary to adjust some of the recommendations to suit the data. These issues are discussed in Appendix 2.A together with a more detailed description of the heath implications of each of the recommendations.

Theoretical construction

The HEI is a weighted measure of how well the households perform in relation to several recommendations simultaneously. First the household's performance in relation to each of the recommendations is measured as the degree to which the household fulfils this particular recommendation. The minimum score, zero, is assigned to the worst possible consumption. This is equal to 0 for a "positive" food or nutrient like e.g. fibres or fruit and vegetables and a maximum possible consumption for a "negative" food or nutrient e.g. saturated fat or sugar. The minimum or maximum possible consumption is defined as the observed minimum or maximum consumption of a specific nutrient or food in the panel[10], due to the fact that it is not possible to have an energy share of e.g. fat equal to 100 per cent during a month. The maximum score is 10 indicating that the recommendation is met[11]. Households with intake between the recommended and the minimum or maximum possible consumption are assigned

[10] The maksimum possible is an energy share at 0.46 for saturated fat, 0.81 for fat and 0.66 for sugar

[11] The number 10 is just an index number and could just as easily have been 100 or 1, which would only have the effect of up- or down-scaling the index.

scores proportionally according to a linear scale. This gives the following formula for measuring the performance score for each of the recommendations:

$$Score_t = 10 \cdot \left(\frac{s_{it} - s_0}{s_{recommended} - s_0} \right) \tag{2.1}$$

where s_{it} is actual consumption of nutrient *i* at time *t* and s_0 represents the consumption which will give a 0 score. It is of course questionable to choose a linear scale since the health effects of changing food habits often follow a non-linear pattern. One problem of using a non-linear scale is to weigh different positive health outcomes together (e.g. a decrease in cardiovascular disease versus a decrease in cancer) in a reasonable way, another problem is that in most cases the exact functional form of the health effects of changed dietary patterns is unknown.

The individual scores are weighted together using an Euclidean distance measure, where *i* is the number of components in the HEI measure and *n* is household number, *t* is time.

$$HEI_{nt} = \sqrt{\sum_{i=1}^{I} \left(score_{int} \right)^2} \tag{2.2}$$

$HEI \in \left[0; \sqrt{600}\right]$. A household fulfilling all the recommendations will get a $HEI = \sqrt{(10)^2 + (10)^2 + (10)^2 + (10)^2 + (10)^2 + (10)^2} \approx 24.5$. A household being half way between the worst possible and the recommended level for all 6 recommendations will get a $HEI = \sqrt{(5)^2 + (5)^2 + (5)^2 + (5)^2 + (5)^2 + (5)^2} \approx 12.2$. It can be discussed whether it is reasonable to weigh all components equally. But as above there will be problems in weighting a decrease in one type of illness against another and to combine a potentially non-linear decrease in the incidence of cardiovascular disease due to a higher fruit and vegetable consumption with another type of non-linear decrease in the incidence of cardiovascular disease due to a decrease in the consumption of saturated fat. To make these comparisons is beyond the scope of this paper. This means that the HEI value developed will merely measure health behaviour than actual health effects.

Data

The HEI measure will be applied to a dataset provided by GfK Denmark which contains weekly records (aggregated to monthly observations) for on average 2500 households' purchases of foods including juice and soft drinks. The data are at household level and the purchased food, in terms of quantity and value, are registered at a very detailed level, for some food groups even at brand level. The data covers the period 1999-2004. The purchase data are concatenated with nutrition data from the Food Composition Databank from the Danish Institute for Food and Veterinary Research.[12] The matrices in the Food Composition Databank give detailed information about the content of energy, 10 macro nutrients (protein, fats, carbohydrates, fibres …etc.), 18 vitamins (vitamin A, vitamin b12…etc.) and 13 minerals (calcium, sodium…etc.) for each food item in the databank[13]. Concatenating purchase data with information from the nutrient database at the most detailed level possible leads to data on purchased quantities of many different nutrients for the individual households. Additional to the purchase registrations the households fill in an annual questionnaire concerning their social and demographic characteristics (e.g. family size, age and level of education of each household member, number of children, region, and income) media habits (e.g. preferred newspapers and magazines) and height, weight and exercise habits[14]. The panel is unbalanced since there is a replacement of 20 per cent of the households each year. For some purposes a subset of the data is created consisting only of the households which stay there for the entire period (1999-2004), approximately 10 per cent of the entire panel. The two versions of the panel are referred to as the unbalanced and balanced panel, respectively.

Empirical construction

As the data at hand only cover in-house consumption of foods, it is necessary to state a set of assumptions about the composition of food consumed outside the house, i.e. to assume how food eaten in canteens, restaurants and fast-food outlets differs compared to food eaten at home. In general, despite the rising share of food eaten away from home, most food is consumed at home (Groth and Fagt, 2003). The most commonly eaten meal away from home is lunch. Research, based on data including food consumed away from home, shows that the nutritional composition of lunch does not vary substantially from the nutritional composition of breakfast and dinner (Groth et al., 1999). For a more rigorous discussion see Appendix 2.B. An exception is that young individuals consume a larger share of added sugar at in-between meals than at ordinary meals. This might cause minor biases in the results even though the data are adjusted for the smaller intake of sugar in the GfK panel compared to data from other

[12] http://www.foodcomp.dk/fcdb_default.asp
[13] In total there were 1032 items in the databank in 2005.
[14] Height, weight and exercise questions for each individual in the household are only filled in for the year 2004.

sources (see below). Based on this we assume that individuals having lunch outside home in canteens (as compared to bringing a lunch bag) compose their meal nutritionally more or less equal to their consumption at home.

Furthermore, the data cover only household purchases and not individual consumption. A refined way to cope with the fact that it is individual consumption which counts, but household data that are in hand, would be to use the method developed by Chesher (1999). In that study a non-parametric regression is used to smooth consumption and thereby jump from household data to individual consumption. Due to the fact that most households eat most meals together[15] we assume, that individuals belonging to the same household compose their diet in the same way. It is a rather different approach to use purchase data at household level to describe diet quality since most other studies describing diet quality are based on 24 dietary recall or 1-2 days of food records. One advantage of the dataset at hand is that it is possible to follow the same households over a long period and that price and total expenditure are included in the data. Moreover, the household reports exact weight of the purchased foods, instead of estimated portion sizes as normally used in food records and dietary recalls. Deliberate misreporting and especially underreporting of the consumption of unhealthy food is a problem in food intake studies based on self-reported consumption (Lafay et al., 2000; Krebs-Smith et al., 2000). In the data at hand the diary keeper reports the purchases for the whole household and for a longer period of time (most households purchase foods for more than one day at a time). This means that the connection between the nutritional value of the purchased foods and the single meals is less obvious and therefore deliberate adjusting is assumed to be diminished. We do not include the overall calorie intake in the HEI, which is due to several causes. First of all it is not known how much the individuals exercise over time[16], which is one of the principal determinants for a precise recommended calorie intake. Furthermore, the amount of calories might be a determinant of whether there is an excess intake of calories, but not a determinant of how healthy a diet is, since no correlation is found between diet quality and total calorie intake (Patterson et al., 1994; Haines et al., 1999).

[15] 41 per cent of the Danish households eat breakfast together and 81 per cent eat dinner together every day (Groth et al., 1999). The same numbers for two-person households in our panel dataset are 42 per cent and 84 per cent, respectively. Of the remaining part of the panel 12 per cent eat dinner together 3-4 days a week, and 10 per cent eat breakfast together 3-4 days a week.

[16] The height, weight and exercise questions are only collected for 2004.

The six included recommendations are shown in the last column of Table 2.1. All recommend levels are computed in energy shares due to the necessary assumption that individuals belonging to the same household compose their diets in the same way. Furthermore, the average intake of various nutrients, fish, fruit and vegetables between the panel from GfK data and other data sources is compared both in grams and in energy shares in order to validate the data.

Table 2.1: Nutrient and food intake

	Average in GfK data – unbalanced panel (2002)	**Average from other sources**	**Official guideline**
Fruit and vegetables	245 g. a day per person (3.9 % of total energy intake)	379 g. a day per person (5.6 % of total energy intake) (Fagt et al., 2002)	600 g. daily per adult, 3-400 g. per child (8.5 % of total energy intake)[1]
Fibre	1.7 % of total energy intake	22 g. per day (1.98 % of total energy intake) (Fagt et al., 2004)	3 g. per MJ = 2.4 % of total daily energy intake
Fats	35.2 % of total energy intake	33 % of total energy intake (Fagt et al., 2002)	Maximum 30 % of total energy intake
Saturated fat	13.0 % of total energy intake	15 % of total energy intake (Fagt et al., 2002)	Maximum 10 % of total energy intake
Added sugar	6.4 % of total energy intake	9 % of total energy consumption (Mølgaard et al., 2003)	Maximum 10 % of total energy intake
Fish	1.5 % of total energy intake	20.5 g. a day per person (1.5 % of total energy intake) (Groth and Fagt, 2003)	200-300 g. a week (3.1 % of total daily energy intake)[2]
Not in the HEI			
Protein	14.5 % of total energy intake	16 % of total energy intake (Fagt et al., 2002)	
Carbohydrates	48.6 % of total energy intake	48 % of total energy intake (Fagt et al., 2004)	

1 This number is calculated as the average energy density for fruit and vegetables in the purchase data (1.45 KJ/g) multiplied by the recommended 600 g. per day and divided by the average recommended intake of kilojoules per day per person (10,235 KJ). In the HEI calculation household-specific energy densities according to type of vegetable consumed in the household are used together with approximated specific energy requirements for this particular household type.

2 This number is calculated as the average energy density for fish in the purchase data (7.42 KJ/g) multiplied by the recommended 200 g. per week and divided by the average recommended intake of kilojoules per day per person (10,235 KJ). In the HEI calculation household-specific energy densities according to the type of fish consumed in the household are used together with approximated specific energy requirements for this particular household type.

The numbers for sugar and fruit and vegetables are lower in GfK data than the numbers from other sources. This might be due to that GfK data are mainly based on in-house consumption and therefore only to a limited extent cover in-between-meals. Fruit are often eaten in-between-meals, and in-between-meals have a somewhat larger share of sugar than other meals as explained in Appendix 2.B. Due to this, the recommended levels for the consumption of added sugar, fruit and vegetables are revised to match the lower coverage degree in the GfK data. The adjusted recommended level for fruit and vegetables is equal to the average in the panel divided with the average in the comparable data source multiplied with the original recommended level:

$$recommended_{adjusted} = {average_{panel}}\Big/{average_{other}} \cdot 600 \qquad (2.3)$$

Likewise for sugar:

$$recommended_{adjusted} = {average_{panel}}\Big/{average_{other}} \cdot 10 \qquad (2.4)$$

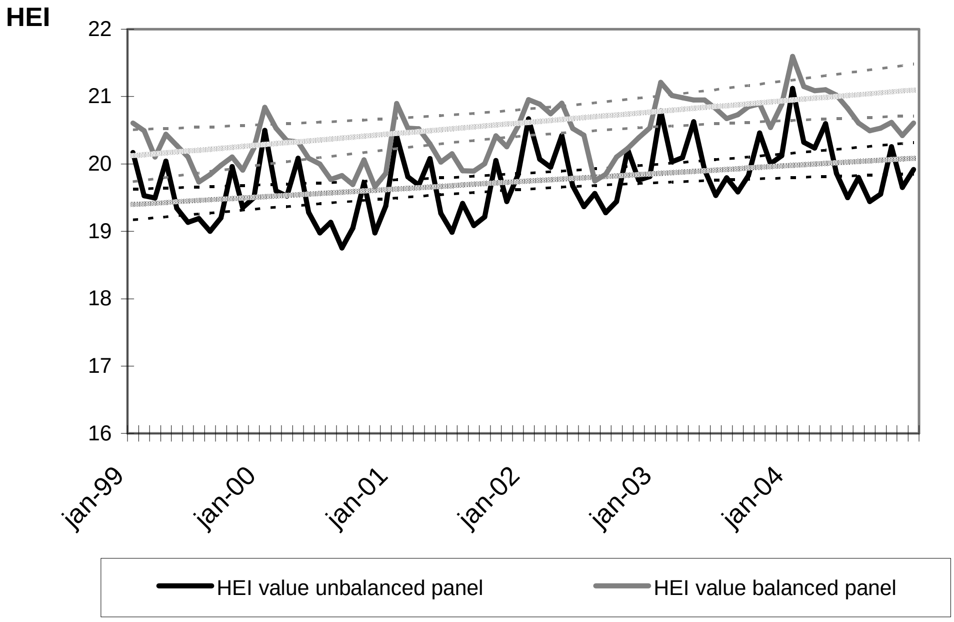
The recommended level of fish, fruit and vegetables is furthermore transformed into recommended energy shares. First, the recommended level of fish, fruit and vegetables in g. is calculated for the individual households according to the number of adults and children in the household. This is transformed into energy shares through a multiplication with the average energy density (KJ/g) for fish, fruit and vegetables, respectively, in this household. This number is then divided by the average gender and age specific recommended energy intake in KJ for the individual households (Kostplanen, 2008).

2.3 Development of the HEI value over time

This section describes the development in both the aggregated HEI value and the scores for each of the recommendations for both a balanced panel, consisting of households which stay for the whole period, and an unbalanced panel, consisting of all households[17]. From January 1999 to the end of 2004 there is a small increase in the index value as seen in Figure 2.1. More distinct is the strong seasonal variation with a peak in January each year, followed by a large decrease in the index value in the following months. This pattern is followed by all type of households. It is remarkable that the households in the balanced panel (i.e. households which are there for the entire period) eat healthier than households from the unbalanced panel. One explanation for this would be that the households which stay there for a long time are the households with a high level of self-control, which, apart from being able to register their purchases each week, are able to control their food intake too. Another explanation would be that this group of households is influenced during their stay in the panel. As the questions posed to the panel during our data period are not concentrated around how healthy their diet is we assume that the first explanation the most plausible.

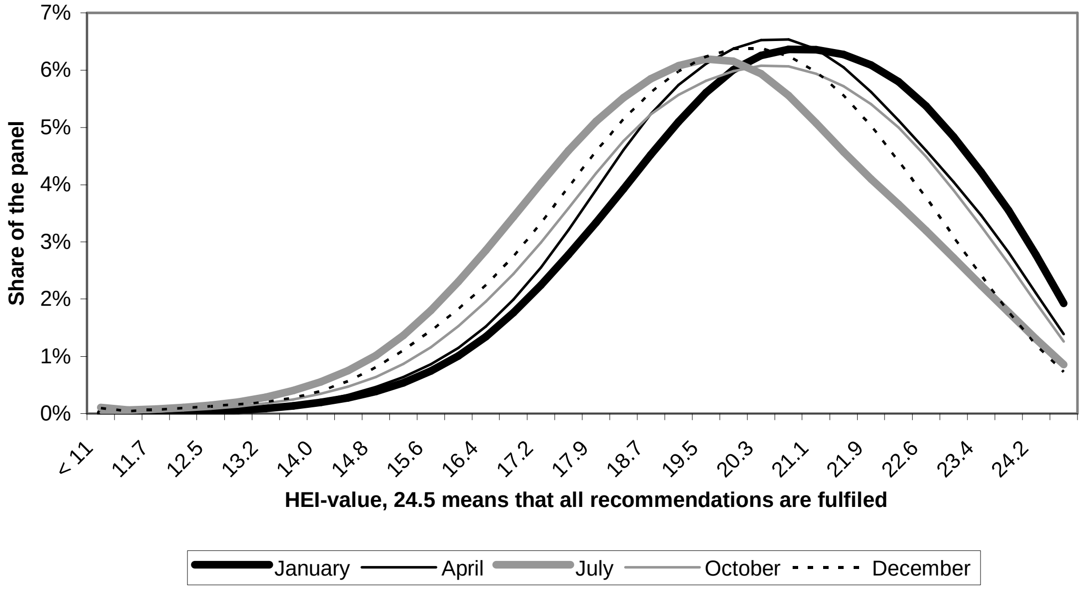
[17] In the dataset described in section 2.2. there is a 20% replacement of households each year

Figure 2.1: Developments in the HEI value over time



The observed increase in the HEI value is based on an increase in the scores for fat, saturated fat and fruit and vegetables and the seasonal variation, e.g. the peak in January is based on seasonality in all components despite fish. The lowest value of the HEI seems to be during summer. This is especially driven by a large seasonality in the score for fruit and vegetables which is considerably smaller during summer. This might be due to self-production, i.e. missing reporting, but as the lowest value of the HEI is found in May and June where there is only a smaller amount of Danish produced vegetables this does not seem plausible. In Appendix 2.C. figures for the average scores over time for each of the six recommendations are shown. Figure 2.2 shows the distribution of the panel over different HEI values for each month of the year 2002 to underpin the seasonal variation. The distributions are calculated as kernel densities with a Gaussian kernel (Blundell and Duncan, 1998). For the clarity of the figure only five months are illustrated. These are January, April, July, October and finally, December. This clearly shows that especially in January there is a larger share of the panel having a high HEI value, while there is a much larger share of the panel in the lower end of the HEI scale during July. December is almost as bad as July. What is even more interesting is that almost all households change over the year, since it is the whole distribution that moves, not only those who perform well or those who perform badly in relation to the diet recommendations.

Figure 2.2: The development in the distribution of the HEI value over the year



It is easier to fulfil the dietary recommendations for one day, one week or one month than it is to fulfil it for all months in a year. The figures below show the yearly average HEI values i.e. to fulfil the recommendation in a year they have to be fulfilled each month in that year. Studies describing dietary patterns are often based on 24 dietary recall or 1-2 days of food records. So in order to be able to compare the computed numbers with other research the share of the panel fulfilling the recommendations in an average month is likewise mentioned, but not illustrated. Less than 1 per cent of the panel fulfil all the diet recommendations in 2004 if the yearly average HEI value is calculated (Figure 2.2). This number has increased a little bit from 1999 to 2004. To compare; 1.2 per cent of the unbalanced and 1.5 per cent of the balanced panel fulfil all the diet recommendations in an average month. This is in accordance with Patterson et al. (1994) who find that 2 per cent of the respondents fulfil the recommendations according to the diet quality index based on data from a 3-day food record. Average scores for each recommendation are calculated for 2004 and 1999 for each household. Figures 2.3 to 2.5 show distributions over different values for both the balanced and the unbalanced panel for selected recommendations. Again due to that it is easier to fulfil the dietary recommendation for one day, one week or one month than over a whole year the monthly average is mentioned in the text to compare with other studies even though it is yearly averages shown in the figures. Only 5 per cent in the unbalanced and 6 per cent of the balanced panel fulfil the recommendation in 2004 if the yearly average scores are calculated as illustrated in Figure 2.3. This number has doubled from 1999 to 2004. In an average month (not illustrated) around one third of both the balanced and the unbalanced panel fulfil the diet recommendation of maximum 30 energy per cent from total fat. This can be compared to Basiostis et al. (2002) who find that 38 per cent of the Americans fulfil the recommendations of only 30 energy per cent from fat in a study based on a 1-day dietary recall. There is also an increase in the score for saturated fat[18] for both the unbalanced and the balanced panel. Around 6 per cent of both the unbalanced and the balanced panel fulfil the recommendations for saturated fat in 2004 if the yearly average scores are calculated as illustrated in figure 2.4. In an average month (not illustrated) around 21 per cent of the households in the balanced panel fulfil the recommendation of max. 10 energy per cent from saturated fat where Basiostis et al. (2002) find that 41 per cent fulfil the recommendations based on a 1-day dietary recall.

Figure 2.3: The development in the distribution of the score for fat

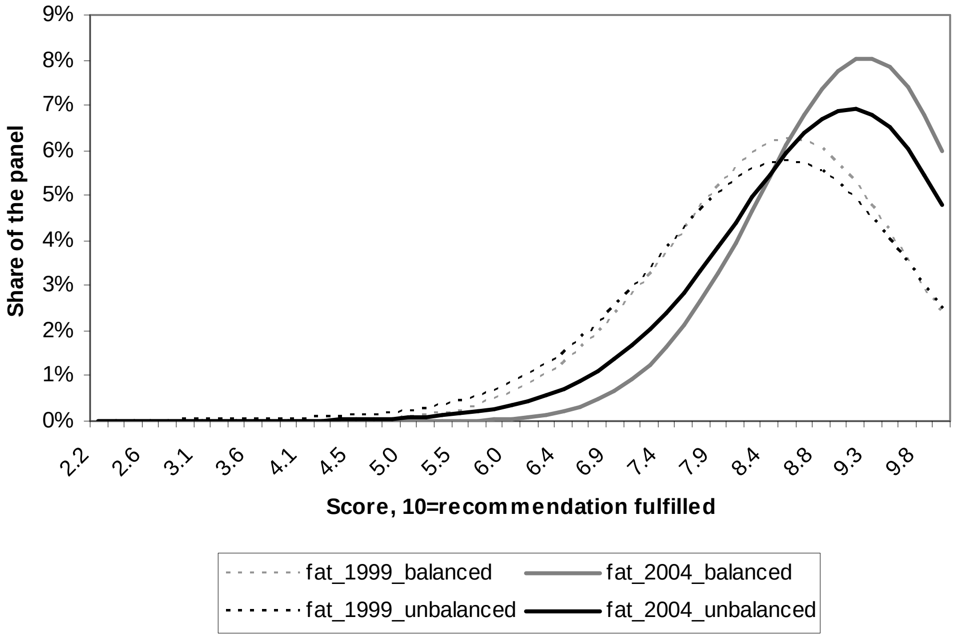


Figure 2.4 shows the distribution for the scores for fish. In 2002 there are more households being close to fulfilling the recommendations than in 1999, but there are also more that are far from fulfilling. This applies to both the unbalanced and the balanced panel. In an average month (not illustrated) a little more than one third of the panel fulfil the recommendation of 2-300 gram of fish each week, but only 2.8 per cent and 1.7 per cent of the balanced and unbalanced panel, respectively, fulfil the recommendation in 2002 if the yearly average scores are calculated. This number has increased for the balanced panel from 1999 and to 2002, but not for the unbalanced.

[18] Figures are shown in Appendix 2.D

Figure 2.4: The development in the distribution of the score for fish

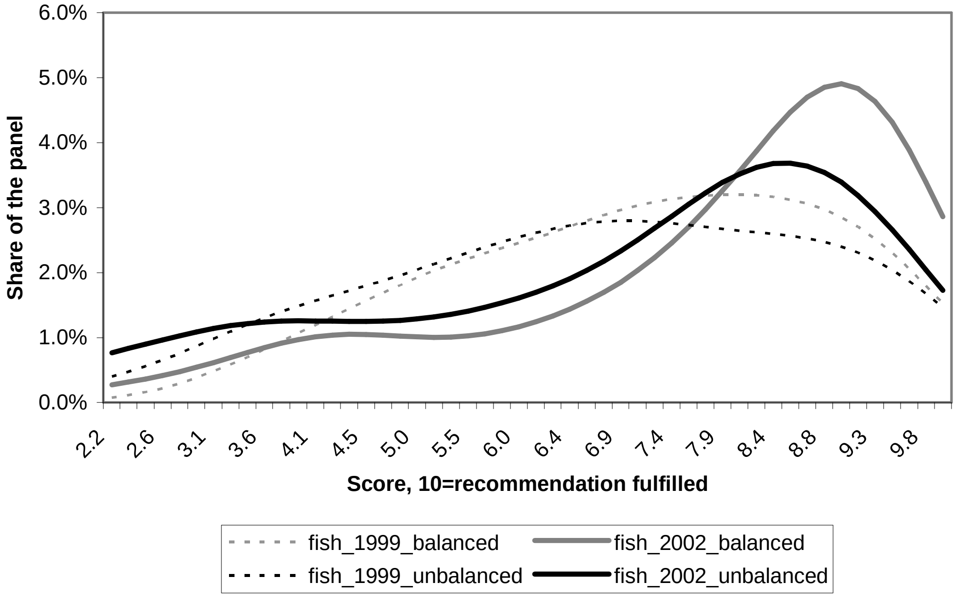
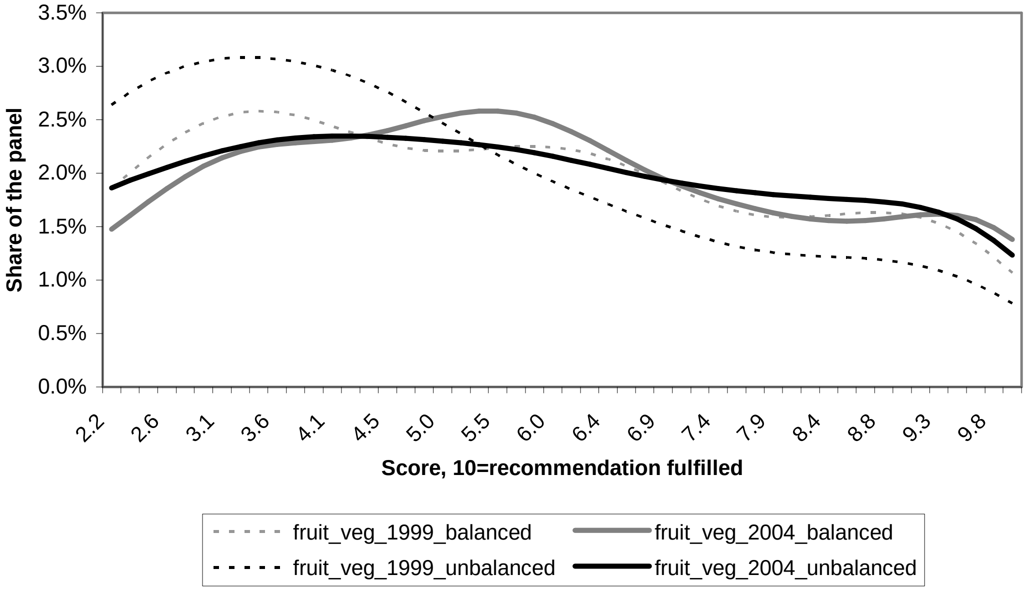


Figure 2.5 shows the distribution of the panel over the score for fruit and vegetables. There have become more consumers having a high consumption of fruit and vegetables from 1999 to 2004 and it seems like the households that had a very low consumption have moved to a more moderate consumption. There are slightly more households fulfilling the recommendation of 600 gram a day in 2004 than in 1999. 1.4 per cent and 1.2 per cent of the balanced and unbalanced panel, respectively, fulfil the recommendation in 2004 if the average scores are calculated, while around 20 per cent of the panel fulfil the recommendation in an average month (not illustrated). This can be compared with Basiotis et al. (2002) who find that 45 per cent of the Americans fulfil the recommended number of servings of fruit and vegetables (in the US the recommended level is 5 a day) in a 1-day dietary recall study. The consumption of fibres (not illustrated) has decreased while the consumption of sugar is almost stable from 1999 to 2004[19]. Very few households (below 1 per cent) fulfil the recommendation for fibres in 1999 and 2004 if the average scores are calculated, while almost 10 per cent of the panel fulfils the recommendation of 3 gram of fibres per MJ in an average month (not illustrated). In an average month 75 per cent of the balanced panel fulfil the adjusted recommendation for the intake of sugar (not illustrated), while only 7.7 and 8.4 per cent of the balanced and unbalanced panel respectively fulfil the recommendation in 2004 if the yearly average scores are calculated.

[19] Figures are shown in Appendix 2.D.

Figure 2.5: The development in the distribution of the score for fruit and vegetables



2.4 Variation of the HEI over social and demographic variables

In this section we look at the variation in how well different social and demographic groups perform in relation to the diet recommendations and how their diets change from 2000 to 2004. Generally, a lower HEI value is found among rural households, lower or non-educated and younger age groups. This is equivalent to findings based on the use of the USDA's Healthy Eating Index (see e.g. Guo et al., 2004; Variyam et al.,1998; Basiotis et al., 2002). In Table 2.2 the share of the panel which fulfils all the dietary recommendations in an average month is calculated. On average, around 12.1 per cent of the panel are found to have a diet close to all the diet recommendations (HEI value at 22.5 or above, which is equivalent to having a score of 9.2 for each recommendation), and 1.2 per cent of the panel fulfil all the recommendations. Households with any kind of theoretical education seem to be more likely to fulfil the recommendations as there are 18.4 per cent of the households almost fulfilling the recommendations. There is one fifth in the bad end for households where the head has no further education or a vocational education. Households in the capital have a much larger share of households in the better end; 1.9 per cent of the households fulfil all the recommendations and only 12.3 per cent are in the lower end. Apart from this there is a tendency that in rural areas there are more households in the lower end than in urban areas. A clear age pattern is also seen in Table 2.2 as more than 20 per cent of the younger households are in the lower end of the scale and less than 1 per cent fulfils all the recommendations.

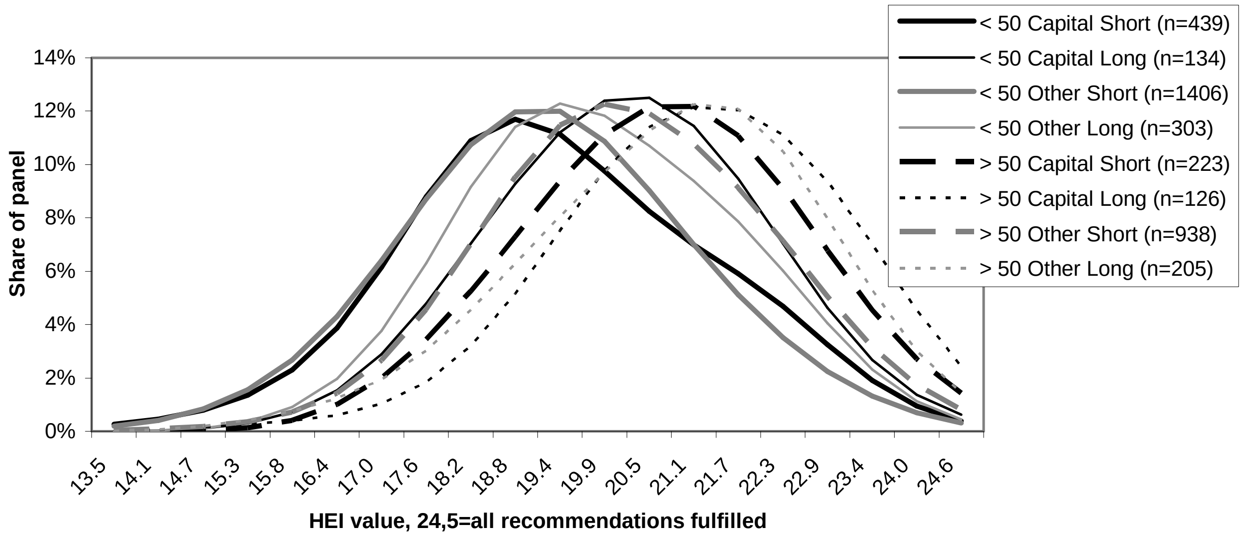
Table 2.2: Consumers with low, middle-low, middle-high and high HEI values

	Below 17.5	17.5-19.99	20.0-22.49	22.5-24.49	All recommenda-tions met
All	16.6 %	35.8 %	34.3 %	12.1 %	1.2 %
Education of the diary keeper					
None	18.1 %	36.9 %	33.8 %	10.3 %	0.9 %
Vocational	18.3 %	37.3 %	32.7 %	10.7 %	1.0 %
Short theoretical	16.0 %	35.6 %	34.3 %	12.4 %	1.7 %
Medium theoretical	10.9 %	31.1 %	37.9 %	18.0 %	2.0 %
Long theoretical	10.9 %	28.3 %	41.0 %	18.4 %	1.5 %
Regional location					
Capital area	13.4 %	32.3 %	36.5 %	15.9 %	1.9 %
Urban, East	15.6 %	35.6 %	35.3 %	12.4 %	1.2 %
Rural, East	18.1 %	37.1 %	32.7 %	10.9 %	1.2 %
Urban, West	16.8 %	36.4 %	34.4 %	11.6 %	1.0 %
Rural, West	20.0 %	38.3 %	31.8 %	9.1 %	0.8 %
Age of diary keeper					
Below 30 yrs	21.9 %	39.7 %	31.1 %	6.8 %	0.6 %
30-39 yrs	23.6 %	41.8 %	27.8 %	6.4 %	0.5 %
40-49 yrs	21.1 %	41.1 %	29.5 %	7.5 %	0.7 %
50-59 yrs	13.0 %	32.3 %	37.3 %	15.1 %	1.8 %
60 yrs and above	12.3 %	31.5 %	38.6 %	16.1 %	1.5 %

Initial analyses have been used to define eight main types of households according to how healthy they eat.[20] These are longer educated (medium and long further education) versus shorter educated (no, vocational or short further education), households aged 50 yrs or above versus households below 50 years and finally capital location versus other location. The mean values for each of these groups are compared in Appendix 2.E. Figure 2.6 shows the distribution of the panel over HEI values in 2004 for these eight socio-demographic groups. Three distinct groups manifest themselves in Figure 2.6. These are longer educated households above 50 yrs, who eat healthier than the rest, short educated households below 50 which eat unhealthiest of all households and all the others in the middle. For each age and regional group it is clear that education matters in a healthier direction and there are also significant differences between means as shown in appendix 2.E. Also age is a determinant for the healthiness of diets. In each region and educational group older have a healthier diet than younger and there are also significant differences between means as shown in appendix 2.E. It is remarkable that the healthiest of the youngest (the long educated) eat just as healthy as the unhealthiest of the oldest (the short educated). This indicates that diet habits will worsen over time if this is a cohort effect and the younger does not change habits as they grow older.

[20] This analysis is seen in Appendix 2.E

Figure 2.6: The density of the HEI value for selected social and demographic groups



In the following figures the average scores in 2004 are shown for selected recommendations for these eight main types of households. Figure 2.7 shows the density over the panel for the score for fish. It is clear that the older are closer to the recommendation for fish than the younger ones and that households in the capital are generally closer to the recommended level than households in other regions. Education does not matter much for the proximity to the recommendation for fish especially not for the older households, for younger there is a tendency that longer educated are closer to the recommendations than shorter educated.

Figure 2.7: The density of the score for fish in selected social and demographic groups

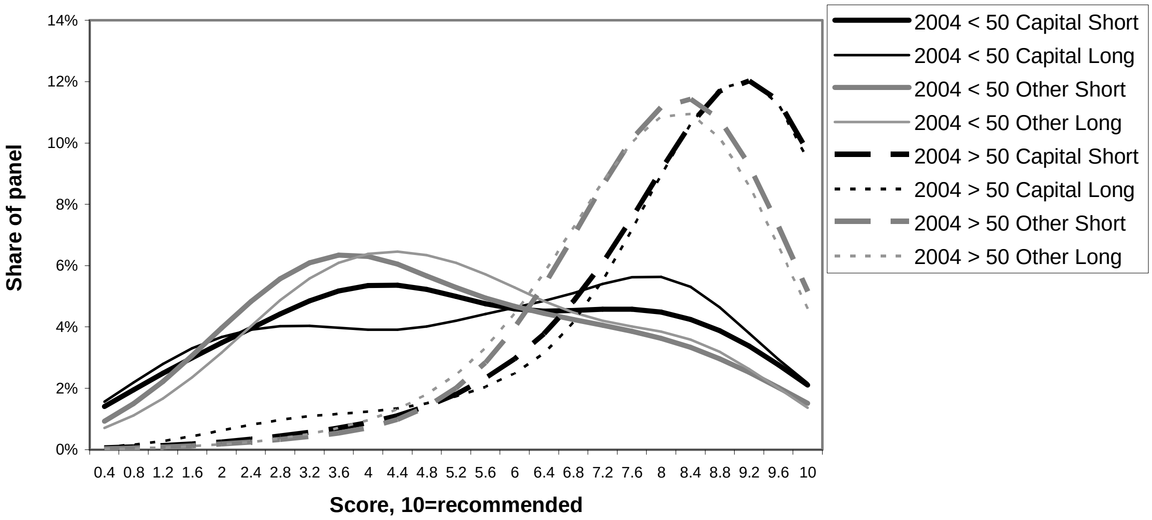
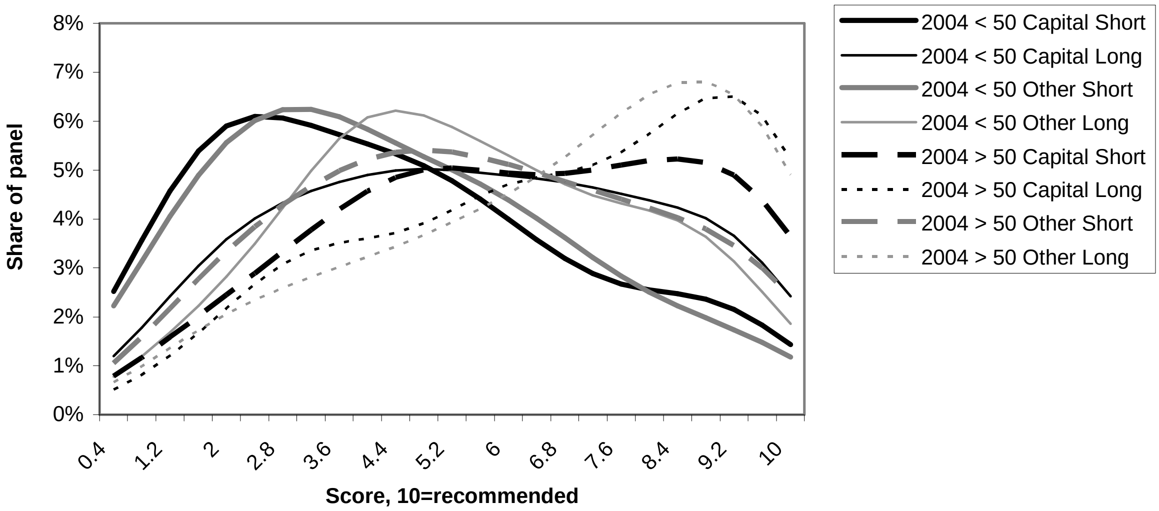


Figure 2.8 shows the density of the panel for the score for fruit and vegetables. Education matters a lot with longer educated having far more households in the better end of the scale

than the shorter educated. Location does not really matter for the proximity to the recommendation for fruit and vegetables. Older households also perform better than younger households everything else equal.

Figure 2.8: The density of the score for fruit and vegetables, selected social and demographic groups



From Figures 2.9 and 2.10 it is evident that it is the longest educated households which are closest to fulfil the recommendations for fat and fibres. There is also a tendency that younger eats less fat than older households.

Figure 2.9: The density of the score for fat in selected social and demographic groups

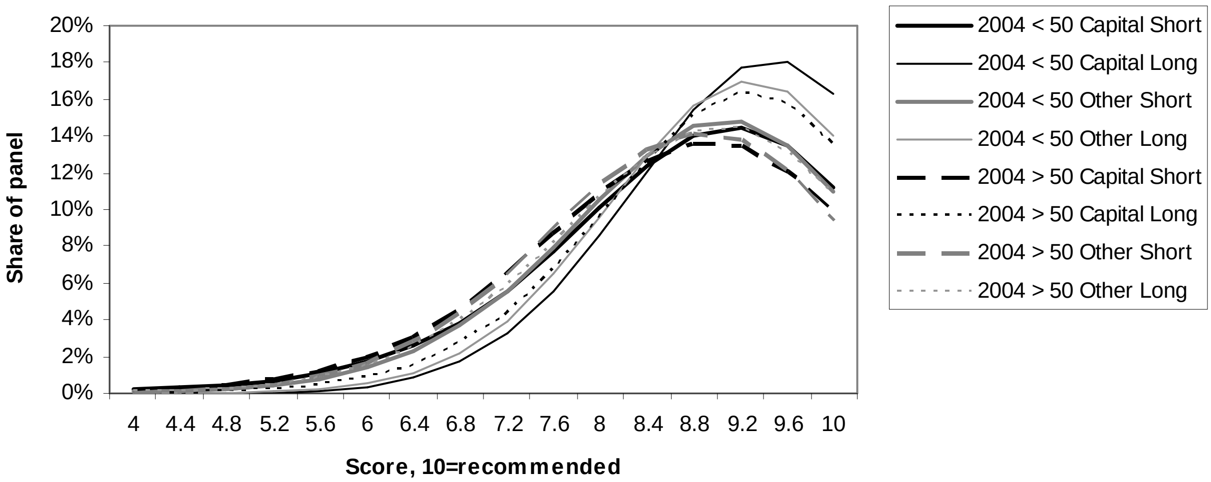
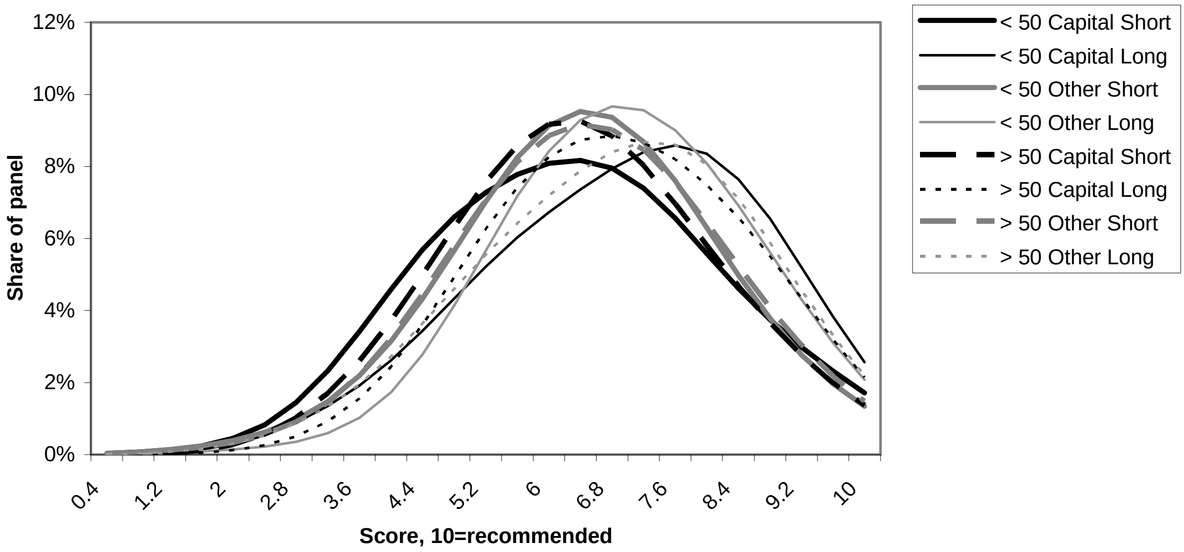


Figure 2.10: The density of the score for fibre in selected social and demographic groups



In the following figures the change in average scores and HEI value from 2000 to 2004 are calculated. 2000 is chosen as the basic year since it is necessary that the households stay in the whole period to calculate the change over time (*n* = 961). The 10th percentile, the mean and the 90th percentile are shown. The households' age in 2000 determines which age group the household belongs to. Only a minor part of the households changes location or education. These are deleted. Figure 2.11 shows the mean change in HEI value for the socio-demographic groups together with the 10 per cent with the lowest change value and the 10 per cent with the largest change. This gives a picture of in which types of households diets are improving or worsening most. There is no significant change in means for the younger households (the value is not significantly different from 0) while older households on average have improved their diet. Largest decreases are found among the younger and especially among the shortest educated younger households, while largest improvements are found between older households. Generally diets have improved less in those groups who initially were identified as being farthest away from the recommendations and improved most for those who were identified to have the healthiest diet. The changes in the scores for fish in Figure 2.12 are rather unambiguous. The mean change is positive for the older and negative for the younger households (the decline in consumption for the younger with a long education is not significantly different from 0). Educational differences are significant different from each other in the capital, while no educational differences are seen in other regions. In general, those who are identified as consuming the least fish, younger and short educated have had a large decline in consumption.

Figure 2.11: Change in average HEI value from 2000 to 2004

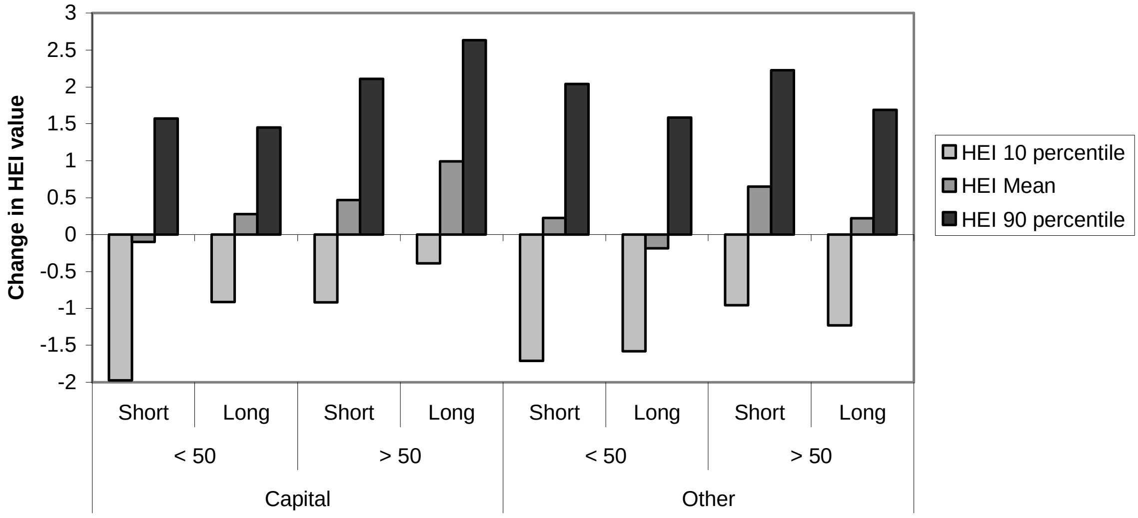


Figure 2.12: Change in average score for fish from 2000 to 2002

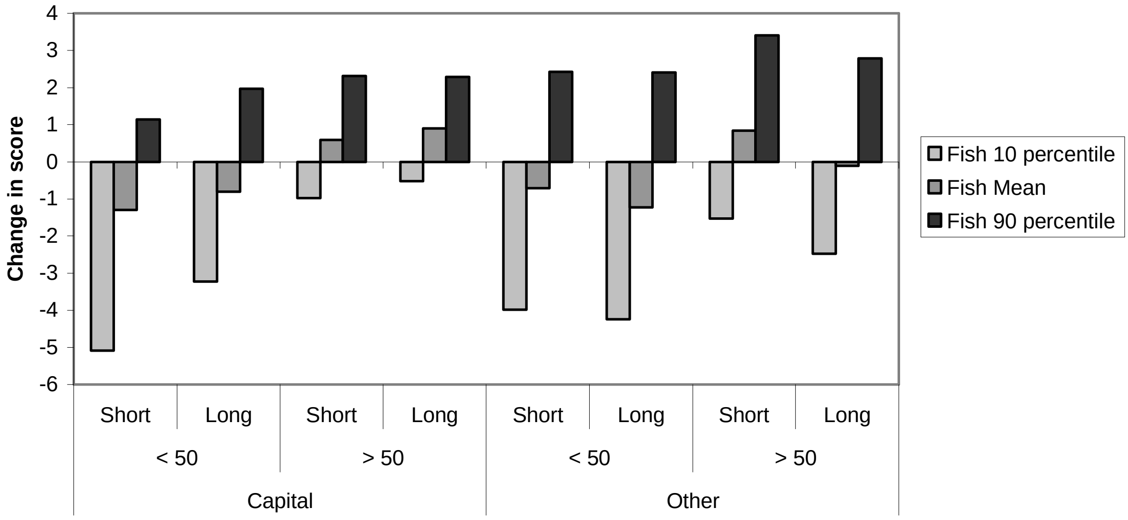


Figure 2.13 shows the changes in the scores for fruit and vegetables. There are rather strong regional differences. On average, there has been an increase for longer educated households in the capital, but this is not the case in other regions where the mean increase has been larger for the shortest educated. The 10 per cent of the older households in the capital with the largest increase in consumption have had a considerable improvement. In other regions it is the shortest educated who have the largest increase in fruit and vegetable consumption. Generally the mean changes are found statistically different from each other for educational and age groups. Those with the largest decline in consumption are the shortest educated in the capital and the younger in other regions.

Figure 2.13: Change in average score for fruit and vegetables from 2000 to 2004

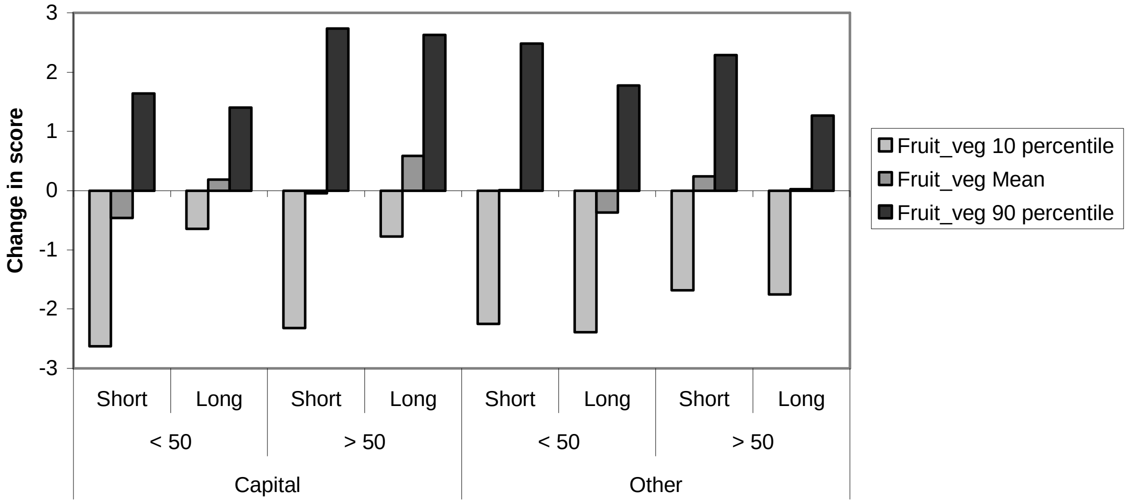
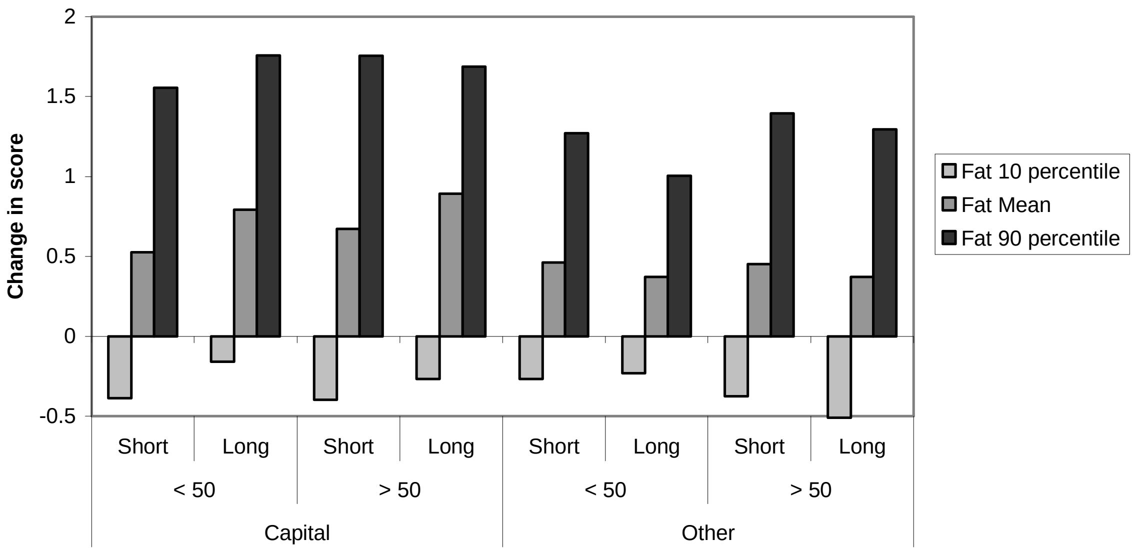


Figure 2.14 shows the change in the score for fat. The mean shows an increase (decreased consumption) for all types of households and this increase is significantly different from 0 for all types of households. In the capital region the mean shows a significant larger increase for the longer educated versus the shorter educated. Regional patterns are found for the mean and the 10 per cent with the largest increase are found for households in the capital compared to households in other regions. Generally, there are larger improvements in the diets concerning fat than for any of the other diet recommendations. This applies to all types of households. No specific patterns are found for the change in the scores for fibres, sugar and saturated fat. Generally, the mean is positive in all groups for saturated fat while the mean is close to 0 for sugar and negative for fibres.

Figure 2.14: Change in average score for fat from 2000 to 2004



2.5 Cross-section estimation with HEI as dependent variable

The tables and figures above describe very detailed the differences in how well different social and demographic groups perform according to the diet recommendations and show the variation within groups. The analysis revealed that a household's HEI improves with age, education and urbanity. But as the analyses do not control for the influence from other social and demographic variables, these findings may change in a statistical analysis of the data where the influence of one variable can be separated from the effects of other variables. In order to isolate the effects a simple cross-section on yearly averages for 2004 is estimated. The dependent variable, the HEI value, takes on values ranging from 0 to $\sqrt{600}$. Given this fixed range, a question may arise as to whether a transformation of HEI is necessary to ensure that the predicted values are also bound within this range. This is, however, not applicable here because there is no natural way of introducing the limits since the HEI is a number not a proportion, a probability or a percentage. The equation to be estimated is as follows:

$$HEI_n = \sum^{J} \alpha_j \cdot reg_{nj} + \sum^{J}\sum^{M} \beta_{mj} \cdot C_{nm} \cdot reg_{nj} + \sum^{J}\sum^{M} \beta_{2mj} \cdot (C_{nm})^2 \cdot reg_{nj} + \sum^{J}\sum^{R} \gamma_{rj} \cdot D_{nr} \cdot reg_{nj} + \varepsilon_n \qquad (2.5)$$

where the HEI for the individual households *n* is a function of *J* region-specific dummies $j \in (Capital, other)$[21], *M* continuous variables (named *C*) (the continuous variables are squared in initial regressions) and *R* different dummy variables (named *D*) all crossed with a region-specific dummy. This is in order to investigate whether there are systematic differences in the effects from socio-demographic characteristics in different regions. An overview of the explanatory variables is presented in Table 2.3 below. Since not all households have answered the questions concerning BMI, exercise habits and attitude towards own weight and dieting practices two different versions of the model are estimated. Model 1 where the BMI questions on individual basis are coupled with the household information. The estimations are done at individual level $(i = 2624)$. Model 2 where the BMI questions are excluded and the estimations are made at household level $(i = 2599)$. Wald tests are performed to test for equality of parameters across regions (i.e. $\alpha_j = \alpha_k, \beta_j = \beta_k$ and $\gamma_j = \gamma_k$ in equation 2.5).

[21] In the original estimations there were 5 different regions; capital, urban east, rural east, urban west, rural west, but there were not significant differences between the parameters for each of the regions outside the capital.

Table 2.3: The variables used in the cross-section estimation of the HEI

Variable	**Basis**	**Label**			
Dummy for whole-period stay in panel	Not in the panel the whole period	Being in the panel for the whole period			
Children	No children	Children 0 to 6 years old	Children 7 to 20 years old		
Sex of main buyer	Woman	Man	Both		
School education of diary keeper [1]	7-10 years	High school or equivalent			
Further education of diary keeper	None	Vocational oriented	Short theoretical	Medium theoretical	Long theoretical
Number of persons in the family	Single female	Single male	Couple		
Occupational status	Employed	Early pensioner[2]	Unemployed		
Age of diary keeper	Continuous				
Food expenditure per person	Continuous				
Extra variables only used in model 1					
BMI (only adults)[3]	Continuous				
Exercise in minutes[4] *(only adults)*	Continuous				
Have you changed your diet to lose weight	No	Yes			
How is your attitude to your own weight	No problems with it	Ought to do something			

1 There is a diary keeper in each household. Age, education and other socio-demographic variables are registered for both the diary keeper and other members of the household. The diary keeper is mostly also the main buyer.
2 Occupation as pensioner is inserted only for those becoming a pensioner before the natural retirement age at 65 years

4. BMI is calculated as: $BMI = \frac{weight(kg)}{height(m) * height(m)}$.

3. This variable is composed of two variables, sports and moderate exercise and is measured as minutes/week

For some variables the effect from social and demographic variables varies between the capital and the other regions, as it appears from the final estimation results in Table 2.4. The dummy for being in the panel for the whole period shows that those who stay the entire period 1999-2004 eat healthier than those who are in the panel a shorter period of time, but this is not significant in model 2. The presence of a male in the households decreases the healthiness of the diet since households consisting of a single male eat less healthy than the basis (a single female) and also households consisting of a couple eat less healthy than a single woman.

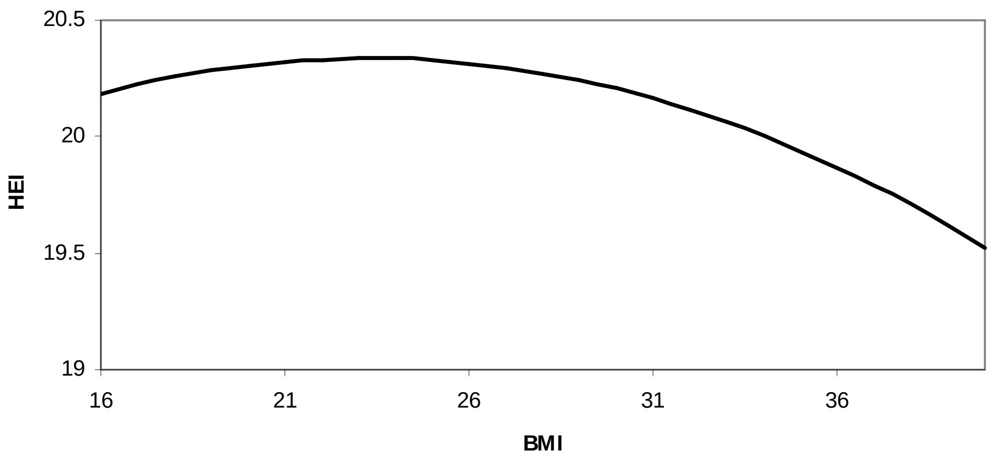
Table 2.4: Parameter values for a cross section estimation with the HEI as dependent variable

	MODEL 1			MODEL 2		
	Basis: Single female, no children, no high school, no further education, employed, female main buyer, satisfied with own weight, no dieting practice, no problem with own weight			Basis: Single female, no children, no high school, no further education, employed, female main buyer		
Parameter	Estimate	Std Err	Pr>|t|	Estimate	Std Err	Pr>|t|
Constant	16.1773	0.8300	<.0001	17.4754	0.2066	<.0001
Dummy for whole-period stay in panel	0.2163	0.0893	0.0155	0.0932	0.1130	0.4097
Single male	-0.5241	0.2232	0.019	-0.4931	0.2233	0.0273
Spouse	-0.5069	0.0821	<.0001	-0.4981	0.1354	0.0002
child_0_to_6	-0.1796	0.1033	0.0821	-0.0911	0.1061	0.3908
child_7_to_20	-0.3124	0.0681	<.0001	-0.2895	0.1670	0.0832
High_school	-0.1121	0.0811	0.1673	0.0141	0.0848	0.8676
Edu_vocational	-0.0017	0.0818	0.9838	0.0420	0.0846	0.6199
Edu_short_theoretical	0.1193	0.1036	0.2496	0.1694	0.1079	0.1168
Edu_medium_theoretical	0.2483	0.1039	0.017	0.3584	0.1087	0.0010
Edu_long_theoretical	0.4663	0.1598	0.0036	0.4708	0.1575	0.0028
Occupation_early_pensioner	-0.1393	0.0881	0.1138	-0.2160	0.0979	0.0275
Occupation_unemployed	-0.1359	0.1311	0.2999	-0.1956	0.1378	0.1560
Age_diarykeeper_capital*	0.0160	0.0046	0.0005	0.0241	0.0051	<.0001
Age_diarykeeper_Other*	0.0095	0.0038	0.0124	0.0080	0.0040	0.0436
Both_buyer	-0.1843	0.0766	0.0162	-0.1508	0.0855	0.0779
Male_buyer_capital	-0.5372	0.1486	0.0003	-0.4750	0.1248	0.0001
Male_buyer_Other	-0.2411	0.2538	0.3423	-0.7488	0.1867	<.0001
Expenditure_pers_capital*	0.0034	0.0003	<.0001	0.0030	0.0003	<.0001
Expenditure_pers_Other*	0.0043	0.0002	<.0001	0.0044	0.0002	<.0001
Body_Mass_Index*	0.1350	0.0617	0.0288			
Body_Mass_Index_sq*	-0.0029	0.0012	0.012			
Min_of_exercise*	0.0000	0.0001	0.9282			
Attitude_own_weight	0.1807	0.2376	0.4471			
Active_change_weight	0.1094	0.0597	0.0672			
	Adjusted R^2=0.4380 n=2624			Adjusted R^2=0.4157 n=2599		

* Continuous variable.

The effect of being in a couple is almost as negative as being a single male. This indicates that males have a negative influence on how healthy the female's diet is. Small children (0-6 years) have no significant effect on how healthy the households eat (not having children is basis), while children 7 to 20 years old have a significantly negative effect. Having a high-school education is not significantly different from having only 10 years of schooling (7-10 years of schooling is basis) and vocationally oriented and short theoretical education are not significantly different from no further education (no further education is basis). Medium and longer theoretical educated eat more healthy. Early pensioners eat less healthy than those who are employed (employed is basis) and unemployed also eat less healthy, but this is not significant. Healthy eating increases with age and this effect is double as large in the capital as in other regions. Male buyers have a negative influence on healthiness. If both the male and the female are main buyers, healthiness decreases (a female main buyer is the basis) and if only the male is the main buyer healthiness decreases even more. This effect is larger in other regions than in the capital. There seems to be a rather strong positive correlation between the amount of money in DKK spent per person on food and the healthiness of the diet (the R^2 is only equal to 0.2180 if expenditure is removed from the equation in model 1). This connection is stronger in other regions than in the capital. There is a quadratic correlation between BMI and the healthiness of the diet. Up to a BMI of 23.47 the HEI is increasing then the HEI is decreasing as shown in Figure 2.15.

Figure 2.15: The correlation between HEI and BMI for a basic household



The amount of minutes of exercise on a weekly basis and the attitude towards own weight (no problems with own weight is basis) do not significantly influence the diet; while those who state they have done something actively to change their diet in order to lose weight actually eat healthier (not doing anything is basis). If the basic household, a single woman, with no children, fulltime occupied, only 10 years of schooling and no further education has a BMI equal to 22, is 21 years old (the minimum age for households in the panel), uses 1000 DKK a month on food and is living in the capital she will have an HEI value at 21.5. If she had a spouse who were the main buyer, children above 6 years and a BMI of 30 she would instead have had an HEI value at 20.4.These results can be compared to the results in Variyam et al. (1998) who have done a cross-section estimation on the US-HEI. They find a positive significant influence of income, education and age on healthy eating. They also find that females eat healthier than males. A positive effect of being either part-time employed or unemployed is also found in some of the estimations in Variyam et al. (1998). The difference may be explained by differences in family structures between Denmark and the States with almost no women being housewives in Denmark. In the US women with part time or no employment might indicate a household with a good economic status. A negative and significant effect on healthiness from BMI is also found in the US-HEI study together with a negative but not significant influence from having children.

2.6 Discussions and conclusion

There seems to have been a minor improvement in the diet of the Danes from 1999 to 2004, but despite massive campaign activities from the Danish authorities still as few as 1.5 per cent fulfil all the diet recommendations. This number is considerably larger in January where as much as 2.2 per cent fulfil the dietary recommendations as compared to December where the number is as small as 0.1 per cent. The bad performance in December is especially caused by an excessive purchase of saturated fat and might be due to traditional eating during Christmas. There is a social gradient in obesity and other health-related illnesses, with lower educated, lower social classes and rural population having a larger prevalence. The same social gradient is found in diets since lower educated eat less healthy compared to longer educated and households in the capital are healthier than households elsewhere. There is found a significant quadratic relation between HEI and BMI. This indicates that the composition of diets matters when it comes to obesity not only the amount of calories in the diet and differences in dietary status might explain some of the social bias in obesity. Furthermore, there seems to be a rather strong gender influence with females eating healthier than males. The male influence seems to be stronger than the female influence since the effects for a woman of being in a couple is almost as strongly negative as the effect of being a male instead of female when single. This is correlated with the BMI as the BMI for men and women is equal at 18 years of age, but the BMI has a steeper increase with age for men than for women and reverses at a higher age for men than for women. The gender differences in healthiness of diet might reflect that generally women are more prone to react to information about healthy eating than men.

What is even more important is that older households are closer to fulfilling the diet recommendations than younger households. The effect of age is so dominating that the best educated of the younger households are at the same level of healthiness as the shortest educated of the older. Younger households also seem to have smaller improvements in the healthiness of diets in the studied period than older households. This is especially true for the short educated younger population. Also the presence of older children in households seems to decrease the healthiness of diets. This is rather important. If the negative generational effect overrules the general improvement in healthiness, this will have a bad influence on future health since the health effects of unhealthy eating are well documented. The differences in healthiness of diets across educational groups might reflect the fact that better educated households react more to general health information than shorter educated households or this might be caused by a less restrictive budget among higher income groups. The relation between food expenditure and healthiness of diets seems to be significant in the estimations and other research finds that lower educated and lower social classes are more responsive to changed relative prices between unhealthy foods and healthy foods. This indicates that changed relative prices might be a solution for improving the diet of the least well-off. But what explains the generational differences in healthiness? One suggestion might be to look at the increased use of convenience food especially among younger households due to time constraints and the rather poor variety of healthy variants of this type of food. Another suggestion might be that young households lack the skills of cooking. The findings in this paper give no answer to these questions and no solution, but this might be an important route of further research.

References

Andersen J.K., A. Büchert, B. Koch, O. Ladefoged, T. Leth, D. Licht and L. Ovesen (2003): *Helhedssyn på fisk og fiskevarer*. Fødevaredirektoratet, Copenhagen

Holm L., A. Astrup, A.W. Dynesen and J. Haraldsdóttir (2002): De store ernæringskampagner virker faktisk *Ugeskrift for læger*, 164(5): 649-651

Astrup A., N.L. Andersen, S. Stender and E. Trolle (2005): *Kostrådene 2005*. Publication no. 36, Ernæringsrådet & Danmarks Fødevareforskning, Copenhagen

Basiotis P.P., A. Carlson S.A. Gerrior, W.Y. Juan and M. Lino (2002): *The Healthy Eating Index: 1999-2000*. Working Paper, Centre for nutrition policy and promotion. USDA, USA

Blundell R. and A. Duncan (1998): Kernel regression in empirical microeconomics, *Journal of Human Resources*, 33(1):62-87

Brønnum-Hansen H., O. Andersen, M. Kjøller and N. K. Rasmussen (2004): Social gradient in life expectancy and health expectancy in Denmark. *Social and Preventive medicine*, 49(1):36-41

Chesher A. (1999): Individual demands from household aggregates: time and age variation in the composition of diet. *Journal of Applied Econometrics*, 13(5): 505-524..

Ekholm O., M. Kjøller, M. Davidsen, U. Hesse, L. Eriksen, A.I. Christensen and M. Grønbæk (2006): *Sundhed og sygelighed i Danmark, og udviklingen siden 1987*. Statens Institut for Folkesundhed, Copenhagen

Fagt S., J. Matthiessen, E.Trolle, N. Lyhne, T. Christensen, H.-J. Hinsch, H. Hartkopp, A. Biltoft-Jensen, A. Møller og A-S. Daae (2002): *Danskernes kostvaner 2000-2001 – Udviklingen i danskernes kost – forbrug, indkøb og vaner*. Fødevaredirektoratet Afdeling for ernæring, Søborg

Fagt S., J. Matthiessen, A. Biltoft-Jensen, M.G. Groth, T. Christensen, H.J. Hinch, H. Hartkopp,E. Trolle, N. Lyhne and A. Møller (2004): *Udviklingen i danskernes kost 1985-2001*. Danmarks Fødevare- og Veterinær Forskning, Copenhagen

Fagt S., M. Groth and A. Lyhne Andersen (2004): *Danskernes kostvaner 1995. Mad og Måltider*.Fødevaredirektoratet, Copenhagen

Fung T.T., E.B. Rimm, D. Spiegelmann, N. Rifai, G.H. Tofler, W.C. Willett and F.B. Hung (2001): Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *American Journal of Clinical Nutrition*, 73: 61-67

Groth M.V., N.L. Andersen, S. Fagt and L. Ovesen. (1999): *Danskernes kostvaner 1995. Mad og måltider*. Fødevaredirektoratet, Copenhagen

Groth M.V. and S. Fagt (2003): *Danskernes kostvaner – Måltidsvaner, holdninger, sociale forskelle og sammenhæng med anden livsstil*. Report no. 9, Fødevaredirektoratet, Copenhagen

Guo X., B.A. Warden, S. Paeratakul and G.A. Bray (2004): Healthy Eating Index and obesity. *European Journal of Clinical Nutrition*, 58 (12): 1580-1586

Haines P.S., A.M. Siega-Riz and B.M. Popkin (1999): The diet quality index revised: A measurement instrument for populations. *Journal of American Dietetic Association*, 99(6): 697-704

Haraldsdöttir J., L. Holm and M.L. Larsen (2005): *Befolkningens fødevareforbrug i 2004 sammenlignet med 1995, 1998 og 2001*. Rapport no. 37, Ernæringsrådet, Copenhagen

Hu F.B., E.B. Rimm, M.J. Stampfer, A. Ascherio, D. Spiegelman and W.C. Willett (2000): Prospective study of major dietary patterns and risk of coronary heart diseases in men. *American Journal of Clinical Nutrition*, 72(4): 912-921

Kant A.K. (1996): Indexes of overall diet quality - A review. *Journal of American Dietetic Association*, 96(8): 785-791

Key T.J., A. Schatzkin, W.C. Willet, N.E. Allen, E.A. Spencer and R.C. Travis (2004): Diet, nutrition and prevention of cancer. *Public Health Nutrition*, 7(1A): 187-200

Kostplanen (2008): http://www.kostplanen.dk/anbefalet.php

Krebs-Smith S.M., B.I. Graubard, L.L. Kahle, A.F. Subar, L.E. Cleveland and R. Ballard-Barbash (2000): Low energy reporters vs. others: A comparison of reported food intakes. *European Journal of Clinical Nutrition*, 54(4): 281–287

Lafay L., L. Mennen, A. Basdevant, M.A. Charles, J.M. Borys, E. Eschwege *et al*. (2000): Does energy intake underreporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Santé (FLVS) study. *International Journal of Obesity and Related Metabolic Disorders*, 24(11): 1500–1506

Lynch J., G. D. Smith, S Harper and K. Brainbridge (2006): Explaining the social gradient in coronary heart disease: comparing relative and absolute risk approaches. *Journal of Epidemiology and Community Health*, 60: 436-441

Maskarinec G., R. Novotny and K. Tasaki (2000): Dietary patterns are associated with body mass index in multiethnic women. *Journal of Nutrition*, 130: 3068-3072

Mølgaard C., N.L. Andersen, V. Barkholt, N. Grunnet, K. Hermansen, B. Nyvad, B.K. Pedersen, A. Raben and S. Stender (2003):*Sukkers sundhedsmæssige betydning*. Report no. 33, Ernæringsrådet, Copenhagen

Newby P.K., D. Muller, J. Hallfrisch, N. Qiao, R. Andres and K.L. Tucker (2003): Dietary patterns and changes in body mass index and waist circumference in adults. *American Journal of Clinical Nutrition*, 77(6): 1417-1425

Nordisk Ministerråd (2004): *Nordic nutrient recommendations NNR 2004 – integrating nutrition and physical activity*. Nordisk Ministerråd, Copenhagen, Denmark

Ovesen L., N.L. Andersen, L.O. Dragsted, J. Godtfredsen J. Haraldsdóttir, S. Stender, K. Sølling, A. Tjønneland and E. Trolle (2002): *Frugt, grønt og helbred – opdatering af vidensgrundlaget*. Report no. 22, Fødevaredirektoratet, Copenhagen

Patterson R.E., P.S. Haines and B.M. Popkin (1994): Diet quality index: capturing a multidimensional behaviour. *Journal of American Dietetic Association*, 94(1): 57-64

Richelsen B., A. Astrup, G.L. Hansen, H.S. Hansen, B. Heitman, L. Holm, M. Kjær, S.A. Madsen, K.F. Michaelsen and S.F. Olsen (2003): *Den danske fedme epidemi, Oplæg til en forebyggelsesindsats*. Report no. 30, Ernæringsrådet, Copenhagen

Richelsen B., N.L. Andersen, A. Flint, K. Hermansen, P. Marckmann, M. Osler and B.K. Pedersen (2005): *En faglig vurdering af den "omvendte" kostpyramide*. Ernæringrådet, report no. 35, Copenhagen

Smed S. (2008): Describing diets with purchase data, Chapter 1, this thesis

Variyam J.N., J. Blaylock, D. Smallwood and P.P. Basiostis (1998): *USDA's Healthy Eating Index and Nutrition Information*. United States Department of Agriculture/Economic Research Service; 1998. Technical Bulletin nº1866 Washington (DC)

WHO (2002): *Reducing Risks, Promoting Healthy Life. The World Health*. Report 2002, Geneva

WHO (2004): *Global strategy on diet, physical activity and health*. Geneva

Willett W.C. and P.J. Skerrett and E.L. Giovannucci (2001): *Eat, Drink, and Be Healthy: The Harvard Medical School Guide to Healthy Eating*. Simon & Schuster, New York

Willett W.C. and M.J. Stampfer (2003): Rebuilding the food pyramid. *Scientific American*, 288: 64-71

Appendix 2.A: Detailed description of the diet recommendations

This appendix gives a detailed description of the official Danish diet recommendations and discusses its relation to alternative ways of defining a healthy diet.

Ad 1) Eat 600 g. of fruit and vegetables

Fruit and vegetables contain vitamins, minerals, fibres and a lot of other healthy nutrients. On average, Danes eat almost 400 g. a day. This number has been rising during the last decade, but it is still too low and has been stagnating during the last couple of years. What is even more important is that one out of four eats less than 200 g. a day (Astrup et al., 2005). An increase in the consumption of fruit and vegetables will reduce the prevalence of cardiovascular diseases (Ovesen et al., 2002) and cancer (Key et al., 2004). In the relation to obesity a meal where fruit and vegetables substitute other forms of food will decrease the prevalence of obesity (WHO, 2004). In the healthy eating pyramid by Walter Willets fruit and vegetables are amongst the foods in the bottom implying that a large intake is recommended (Richelsen et al., 2005).

Ad 2) Eat fish several times a week

Fish contains healthy fish oils (omega-3 fatty acids) and among others vitamins and minerals, in particular vitamin D, iodine and selenium (Astrup et al., 2005). Danes eat too little fish since only about 15 per cent of the population eat the recommended 30 g. a day, and consumption is decreasing (Fagt et al., 2003). A rise in the consumption of fish implies a decrease in cardiovascular diseases as e.g. ischemic heart disease (Andersen et al., 2003). In relation to obesity fish contains relatively little fat so to substitute a meat meal with a fish meal will improve the general diet. It is recommended to eat a variety of fish types. The recommendation about eating fish is also given in the healthy eating pyramid by Walter Willets (Richelsen et al., 2005). The data on fresh fish in the GfK data are missing from January 2003 and onwards and the data on processed fish from July to December 2002. The purchase of fish is therefore approximated based on estimated intakes for the missing data periods (See Smed, 2008).

Ad 3) Eat potatoes, rice, pasta or brown bread every day

Potatoes, rice, pasta and bread contain carbohydrates, vitamins, minerals and fibres. Danes get around 48 per cent of their energy from carbohydrates. The consumption of bread and cereals is decreasing, while the consumption of potatoes has been stable between 1995 until today (Fagt et al., 2003). According to the official diet recommendations a diet containing 50-60 per cent of total daily energy intake from carbohydrates will protect against cardiovascular diseases by lowering the intake of fat. This is also the reason why a carbohydrate containing diet is said to prevent obesity. Fibres are known to lower the incidence of several types of cancer. The consumption of potatoes, rice, pasta and white bread is controversial since the healthy eating pyramid by Walter Willett recommends a lower intake of “fast” carbohydrates. These are rice, potatoes, pasta and white bread. Both the official recommendations and the healthy eating pyramid recommend the intake of brown bread compared to white bread (and whole grain flour compared to white flour) and there is consensus that fibres are necessary in a healthy diet (Astrup et al., 2005; Richelsen et al., 2005). Since there are different opinions on the consumption of carbohydrates it is chosen only to focus on the consumption of fibres.

Ad 4) Reduce the consumption of sugar – especially from soft drinks, sweets and cakes

Sugar contains energy, but not any vitamins or minerals. The Danes eat too much sugar; children get on average 14 per cent of their energy from sugar, and around 80 per cent of all children get more than the recommended maximum of 10 per cent. The average adult person gets 9 per cent of total energy from sugar, which is below the recommended level, but 40 per cent get more. Most of the sugar comes from soft drinks, sweets, chocolate and cakes. A decrease in the intake of sugar is especially recommended in relation to obesity and diabetes (Mølgaard et al., 2003).

Ad 5) Reduce the consumption of fat – especially saturated fat

Meat and dairy products contain fat, especially saturated fat. The Danes have reduced their intake of fat during the latest years. From 1995 to 2001 the total intake has decreased from 39 per cent of total energy intake to 35 per cent for adults and from 35 per cent to 34 per cent for children. About 15 per cent of total energy intake comes from saturated fat (Astrup et al., 2005). The recommended level of saturated fat is that a maximum of 10 per cent of total energy intake must come from saturated fat. The recommendation for total fat is at least 25 per cent, but no more than 30 per cent (Nordisk Ministerråd, 2004)[22]. The lower bound is to ensure a sufficient intake of essential fatty acids and fat soluble vitamins. A reduction in the intake of fats (and especially saturated fat), which is above the recommended maximum level, can reduce the incidence of cardiovascular diseases. Fat is also important in relation to obesity since fat is very energy dense compared to other nutrients. Also the probability of getting diabetes increases as the intake of especially saturated fat increases. According to the healthy eating pyramid by Willets a liberal consumption of the so-called healthy fats is recommended together with a low consumption of saturated fats. This is not in accordance with the Danish recommendations where there is an upper limit in the total consumption of fats. We choose to include a maximum level of 10 energy per cent for saturated fat and a maximum level of 30 energy per cent for total fat, following the recommendations from the Danish Ministry of Family and Consumer Affairs. As there are very few households going below the lower limit of fat consumption on average the lower level is not taken into account in the HEI.

[22] Some sources recommend 25-35 per cent of daily energy intake from fat instead of a maximum level of 30. We have chosen to focus on a maximum level of 30 per cent.

Appendix 2.B: Food eaten away from home

As the data used for constructing the HEI only cover in-house consumption of foods, it is necessary to state a set of assumptions about the composition of food eaten in canteens, restaurants and fast-food outlets. In general, despite the rising share of food purchased and eaten away from home, most of the food in Denmark is consumed at home and food consumed at different times of the day is almost composed nutritionally equal. The following tables are based on a survey on the dietary habits of the Danish population based on a 7-day dietary registration with fixed answering categories. Therefore, this survey also covers food eaten away from home. The results from this study show that 17 per cent of the respondents eat in canteens (work related) more than 20 times a month, 10 per cent of them 5-20 times a month and 73 per cent less often. 83 per cent, 71 per cent and 75 per cent of woman frequent grill bars, cafeterias and restaurants, respectively, less than once a month, while the same numbers for men are 62 per cent, 60 per cent and 62 per cent (Groth and Fagt, 2003). The largest part of food eaten away from home is therefore consumption in canteens related to work and this only in a limited amount. The distribution of energy intake over the day for various age groups is shown in Table B.1 below. Dinner accounts for a little more of energy intake than the other meals for individuals above the age of 18 and in-between-meals cover a relatively larger shares for younger individuals. But generally, there are only minor differences between the energy shares from different meals for various age groups.

Table B.1: Distribution of energy intake over meals

	Age									
	0-6	7-14	15-18	19-24	25-34	35-44	45-54	55-64	65-74	75-80
Energy										
Breakfast	20	19	16	17	17	18	19	19	20	19
Lunch	22	20	19	20	19	18	19	19	19	18
Dinner	29	32	33	33	39	39	40	39	40	42
In-between	30	29	32	29	26	25	24	21	21	20

Source: Groth et al. (1999).

The following figures are a bit different since they show the energy share from fat over different meals for various age groups, i.e. in Table B.2 an energy share of 32 for breakfast for the 19-24-year-olds means that 32 per cent of the energy consumed for breakfast for this age group comes from fat. There is not much variation in the energy share from fat over the day even though a little more fat is consumed during dinner than the rest of the meals. Looking at the energy share from fat for various age groups shows that the eldest (above 64 yrs) eat a fatty breakfast while younger children 0-6 years of age have lunch and dinner which contain more fat than average.

Table B.2: Percent of total energy intake from fat at various meals

	Age									
	0-6	7-14	15-18	19-24	25-34	35-44	45-54	55-64	65-74	75-80
Fat										
Breakfast	32	30	29	32	35	36	35	35	39	38
Lunch	38	34	34	35	37	37	36	36	37	35
Dinner	41	39	39	39	40	40	39	40	40	39
In-between	30	32	29	28	30	32	31	31	33	32

Source: Groth et al. (1999).

More differences are seen in Tables B.3 and B.4 showing the energy shares from carbohydrates and added sugar. Here the eldest have breakfast and lunch with fewer carbohydrates than the average. The distribution of added sugar is equal across age groups unless for in-between-meals where the young (up to early 30s) get more than 22 per cent of their energy from sugar.

Table B.3: Percent of total energy intake from carbohydrates at various meals

	Age									
	0-6	7-14	15-18	19-24	25-34	35-44	45-54	55-64	65-74	75-80
Carbohydrates										
Breakfast	54	57	56	53	50	49	50	50	45	48
Lunch	49	52	50	49	47	45	45	42	41	40
Dinner	42	43	41	39	37	35	34	35	36	38
In-between	63	60	55	52	52	50	50	51	53	52

Source: Groth et al. (1999).

Table B.4: Percent of total energy intake from added sugar at various meals

	Age									
	0-6	7-14	15-18	19-24	25-34	35-44	45-54	55-64	65-74	75-80
Added sugar										
Breakfast	7	9	8	7	6	7	8	8	8	9
Lunch	6	9	10	9	7	4	4	4	4	5
Dinner	9	9	7	6	4	4	3	4	5	6
In-between	26	27	25	24	22	18	16	17	17	17

Source: Groth et al. (1999).

In this paper it is generally assumed that individuals having lunch outside home in canteens (as compared to bringing a lunch bag) compose their meal nutritionally more or less equal to their consumption at home. This is based on the above tables. Furthermore, it is shown in the literature that 41 per cent of the Danish households eat breakfast together and 81 per cent eat dinner together (Groth et al., 1999). Based on these tables and numbers it is assumed that individuals belonging to the same household compose their meals (in energy shares) more or less equal. A more refined way to cope with the fact that it is individual consumption which counts, but household data that are in hand, would be to use the method developed by Andrew Chesher which uses non-parametric regression to smooth consumption and thereby jump from household data to individual consumption (Chesher, 1999). It is chosen not to use this approach here.

Appendix 2.C: Decomposition of the HEI value

Figures C.1 to C.3 show the development over time in the scores for each of the recommendations for both the balanced and the unbalanced panel[23].

Figure C.1: The development in the scores for fat and fibres

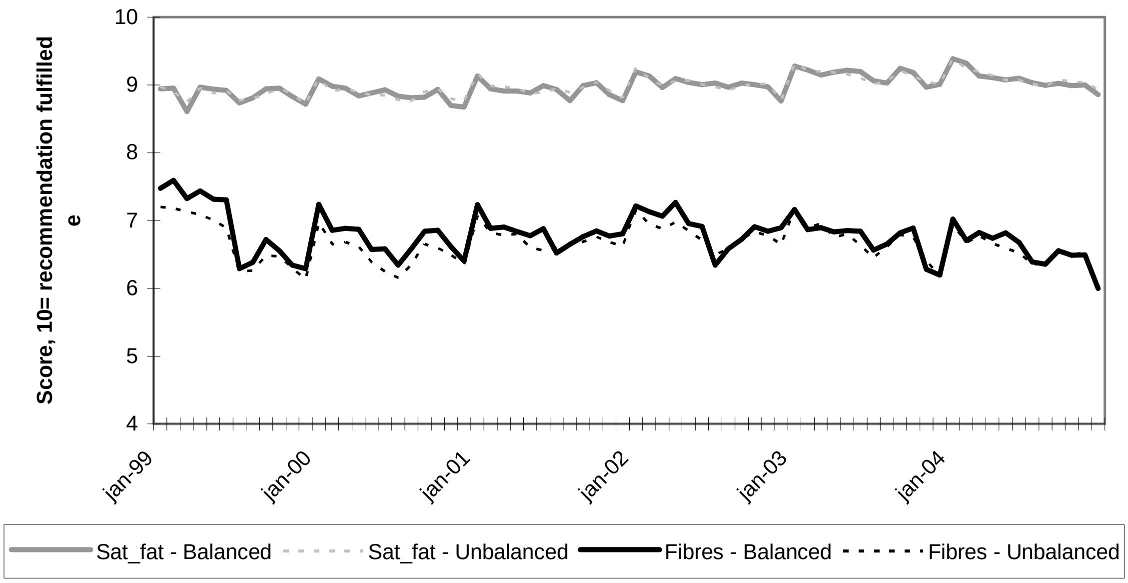
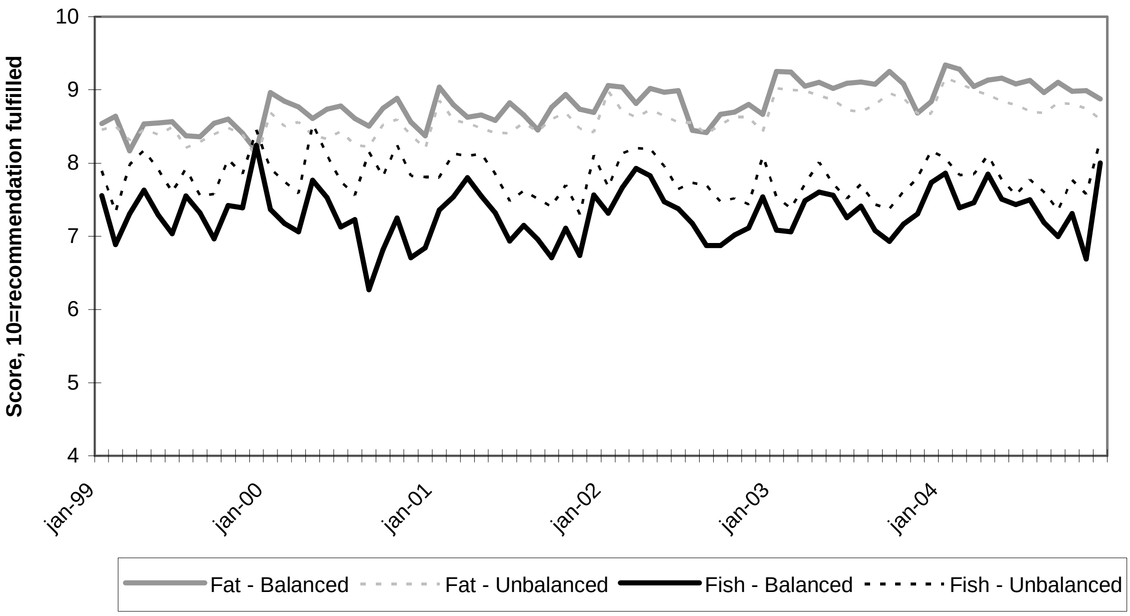
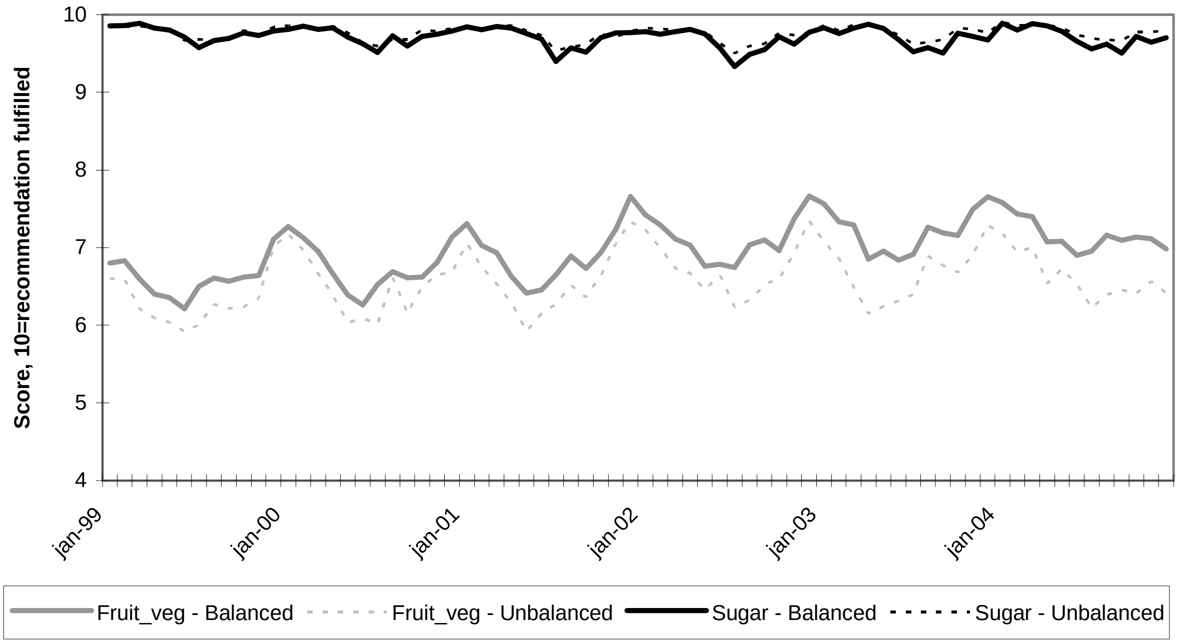


Figure C.2: The development in the scores for saturated fat and fish



[23] The unbalanced panel consists of all households' reporting purchases between January 1999 and December 2004. The balanced panel is the households who participated all six years

Figure C.3: The development in the scores for sugar and fruit and vegetables



Appendix 2.D: Distribution of the scores for each recommendation

The distribution of the average individual scores for saturated fat, fibres and sugar averaged over 1999 compared to 2004 for both the balanced and the unbalanced panel[24].

Figure D.1: The development in the distribution of the score for saturated fat

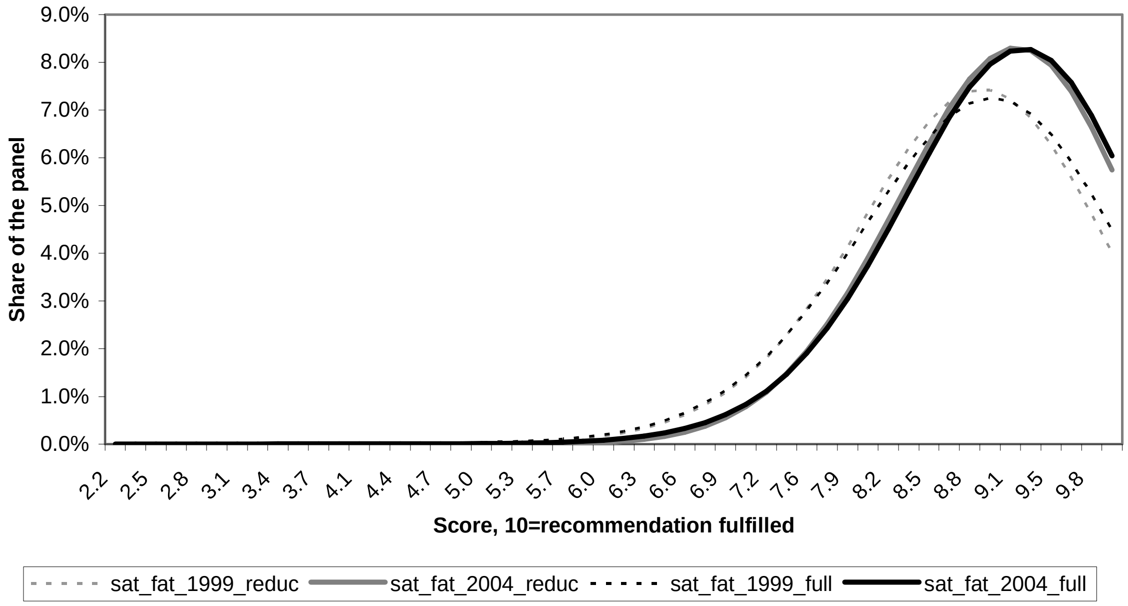
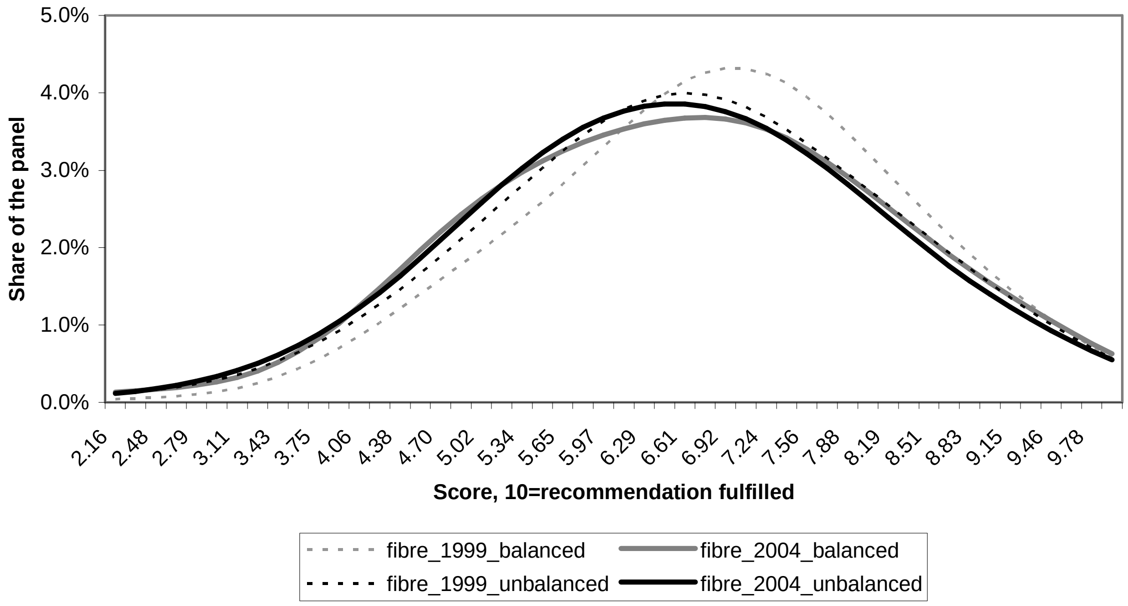
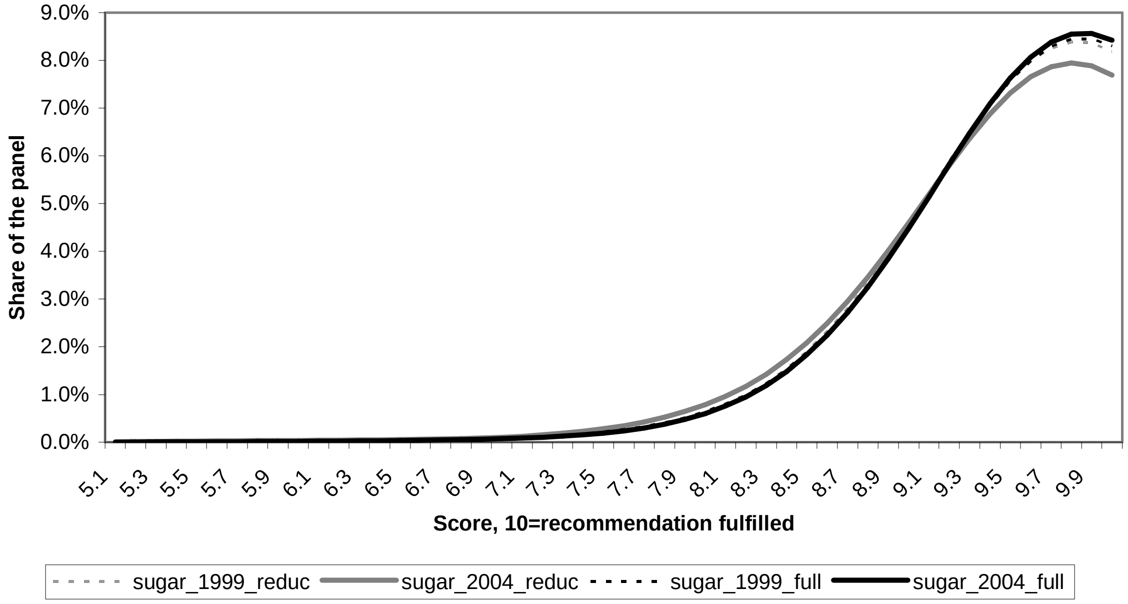


Figure D.2: The development in the distribution of the score for fibre



[24] The unbalanced panel consists of all households' reporting purchases between January 1999 and December 2004. The balanced panel is the households who participated all six years

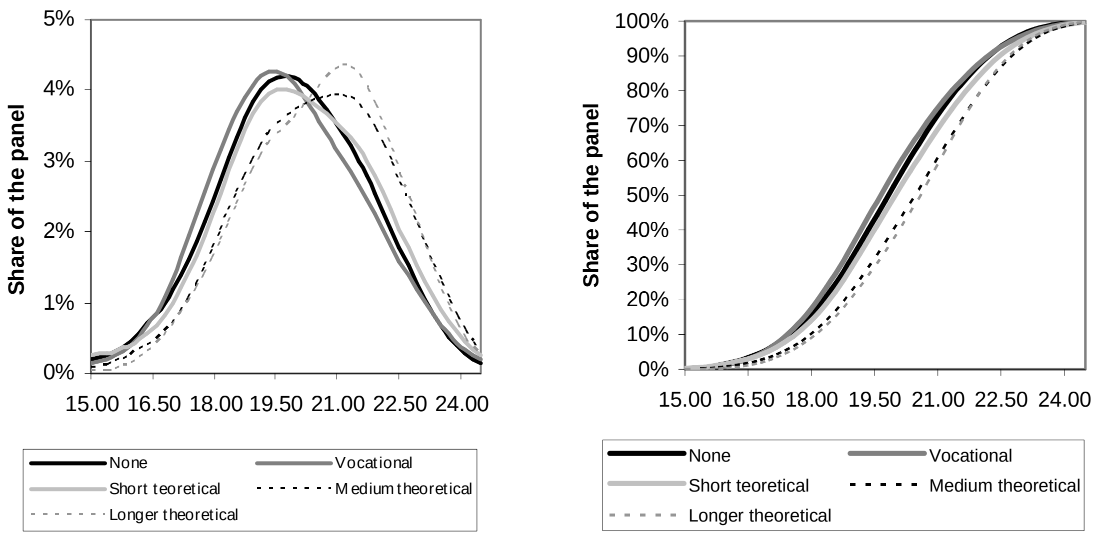
Figure D.3: The development in the distribution of the score for added sugar



Appendix 2.E: Differences in the performance of social and demographic groups

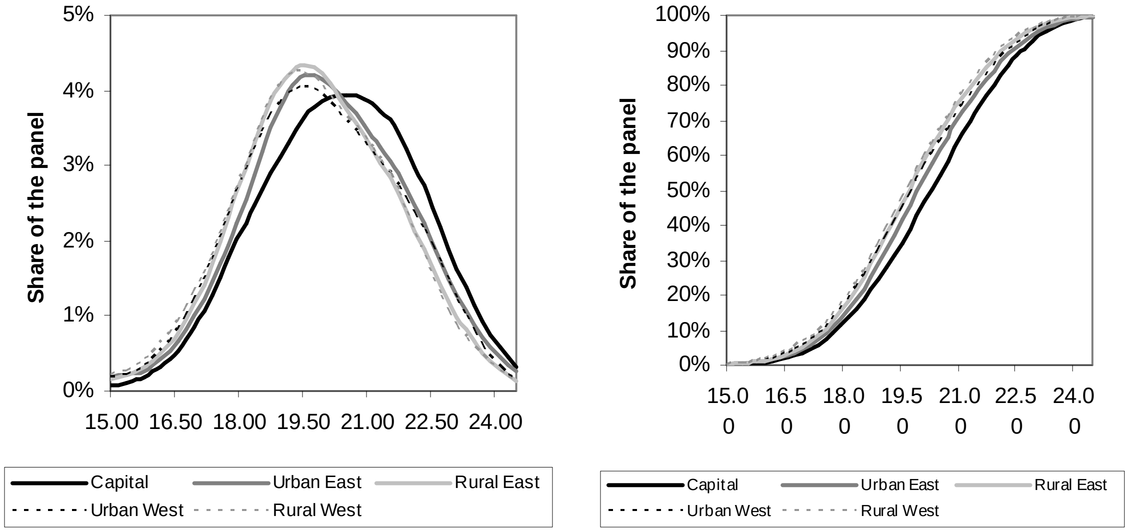
The density of the HEI value averaged over 2004 for various types of households is shown in details in Figures E.1 to E.3. All figures are based on the unbalanced panel. Here we have the average performance over a year for each household meaning that fewer households than shown in Table 2.2 fulfil the recommendations. Starting with Figure E.1 the distribution for educational groups is shown. The distribution for households where the diary keeper has no further education, vocational or shorter theoretical education is pushed towards the left compared to households where the diary keeper has medium or longer theoretical education.

Figure E.1: The density (*left*) and cumulative density (*right*) for the HEI value for different educational groups



The density for households located in different regions of Denmark is shown in Figure E.2. The capital area is pushed towards the right. Furthermore, there is a small tendency that urbanity increases the share of the panel in the better end of the scale. In the lower end a capital, east and west grouping is found.

Figure E.2: The density (*left*) and cumulative density (*right*) for households located in different regions



Finally, the density for various age groups is pictured in Figure E.3. The density for households 60+ and 49-59 years follows each other closely and is pushed towards the right. There is not so much difference in the number of households located in the better end for the remaining households. Households where the head is below 30 years are more likely to be in the lower end. This is to some extent also the case for households 30-39 years old. Households 40-49 years follow their own path.

Figure E.3: The density (*left*) and cumulative density (*right*) for households with different age

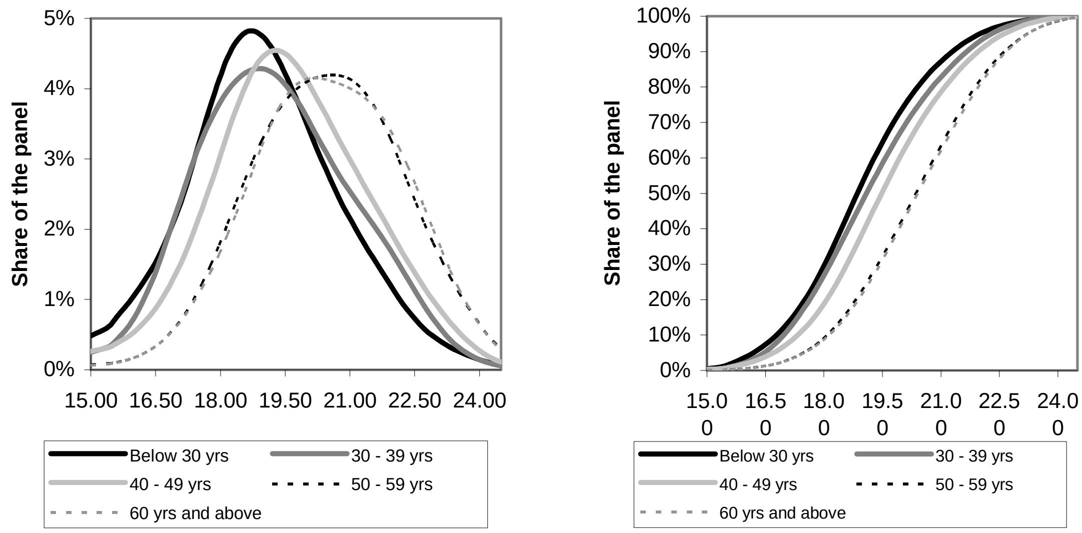


Table E.1: Comparison of means for different groups of consumers

Comparison of young versus old

		HEI value		Fat		Sat_fat		Fruit and veg		Fibre		Fish		Sugar	
		Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value
Capital	Short	-0.99	-27.61	0.16	7.07	0.05	3.04	-1.01	-20.73	-0.03	-1.07	-2.74	-57.23	0.13	11.91
	Long	-0.92	-12.22	0.31	7.14	0.12	4.06	-1.27	-12.13	0.06	0.82	-2.38	-23.36	0.03	1.45
Other regions	Short	-1.24	-16.89	0.08	1.70	0.02	0.70	-1.57	-15.33	-0.08	-1.17	-2.62	-25.56	-0.01	-0.38
	Long	-0.88	-8.80	0.20	3.44	0.03	0.88	-1.06	-7.03	0.15	1.54	-2.51	-17.04	-0.01	-0.27

Comparison of long versus short educated

		HEI value		Fat		Sat_fat		Fruit and veg		Fibre		Fish		Sugar	
		Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value	Diff	t-value
Capital	< 50	-0.68	-10.27	-0.28	-7.86	-0.18	-6.96	-0.93	-10.39	-0.48	-8.31	-0.22	-2.38	0.01	0.45
	> 50	-0.61	-11.89	-0.14	-4.26	-0.11	-5.16	-1.19	-16.54	-0.39	-8.34	0.13	2.15	-0.09	-5.02
Other regions	< 50	-0.83	-8.65	-0.52	-8.77	-0.18	-4.67	-1.08	-7.66	-0.70	-7.42	0.07	0.45	0.02	0.94
	> 50	-0.48	-6.11	-0.40	-8.19	-0.17	-5.51	-0.56	-4.82	-0.47	-6.50	0.18	1.88	0.02	1.01

Chapter 3

Valuation of health

Sinne Smed
AKF
(Danish Institute of Governmental Research)
Copenhagen

Abstract

In this paper consumers' valuation of health and non-nutritional characteristics as e.g. taste and quality are estimated in a hedonic price model. The results show the importance of removing individual heterogeneity and to control for the valuation of non-nutritional characteristics in the estimation. Consumers' valuation of health is estimated in six dimensions each representing one of the official Danish diet recommendations. Consumers are found to value some elements in the health vector positively and others negatively with the overall conclusion that consumers have preferences for energy dense foods. A positive correlation between the valuation of health and the valuation of the non-nutritional characteristics are found indicating that consumers either have a general high valuation of all the characteristics in food or a low valuation. Under certain assumptions, the implicit price of the characteristics estimated in the hedonic model can be interpreted as exogenous. Using this, a positive correlation between healthiness of diets and expenditure is found, suggesting that cost might be a barrier, for some households, to change towards a healthier diet. Cost might especially be a barrier for increasing the consumption of fruit and vegetables

Introduction

The comparison of official diet recommendations with actual Danish dietary practice shows that many people fail to eat according to the recommendations (Smed, 2008) despite well-documented health consequences of not doing so (Astrup et al., 2005). Furthermore, there are large differences in dietary performance between social and demographic groups and observations of dietary changes over time indicate that this bias will increase (Smed, 2008). Lack of knowledge seems not to be the main reason for most households to the absence of healthy dietary patterns since extensive recognition of the dietary health messages has been documented (Pedersen and Kjær, 2007; Holm et al., 2002; Astrup et al. 2005; Struktur-direktoratet, 2000). Since the recommendations are known, at least passively, it is important to understand if there are other possible barriers to improve diets. The relative prices between healthy and unhealthy foods or a restrictive budget constraint are some suggested barriers to a change towards a more healthy diet (see e.g. Smed, 2007 et al.; Darmon et al., 2002; Lennernas et al., 1997; Adelaja et al., 1997). Another suggested explanation is that consumers simply prefer energy dense foods due to, that in ancient times, when food environments were uncertain, the best surviving strategy was to build up bodily fat in order to survive the next period of food scarcity (Drewnowski, 1995; Birch, 1999). Today, food in highly industrialised economies is no longer scarce and the production is highly efficient leading to lower food prices and high availability of processed and pre-prepared foods. According to this explanation, overeating is just a manifestation of the fundamental mismatch between modern environments and ancient times, in which preferences for eating evolved (Smith, 2002a, 2002b; Yanovski, 2003). This means that consumers in modern societies often face a trade-off between preferences for taste and convenience on the one hand and health on the other.

The objective of this paper is to estimate consumers’ valuation of health characteristics and valuation of non-nutritional characteristics (e.g. taste) in food. To find the values consumers attach to food characteristics a hedonic price model is estimated, where six different health characteristics and three non-nutritional characteristics are treated as inseparable parts of an entity, the total diet. Furthermore, under certain assumptions the implicit price of the characteristics estimated in the hedonic model can be interpreted as exogenous. Under this assumption the hedonic prices can be used to identify if a healthier diet costs more and thereby whether budget and prices might be a barrier towards healthy eating. The dataset used for the analysis in this paper is unique since it combines nutrition information with prices and expenditures at household level and follows the same households over a long period of time.

The contributions of this paper are therefore twofold; first, due to richness of our data it is possible to control for individual heterogeneity and the influence of taste for non-nutritional characteristics in the estimation of the valuation of health and secondly, it is possible to follow the valuation of the health characteristics over different types of consumers.

The remainder of the paper is organised as follows: Section 3.1 describes the theory of hedonic pricing. In Section 3.2 we compare the results from a model based on panel data with an existing model based on cross-section data in order to infer the importance of removing individual heterogeneity. This implies that the estimation in this section is based on a fairly restrictive version of the final model. In the next sections some of these restrictions are relaxed one by one. Section 3.3 relaxes the assumption of linear marginal utility of characteristics and the costs of a healthy diet are calculated for the average consumer. Section 3.4 relaxes the assumption of equal valuation of health and non-nutritional characteristics over households and the valuation of health over different types of consumers are compared. Section 3.5 contains a discussions and conclusion of the paper.

3.1 The theory of hedonic price models

The following section explains the theory of hedonic pricing, which is used to determine consumers' valuation (implicit prices) of health characteristics in food. The theory of hedonic pricing has mostly been applied to durable goods, but is recently applied to aggregate categories of foods (e.g. Lenz et al., 1994) and individual food categories (e.g. Shi and Price, 1998). Recently, Ranney and McNamara (2002) infer the implicit market valuation of dietary quality from the total expenditure on food. The implicit values of the health characteristics (or implicit prices) are estimated holding constant the other factors that affect food expenditures. The approach here follows Ladd and Zober (1977), Lenz et al. (1994), Shi and Price (1998) and Ranney and McNamara (2002). Like these papers we consider only demand-side interactions (contrary to Rosen, 1974), since the individual consumer is assumed not to be able to affect the price, nor the health characteristics of each food.[25] A consumer purchases a vector of J (running index j) foods. Let q denote the vector of purchased food quantities. A household is assumed to derive utility from health, h, inherent in this food, the non-nutritional characteristics of food, g, and expenditure on non-food goods, y. In the following derivation

[25] On a long-term basis consumers might be able to influence the health characteristics of food due to the introduction of new products caused by demand effects (as e.g. low fat dairy products), but this effect is assumed to be of minor importance in the time span considered here.

of the hedonic price equation the subscript of household *n* is suppressed due to ease of notation.

$$U = u\left(g, h, y \middle| \Omega\right) \tag{3.1}$$

Ω is a vector of social and demographic variables characterising each household and $y = X_{tot} - pq$ where X_{tot} is total expenditure. *g* is a vector of *M* (running index *m*) non-nutritional characteristics $g = \left(g_1, ... g_M\right)$, which is a function of the actual amount of foods purchased *q*.

$$g = g\left(q\right) \tag{3.2}$$

The health associated with the purchase of foods is a function of the vector $z = \left(z_1, ..., z_i\right)$ describing the content of each of the *I* (running index *i*) nutrients in food. This is a function of the actual amount of foods purchased, *q*.

$$h = h\left(z\left(q\right)\right) \tag{3.3}$$

The connection between the amount of nutrients and the food purchased is assumed to be described through a linear function, the technology matrix *A*.

$$A \equiv goods \begin{cases} 1 \\ \vdots \\ j \\ \vdots \\ J \end{cases} \overset{\overbrace{\begin{matrix} 1 & \ldots & i & \ldots & I \end{matrix}}^{Health-characteristics}}{\begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1I} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{j1} & \cdots & a_{ji} & \cdots & a_{jI}! \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{J1} & \cdots & a_{Ji} & \cdots & a_{JI} \end{pmatrix}} \tag{3.4}$$

We assume constant return to scale and that characteristics is additive over goods i.e. $z = A'q$. This implies that total amount of fat consumed is the sum of the contribution of fat from each of the foods consumed and that the amount of fat consumed from e.g. two litres of milk is the double of the amount of fat consumed from one litre of milk. The *A* matrix is assumed constant over consumers, to have full rank, and to have at least as many goods as characteristics. Together with health the consumer purchases non-nutritional characteristics (like taste). The connection between the amount of non-nutritional characteristics and food purchased is assumed to be described through a matrix *B*, $g = B'q$, also assumed to be constant over consumers, to have full rank and have at least as many goods as characteristics.

$$B \equiv goods \begin{cases} 1 \\ \vdots \\ j \\ \vdots \\ J \end{cases} \overbrace{\begin{matrix} 1 & \dots & m & \dots & M \end{matrix}}^{non-nutritional-characteristics} \begin{pmatrix} b_{11} & \cdots & b_{1m} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots & \cdot^{\cdot^{\cdot}} & \vdots \\ b_{j1} & \cdots & b_{jm} & \cdots & b_{jM}! \\ \vdots & \cdot^{\cdot^{\cdot}} & \vdots & \ddots & \vdots \\ b_{J1} & \cdots & b_{Jm} & \cdots & b_{JM} \end{pmatrix} \tag{3.5}$$

There is substantial empirical evidence that preferences for food can be viewed as separable from preferences from non-food items (e.g. deJanvry, 1966). This assumption is maintained throughout this paper. Furthermore, we assume that the utility of consumption of non-nutritional characteristics of food is additively separable from the utility of the consumption of the health characteristics, which is a condition for the possibility to isolate the two effects from each other. This implies that (3.1) reduces to the following utility function:

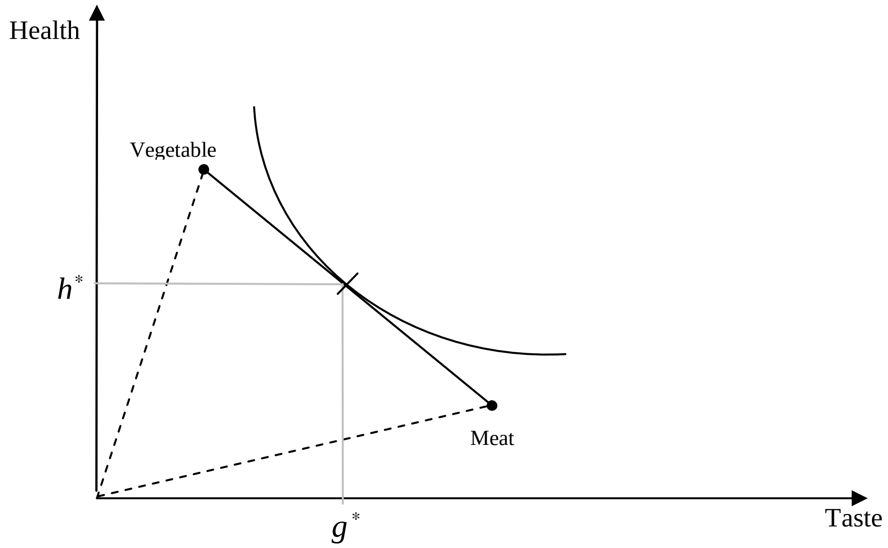
$$U = u_g(g|\Omega) + u_h(h|\Omega) \tag{3.6}$$

If (3.2) and (3.3) is inserted in (3.6), utility is defined as:

$$U = u_g(g(q)|\Omega) + u_h(h(z(q))|\Omega) \tag{3.7}$$

As an example, imagine a world of two goods; say vegetables and meat, and two characteristics; “taste” and “health”. We assume vegetables are high in health and low in “taste” and that the opposite applies to meat. A consumer, who uses his whole food budget to buy only vegetables, will get a high amount of “health” and a low amount of “taste” as illustrated in Figure 3.1. If he uses the whole budget to buy only meat he will get a low amount of “health” and a high amount of “taste”. The possible bundles of characteristics to purchase within the same budget are characterised by the straight line connecting the endpoints for the vegetable line and the meat line in Figure 3.1. When the consumer maximise his utility function subject to a budget constraint, $p'q \leq X$, where *p* is prices of goods and *X* is the budget for foods, he gets the optimal amounts of “health”, h^* , and “taste”, g^* .

Figure 3. 1: Example of preferences and the budget constraint over characteristics



If consumers maximise the utility function (3.7) in a world of *J* goods and *I+M* characteristics the *J* first order conditions are:

$$\frac{\partial u}{\partial g}(g(q))\frac{\partial g}{\partial q_j}+\frac{\partial u}{\partial h}(h(z(q)))\frac{\partial h}{\partial z}(z(q))\frac{\partial z}{\partial q_j}-\frac{\partial u}{\partial X}p_j=0 \qquad (3.8)$$

$$\Leftrightarrow p_j=\frac{\frac{\partial u}{\partial g}(\cdot)}{\frac{\partial u}{\partial X}(\cdot)}\cdot\frac{\partial g}{\partial q_j}(\cdot)+\frac{\frac{\partial u}{\partial h}(\cdot)\frac{\partial h}{\partial z}(\cdot)}{\frac{\partial u}{\partial X}(\cdot)}\cdot\frac{\partial z}{\partial q_j}(\cdot) \qquad (3.9)$$

The derivatives $\frac{\partial z}{\partial q_j}$ and $\frac{\partial g}{\partial q_j}$ (i.e. the amount of health and of non-nutritional characteristics in one unit of food *j*) are described through the elements in the technology matrices *A* and *B*, a_{ji} and b_{jm}. The marginal utility of the budget is defined as: $\frac{\partial u}{\partial X}$. Following the traditional theory of hedonic pricing and holding constant the marginal utility of the budget over time and households (see e.g. Lenz et al., 1994; Ladd and Zober, 1977; Ladd and Suvannunt, 1976) the implicit prices of the health characteristics, π_i, can be defined as: $\frac{\frac{\partial u}{\partial h}(\cdot)\frac{\partial h}{\partial z_i}(\cdot)}{\frac{\partial u}{\partial X}}=MRS$ (marginal rate of substitution) between expenditure and the health characteristics. The implicit price of the non-nutritional characteristics (e.g. taste) of food consumption, τ_m, is defined as:

$$\frac{\frac{\partial u}{\partial g_m}(\cdot)}{\frac{\partial u}{\partial X}} = MRS \text{ between non-nutritional characteristics and expenditure.}$$

The assumption of keeping the marginal utility of the budget constant is restrictive, but perhaps not totally unrealistic when applied on a necessity like food (the budget share used on food is slightly falling in Denmark when income is increasing) compared to using the assumption in relation to luxury goods or in another country where a larger share of the budget is used on foods.

Due to (3.9) the price of food *j* can be divided into a price paid for the non-nutritional characteristics inherent in that food and a price paid for the health characteristics. If there are no restrictions on functional form the implicit price of the health characteristics and the implicit price of the non-nutritional characteristics will both be functions of the amount of characteristics purchased and depend on social and demographic variables as in (3.10) below (the equation is here expressed by subscripts for household *n* in order to specify which variables are dependent on individual household variables).

$$p_{njt} = \sum_{m=1}^{M} \tau_{nmt}\left(g(q)|\Omega_n\right)\cdot b_{jm} + \sum_{i=1}^{I} \tau_{nit}\left(h(z(q))\right)\cdot a_{ji} \quad (3.10)$$

If the *J* first order conditions (3.10) are inserted in the budget constraint (the final first order condition $X_{nt} = \sum_{j=1}^{J} p_{ntj} \cdot q_{ntj}$) we get the following equation:

$$X_{nt} = \sum_{j=1}^{J}\left(\sum_{m=1}^{M} \tau_{nmt}\left(\cdot|\Omega_n\right)\cdot b_{jm} + \sum_{i=1}^{I} \pi_{nit}\left(\cdot|\Omega_n\right)\cdot a_{ji}\right) q_{njt} \quad (3.11)$$

⇕

$$X_{nt} = \sum_{j=1}^{J}\sum_{m=1}^{M} \tau_{nmt}\left(\cdot|\Omega_n\right)\cdot b_{jm}\cdot q_{njt} + \sum_{j=1}^{J}\sum_{i=1}^{I} \pi_{nit}\left(\cdot|\Omega_n\right)\cdot a_{ji}\cdot q_{njt} \quad (3.12)$$

Since $\sum_{j=1}^{J} b_{jm}\cdot q_{njt} = g_{nmt}$ and $\sum_{J=1}^{J} a_{ji}\cdot q_{njt} = z_{nit}$ due to the assumption of additivity of nutrients across goods (3.12) reduces to:

$$X_{nt} = \sum_{j=1}^{J}\sum_{m=1}^{M}\tau_{nmt}\left(\cdot\left|\Omega_n\right.\right)\cdot g_{nmt} + \sum_{j=1}^{J}\sum_{i=1}^{I}\pi_{nit}\left(\cdot\left|\Omega_n\right.\right)\cdot z_{nit} \tag{3.13}$$

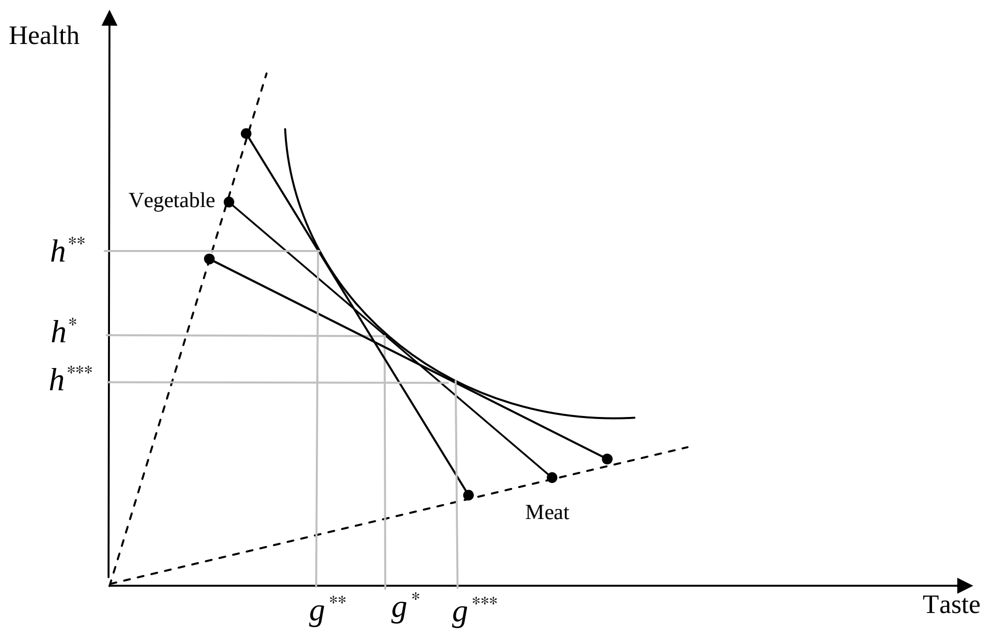
This means that total outlay on foods for households *n* at time *t* is a function of total outlay on non-nutritional characteristics (the implicit prices of non-nutritional characteristics multiplied with the amounts of each of the non-nutritional characteristics bought) and total outlay on health characteristics (the implicit prices of the health characteristics multiplied by the amount of each health characteristic).

In the two goods two characteristics setting this is equal to:

$$\begin{gathered} X_{nt} = \left(\tau_{n,taste}\cdot b_{1m} + \pi_{n,health}\cdot a_{1i}\right)\cdot q_{n1t} + \left(\tau_{n,taste}\cdot b_{2m} + \pi_{n,health}\cdot a_{2i}\right)\cdot q_{n2t} \\ \Updownarrow \\ X_{nt} = \left(\tau_{n,taste}\cdot q_{nt}\right) + \left(\pi_{n,health}\cdot z_{nt}\right) \end{gathered} \tag{3.14}$$

Figure 3.2 shows a geometric illustration of equation (3.14). Changed prices of vegetables and meat, everything else equal, will lead to a changed slope on the budget restriction and thereby changed realisations of the optimal amount of health h^*, h^{**}, h^{***} and taste g^*, g^{**}, g^{***}. With enough realisations this will reveal the parameters of the utility function.

Figure 3. 2: Illustration of the effect of changed food prices on optimal consumption



3.2 Comparison of a panel-data model with a cross-section model

Estimations of hedonic price models are commonly based on cross-section data, while the data used in this paper have the feature of being panel data. This allows for taking the issue of individual heterogeneity into account. In this section a fairly restrictive version of the final model is estimated in order to make the model comparable to the model developed by Ranney and McNamara (2002) based on cross-sectional data. These restrictions are relaxed one by one in the following sections. In this section we assume that *a)* the valuation of the health characteristics is the same across consumer groups i.e. $\pi_{1it} = \pi_{2it} = ... = \pi_{Nit}$, *b)* the valuation of health is stable over time i.e. $\pi_{ni1} = \pi_{ni2} = ... = \pi_{niT}$, and finally *c)* the valuation is independent of the level of consumption. Thereby (3.13) reduces to:

$$X_{nt} = \sum_{m=1}^{M} \tau_m(\Omega_n) \cdot g_{nmt} + \sum_{i=1}^{I} \pi_i \cdot z_{nit} \qquad (3.15)$$

Furthermore, following Ranney and McNamara (2002) the part of expenditure used on non-nutritional factors is assumed to be constant over time and dependent on social and demographic variables, i.e. $\sum_{m=1}^{M} \tau_n(\Omega_n) \cdot g_{nmt} = \Omega_n$. This is a rather restrictive assumption, but the reason to follow this approach is to make it possible to test our extended model with an existing model. The model to be estimated is then:

$$X_{nt} = \Omega_n + \sum_{i=1}^{I} \pi_i \cdot z_{nit} + \varepsilon_{nt} \qquad (3.16)$$

This version of the model is equivalent to assume linearity in both the non-nutritional characteristics and in the health characteristics (assuming that the implicit price of health and non-nutritional characteristics are independent of the level of consumption). Here this also implies homothetic preferences. A problem in this specification is individual heterogeneity. If the social and demographic variables do not cover all individual heterogeneity equation (3.16) will look like (3.17). This will result in biased parameters since unobserved non-nutritional characteristics then will be correlated with the choice of health characteristics as pointed out by Bartik (1987).

$$X_{nt} = \Omega_n + \sum_{i=1}^{I} \pi_i \cdot z_{nit} + \eta_n + \varepsilon_{nt} \qquad (3.17)$$

To solve this problem the function can be estimated in first differences. To the extent that the social and demographic variables are assumed to be constant over time the equation will reduce to:

$$\Delta X_{mt} = \sum_{i}^{I} \pi_{it} \cdot \Delta z_{mjt} + \Delta \varepsilon_{it} \qquad (3.18)$$

When estimating (3.16) and (3.18) total expenditure is inserted as expenditure per person[26] per month in order to make households of different sizes comparable. The amount of health purchased is approximated through the composition of the diet i.e. the energy shares for total fat, added sugar, fruit and vegetables, fish, fibre and saturated fat are inserted on the right hand side. Each of these foods and nutrients are assumed important for a healthy composition of diets and have a recommended maximum or minimum consumption due to the official Danish diet recommendations[27]. The model is estimated in levels and in differences and the results are compared. This part of the model is only estimated for 2004 since this is the only year with data for Body Mass Index (BMI)[28] and questions concerning attitude towards own weight, exercise and dietary practice. Most of the social and demographic variables are inserted as dummies except age, BMI and number of children, which are treated as continuous variables. The variables included in the final model are shown in Table 3.1. Insignificant parameters are removed.

Table 3. 1 Variables included in the estimation

Variable	**Basis**	**Label**	
Sex of main buyer	Woman	Man or both	
Further education of diary keeper	No further education or vocational oriented	Short theoretical	Medium or long theoretical
Geographical location	Rural	Urban	Capital
Number of persons in the family	Single	Couple	
Have you changed your diet to lose weight	No	Yes	
No of children 0-6 years	Continuous		
No of children 7-20 years	Continuous		
Age of main buyer	Continuous		
BMI (only adults)	Continuous		

[26] This variable is calculated using the OECD consumption unit equivalence scale. On this scale, the first adult is given a weight of 1.0 and all other adults a weight of 0.7. Children get a weight of 0.5 (for details, see Atkinson, 1995, pp. 80-81).

[27] In the final versions of the model these are translated into a direct measure of health through a score calculating how close the individual households are from fulfiling the dietary recommendations. See below.

[28] BMI is calculated from questions on height and weight and is defined as: $BMI = \frac{weight(kg)}{height(m) * height(m)}$.

The estimation results are shown in Table 3.2.

Table 3. 2: Parameter values from the estimation of the hedonic price equation

	Model in level, year 2004 (equation 3.16)			*Model in differences, year 2004* (equation 3.18)		
Parameter	Estimate	Std err.	Pr>|t|	Estimate	Std err.	Pr>|t|
Constant	51.63	24.32	0.0338			
Age of main buyer	13.46	0.72	<.0001			
Age of main buyer squared	-0.11	0.01	<.0001			
Main buyer short theoretical education	8.21	4.24	0.053			
Main buyer medium or long theoretical education	49.41	3.59	<.0001			
Couple	-205.69	3.39	<.0001			
Male or both is main buyer	14.61	3.25	<.0001			
Capital area household	60.57	4.04	<.0001			
Urban household	3.62	3.31	0.2734			
No of children below 7 in the household	-75.36	6.10	<.0001			
Children between 7 and 20 in the household	-98.63	5.29	<.0001			
Main buyer is dieting	19.22	3.02	<.0001			
BMI of main buyer	1.28	0.31	<.0001			
Fat	-112.21	26.10	<.0001	-3.93	21.82	0.8572
Fibres	-5741.77	383.60	<.0001	-4380.60	356.60	<.0001
Fish	2173.11	87.13	<.0001	1649.22	66.79	<.0001
Sat_fat	795.40	51.26	<.0001	527.09	42.93	<.0001
Sugar	121.28	24.84	<.0001	142.07	20.93	<.0001
Fruit_Veg	4464.41	49.34	<.0001	4381.85	51.10	<.0001
	R^2=0.4728			R^2=0.3171		

The results show that food expenditures depend on social and demographic variables. Expenditure on food increases with age until the age of 61 and then decreases. Households with a longer or medium length theoretical education spend 49.41 DKK more per person per month on food than a household with no further education. Other types of education have no significantly effect on expenditure. Capital location increases expenditure by 60.57 DKK per person per month compared to other parts of the country. Couples and households with young children use 205.69 and 75.36 DKK less per month per person, respectively, which we assume are due to economies of scale in the household. Those who have taken action to reduce their weight also use more money on food (19.22 DKK per person per month). Expenditures are increasing with BMI (1.28 DKK per BMI unit). A household where the male is main responsible for shopping or where both shop equally often also spends a little more on food than households where the female is the main shopper. The question is whether these social and demographic variables cover all individual heterogeneity. Equation (3.18) is estimated in differences, differentiating out all individual heterogeneity, leaving only the implicit prices for fulfilling the dietary recommendations to be estimated. Comparing these implicit prices estimated in levels with the same estimated in differences show parameters of equal signs, but different magnitudes. Consumers value fish, saturated fat, sugar and fruit and vegetables in the diet positively, while the fat and fibres are valued negatively in both models (the parameter for fat is insignificant in the model in differences). A Wu-Hausman test for exogeneity gives the value 369 with 6 degrees of freedom proposing that the null hypotheses of no correlation between the unobserved non-nutritional characteristics and the health characteristics are rejected (this means that the social and demographic variables do not cover all individual heterogeneity) and the model in differences is the most convenient model of the two suggested. This emphasises the possible risk of getting biased parameters in the estimation due to unobserved heterogeneity. The model in this section was based on rather restrictive assumptions. In the next sections we use the model in differences and relax some of these assumptions.

3.3 The cost of a healthy diet, the average consumer

In this section we release the assumptions that the valuation of characteristics is independent of the level of consumption, and that the amount used on non-nutritional characteristics is the same over time. The first part of this section describes the changed assumptions and empirical considerations concerning the estimation of a less restrictive version of (3.13) while the last part is devoted to the results. The assumptions of equal valuation of the health characteristics and the non-nutritional characteristics across consumer groups are retained in this section (i.e. $\pi_{1it} = \pi_{2it} = ... = \pi_{Nit}$) and that the valuations are stable over time (i.e. $\pi_{ni1} = \pi_{ni2} = ... = \pi_{niT}$). But the valuations are allowed to depend on the level of consumption thereby (3.13) is expressed as:

$$X_{nt} = \sum_{m=1}^{M} \tau_m \left(g_m \right) \cdot g_{nmt} + \sum_{i=1}^{I} \pi_i \left(z_i \right) \cdot z_{nit} \tag{3.19}$$

Again total expenditure for each household on the left hand side is inserted as expenditure per person[29] per month in order to make households of different sizes comparable. The health

[29] This variable is calculated using the OECD consumption unit equivalence scale. On this scale, the first adult is given a weight of 1.0 and all other adults a weight of 0.7. Children get a weight of 0.5 (for details, see Atkinson, 1995, pp. 80-1).

characteristics, z_i, i.e. energy shares for total fat, added sugar, fruit and vegetables, fish, fibre and saturated fat are inserted on the right hand side. In the final results the energy shares are recalculated to represent health through a health score measuring how close the households are from fulfilling the dietary recommendations, which is easier to interpretable in terms of health (See below). Furthermore we assume that the part of total food expenditure used on non-nutritional characteristics consists of a part which is household specific and another part which depends on the level of non-nutritional characteristics consumed:

$$X_{non\_nut,nt} = \sum_{m=1}^{M} \tau_m (g_m) \cdot g_{nmt} + \eta_n \tag{3.20}$$

The household-specific part might be due to that some households like buying foods in more expensive shops than other households or that some households prefer foods of a higher non-observable quality, but with the same amount of non-nutritional characteristics and the same health profile as other households. The valuation of the health characteristics are assumed to depend on the level of consumption following a quadratic form:

$$X_{nt} = \sum_{m=1}^{M} \left(\tau_m \cdot g_{nmt} - \hat{\tau}_m \cdot (g_{nmt})^2 \right) + \sum_{i=1}^{I} \left(\pi_i \cdot z_{nit} - \hat{\pi}_i \cdot (z_{nit})^2 \right) + \eta_n + \varepsilon_{nt} \tag{3.21}$$

The equation is estimated in differences:

$$\Delta X_{nt} = \sum_{m=1}^{M} \left(\tau_m \cdot \Delta g_{nmt} - \hat{\tau}_m \cdot \Delta (g_{nmt})^2 \right) + \sum_{i=1}^{I} \left(\pi_i \cdot \Delta z_{nit} - \hat{\pi}_i \cdot \Delta (z_{nit})^2 \right) + \Delta \varepsilon_{nt} \tag{3.22}$$

Furthermore, dummies are inserted for the months of January and December.[30] Three types of non-nutritional characteristics are chosen. These are characteristics intrinsic to meat and fish, to carbohydrate containing foods and to dairy. One unit (e.g. kg.) of a good belonging to the dairy group will give one unit of the non-nutritional characteristic "dairy" and so on. The content of each of these groups is shown in Appendix 3A. These broad groups are chosen in order to secure independence between the nutritional characteristics and the non-nutritional characteristics. [31] Independence means that the households can both increase and decrease the healthiness of their diet maintaining the same consumption of non-nutritional characteristics.

[30] In the original estimations dummies were inserted for each of the 12 months, but the dummies are insignficant for the other months.

[31] In initial estimations there were 9 groups representing the non-nutritional characteristics. There are no major differences in the valuation of the health characteristics between the models with 9 or 3 non-nutritional characteristics, while both models deviate from the model where the non-nutritional characteristics are not

To infer total utility we use that expenditure is equal to the implicit price (marginal utility) of each characteristic multiplied by the amount of good bought. The values in the parentheses are equal to the marginal utility.

$$X_{nt} = \sum_{m=1}^{M} \left(\tau_m - \hat{\tau}_m \cdot g_{nmt}^{*}\right) \cdot g_{nmt}^{*} + \sum_{i=1}^{I} \left(\pi_i - \hat{\pi}_i \cdot z_{nit}^{*}\right) \cdot z_{nit}^{*} \qquad (3.23)$$

From these marginal utilities it is possible to calculate total utility from each of the characteristics as:

$$U_i = \int_0^{z_i^*} \pi_i - \hat{\pi}_i \cdot z_{nit} \qquad (3.24)$$

Since one DKK spent on food will give you varying amounts of nutrients, dependent on which type of food you choose to buy, the budget constraint in characteristics space is generally nonlinear. This leads to endogenous prices. However, at the optimal point where the indifference curve is a tangent to the budget constraint the separating hyper-plane between these two loci is linear. In this optimal point and under the assumption of constant return to scale, prices can be assumed to be exogenous (Deaton and Muellbauer, 1980). We have constant return to scale due to the linear technology, and the first order equations used in the derivation of the hedonic price equation are a result of consumers optimising their utility. Hence we can infer prices to be exogenous everything else equal. This interpretation of the implicit price is used in the following to calculate expenditure on health for the individual household:

$$X_{nt,health} = \sum_{i=1}^{I} \left(\pi_i - \hat{\pi}_1 \cdot z_{nit}^{*}\right) \cdot z_{nit}^{*} \qquad (3.25)$$

Consumer surplus for health can then be calculated as the difference between total utility and total cost. To represent a more direct measure for health the energy shares can be translated into a health score measuring how close the households are from fulfilling each of the official Danish dietary recommendations.[32] These scores take values between 0 and 10, with 10 representing that the recommendation is fulfilled and 0 representing the longest distance possible from the recommendation.

explicitly modelled. Therefore, we choose the model with only 3 non-nutritional characteristics to ensure independence. Appendix 3B shows the difference in the estimated health values for models with a different number of non-nutritional characteristics.

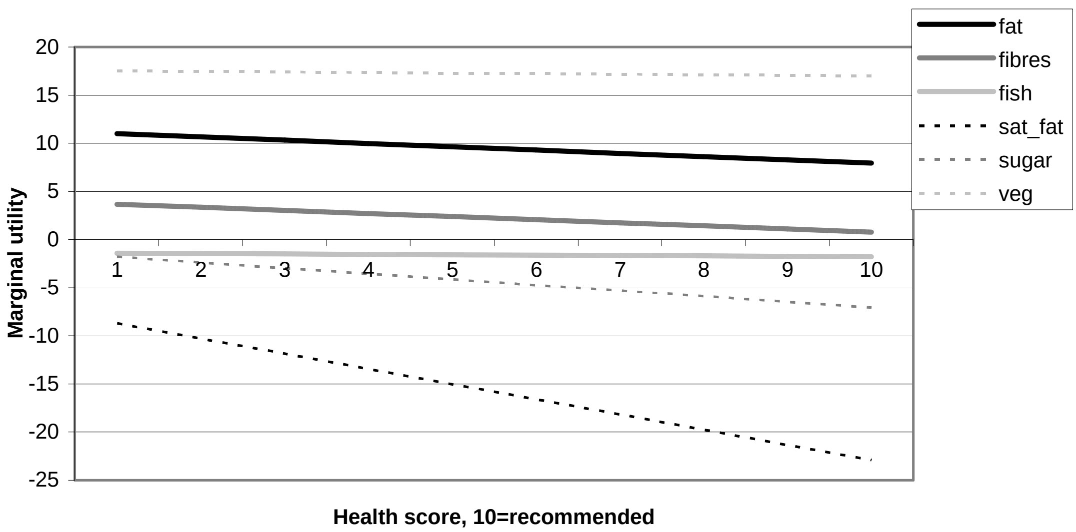
[32] For a detailed description of the diet recommendations and the health scores see Smed (2008)

$$Score = 10 \cdot \left(\frac{s_i - s_{worst}}{s_{recomended} - s_{worst}} \right) \Leftrightarrow s_i = \frac{1}{10}\left(s_{recommended} - s_{worst}\right) \cdot score + s_{worst} \qquad (3.26)$$

Where s_i is actual energy share and $s_{recommended}$ and s_{worst} is the recommended and worst possible energy share, respectively. In each of the measures above (3.26) can be inserted and marginal utility, total cost etc. can be represented via the health score instead of the energy shares. [33]

Figure 3.3 shows the calculated marginal utility of health represented via the health scores. All marginal utilities are decreasing with quantity (the quantity of health increases as the consumers are closer to fulfil the dietary recommendations). The marginal utility of the health scores is positive for fat, fruit and vegetables and fibres and negative for fish, sugar and saturated fat. A negative marginal utility translates into that the consumer values health from this particular recommendation negatively, e.g. they do not want to cut down on the consumption of sugar or saturated fat or increase the consumption of fish as recommended.

Figure 3. 3: Marginal utility for each of the dietary recommendations, the average consumer



[33] A Healthy Eating Index (HEI) has also been used to represent the health characteristics in the initial estimations without changing the main conclusions. The HEI value is a measure telling how close the households are from fulfilling the dietary recommendations for added sugar, total fat, fibres, fruit and vegetables, fish and saturated fat in one aggregate measure. An HEI value of 24.5 represents a household fulfilling all the dietary recommendations, while an HEI value of 0 represents a household with the worst possible diet. A value of e.g. 20 can be obtained by different combinations of proximity to each of the 6 recommendations and are therefore not unique. See Smed (2008) for a more detailed description of the HEI measure. We choose to use the individual health scores since they contain more information.

Figure 3.4 illustrates total cost for each of the recommendations. As it is illustrated total cost declines for some of the recommendations as the recommendation is approached. This is the case for e.g. saturated fat, fish and sugar, meaning that consumers save money as they approach the recommendation. So even though consumers have a negative marginal utility (total utility decreases as the recommendation is approached), the savings in total cost might be even larger, which results in a positive consumer's surplus. As an illustration the total cost for the health score for saturated fat is shown together with total utility and consumer surplus in Figure 3.5.

Figure 3. 4: Total cost for each of the dietary recommendations, the average consumer

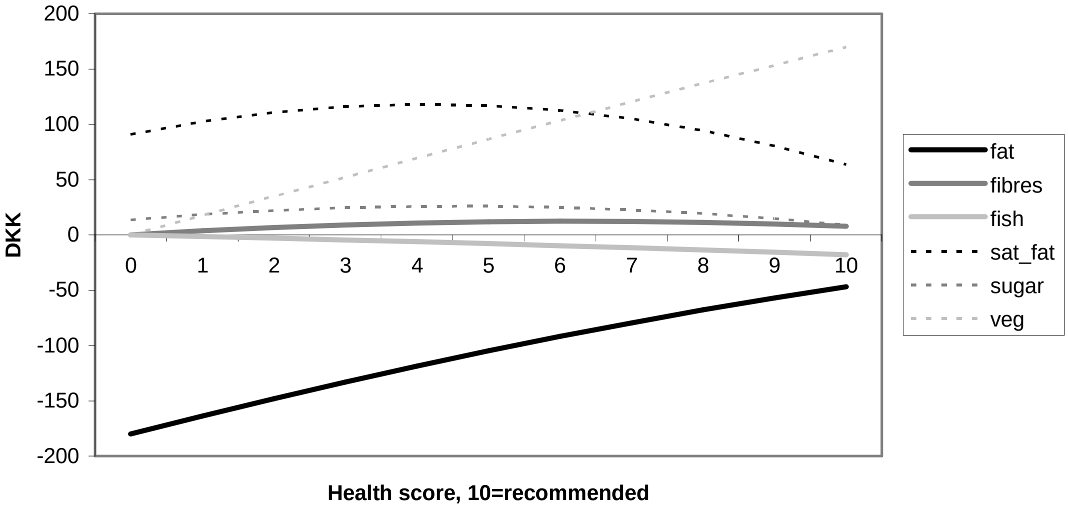
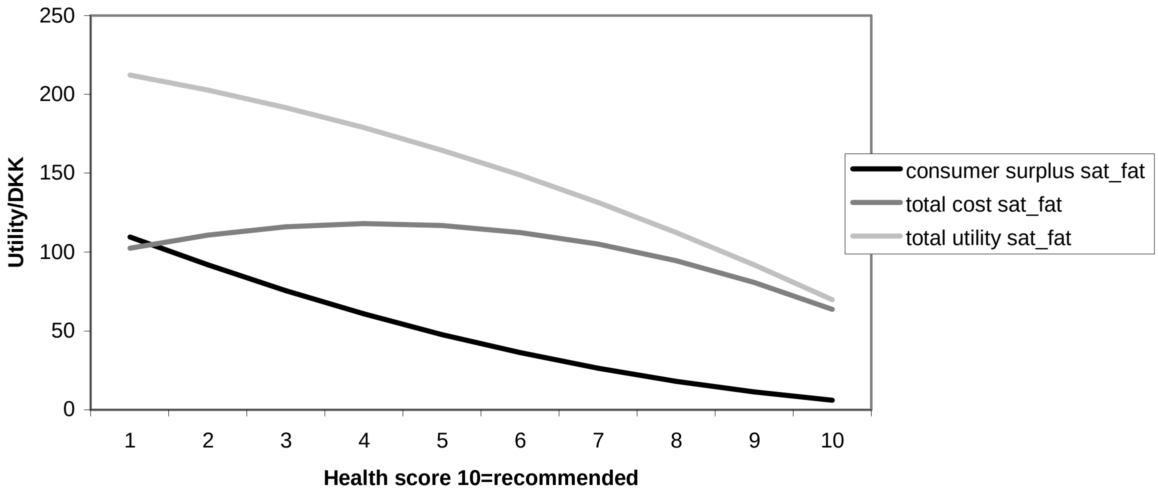


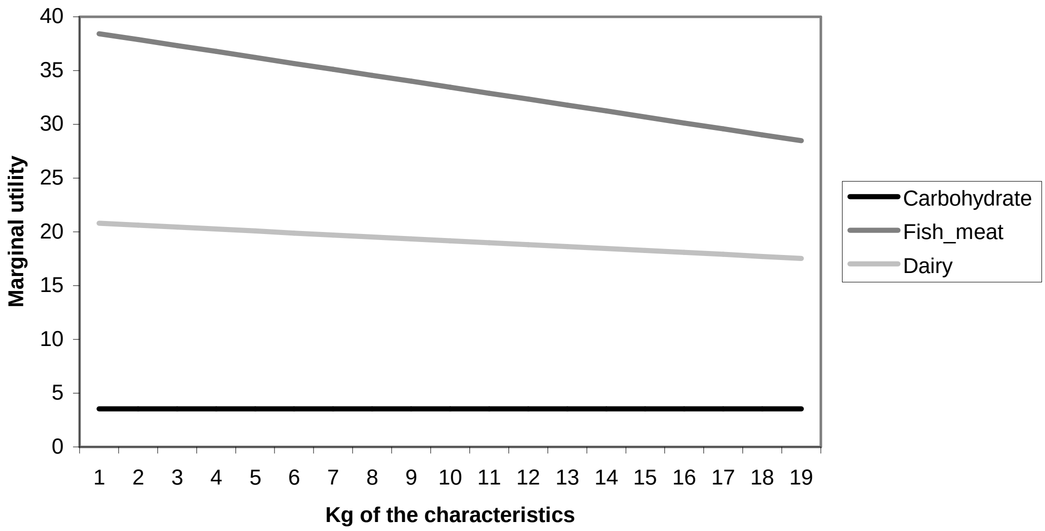
Figure 3. 5: Total cost, utility and consumer surplus for saturated fat, the average consumer



Aggregating the total costs in Figure 3.4 shows that a household fulfilling all the dietary recommendations spend 186 DKK per person monthly on health while a household scoring 5 on all recommendations spends 128 DKK per person. Due to many possible combinations of diets there is not a unique relation between costs and healthiness. E.g. a household that fulfils the recommendation for fruit and vegetables and score 5 on all the other recommendations will use 211 DKK monthly per person on health. A household that fulfil the recommendation for saturated fat and fruit and vegetables and score 5 on all the other recommendations will use 159 DKK on health monthly per person.

The marginal utility of the non-nutritional characteristics in carbohydrate containing foods, meat and fish and dairy are shown in Figure 3.6. The unit used for measuring the amount of non-nutritional characteristics is equal to the unit used for foods (one kg of meat is equal to one kg of non-nutritional characteristic from meat). Consumers value one kg of non-nutritional characteristic from meat and fish high compared to the non-nutritional characteristics from dairy and carbohydrate containing foods. The valuation of the non-nutritional characteristics from carbohydrate containing foods is low and almost linear.

Figure 3. 6: Marginal utility of the non-nutritional characteristics, the average consumer



To investigate the relationship between food expenditure and health, expenditure on health is pictured as a function of total monthly expenditure on food per person in Figure 3.7 and the share of total expenditure used on health as a function of total monthly expenditure on food per person in Figure 3.8. Expenditure on health is increasing as total food expenditure is increasing, but health takes the character of a necessity as the health share increase with a declining speed as expenditures on foods increase. The health share is declining in a nonlinear way and there is, despite the correlation, a huge variation. This is, as mentioned above, caused by that a healthy diet can be obtained in many different ways, i.e. there is not a unique way to obtain a certain level of health. These figures suggest that the costs of a healthy diet might be a barrier for some households towards a more healthy diet.

Figure 3. 7: Expenditure on health as a function of total expenditure on foods

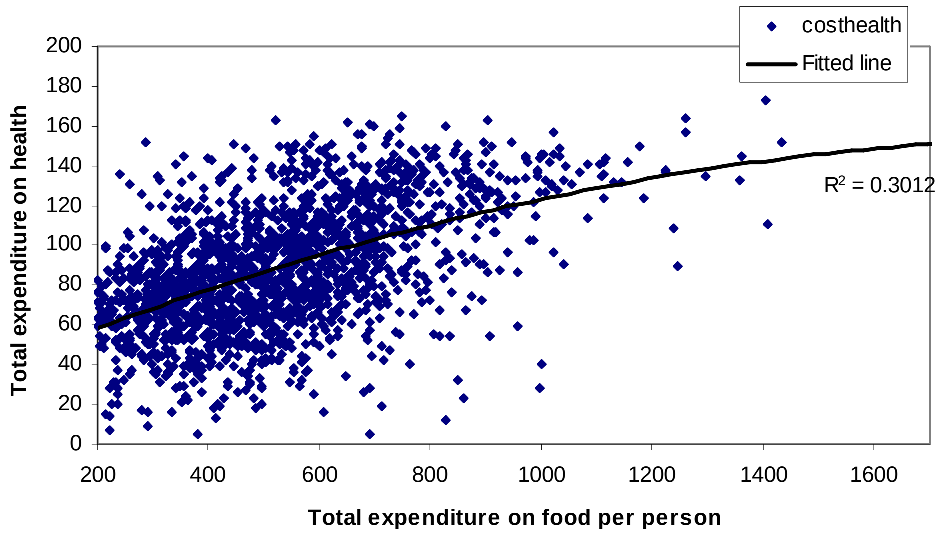
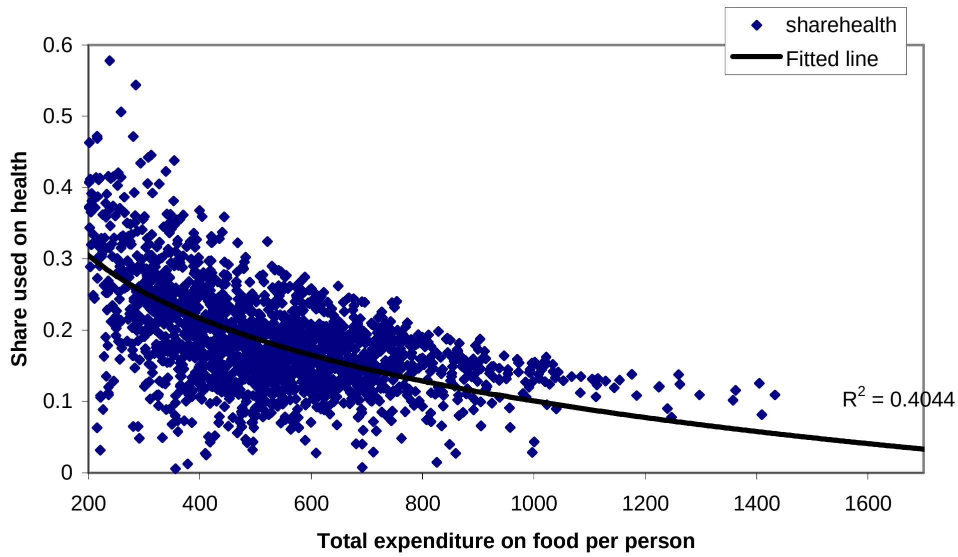


Figure 3. 8: Share of total expenditure used on health



The rest of this section is devoted to show differences in health expenditures for various types of consumers, still with the implicit prices estimated on the basis of estimation on the average consumer. Figure 3.9 shows the distribution of different types of households over values of monthly expenditure on health. Longer educated households generally have larger expenditures on health than short educated households everything else equal, which is expected due to observed food composition patterns in Smed (2008). But younger households also spend more on health than older households and this is not expected since older generally eat healthier than younger. The reason for that the younger households spend more on health might be explained by the fact that younger generally spend more on foods than older households. Another explanation is the way that households compose “healthiness”. The larger consumption of fish and fat amongst the older reduces the costs of obtaining a certain level of healthiness.

Figure 3. 9: Monthly expenditure on health per person, density over household types

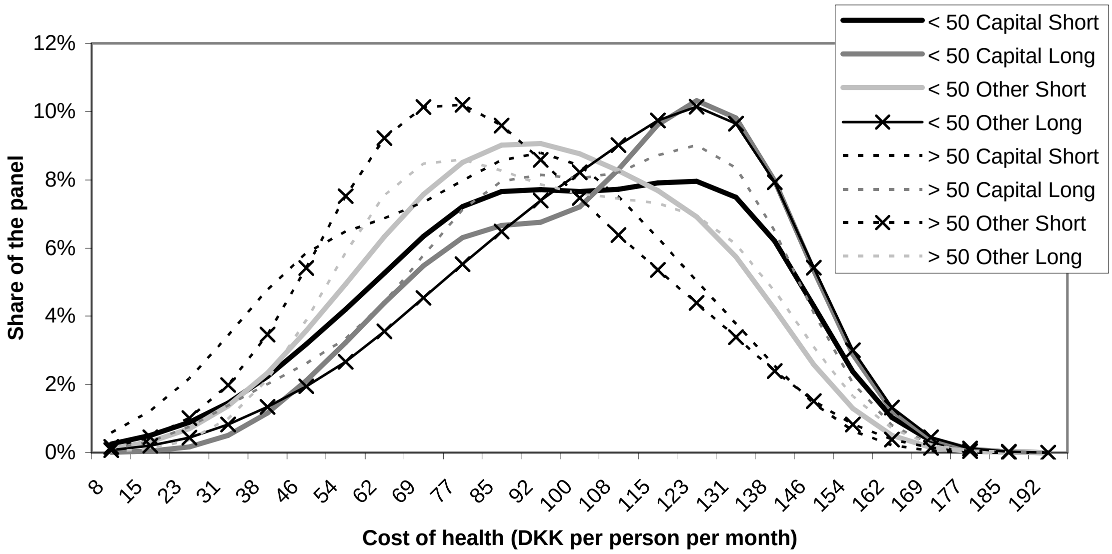
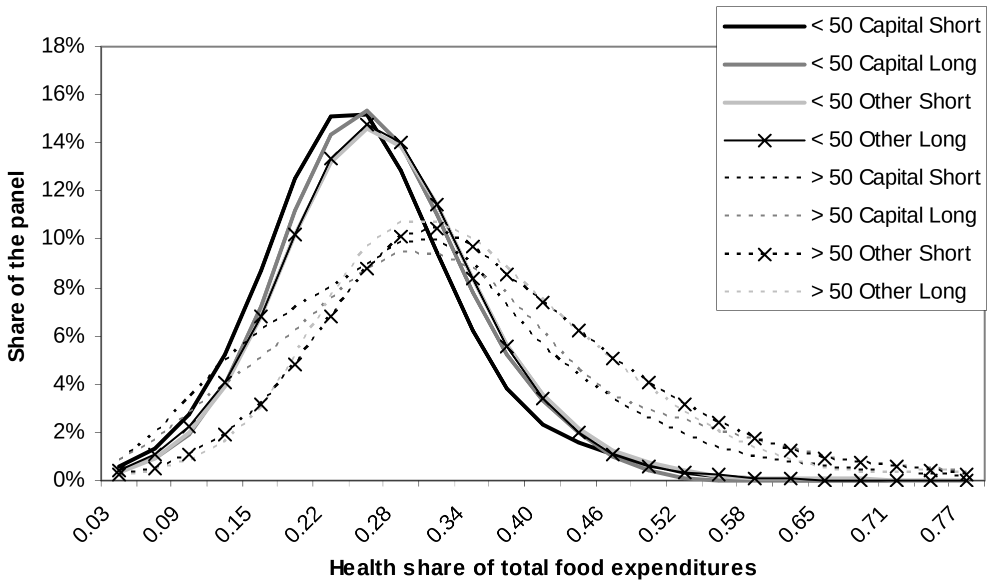


Figure 3.10 shows the density over the panel over different shares of total expenditure used on health. It is clear that older households use a larger share of total food expenditure on health.

Figure 3. 10: Share of total food expenditure used on health, density over household types



This section has assumed the implicit prices to be exogenous and as such has assumed that all consumers have the same valuation of the health and taste components. This assumption is relaxed in the next section.

3.4 Valuation of health over households

Actual dietary patterns reflect preferences for health in foods, but only to a certain extent. Together with preferences for health actual dietary patterns are also determined by budgets, prices and substitution patterns between health and other food characteristics. Preferences for taste have a large influence on healthiness of diets, since e.g. preference for beef instead of pork might diminish the amount of saturated fat consumed together with the meat. Preferences for taste are to a large extent determined by traditions and habits. Finally budgets might drive a wedge in between estimated preferences for health and observed food expenditure patterns. Therefore, in the following the assumptions of equal valuation of health over households are relaxed to uncover preferences for healthiness and for taste for different types of households. This is equivalent to estimate equation (3.21), now allowing $\pi, \hat{\pi}, \tau$ and $\hat{\tau}$ to depend on social and demographic variables:

$$X_{nt} = \sum_{m=1}^{M} \left(\tau_{mn} \cdot g_{nmt} - \hat{\tau}_{mn} \cdot (g_{nmt})^2 \right) + \sum_{i=1}^{I} \left(\pi_{in} \cdot z_{nit} - \hat{\pi}_{in} \cdot (z_{nit})^2 \right) + \eta_n + \varepsilon_{nt} \qquad (3.27)$$

The social and demographic variables chosen are those with the larges influence on healthy eating behaviour according to Smed (2008). These are as follows; below 50 years of age versus above 50, shorter education, (i.e. none, vocational or shorter theoretical education) versus longer education, (i.e. medium or long theoretical education) and finally capital versus other location. The original parameters and t values of the estimation are shown in Appendix 3D. From the estimated parameters it is possible to infer the marginal utility of getting one unit closer to each of the dietary recommendations using the same recalculation procedure as in (3.26).

Figure 3.11 shows the marginal utility of getting closer to the recommendation for fat (cutting back on consumption) everything else equal evaluated at a health score equal to 5 and a health score equal to 10 (10 = recommendation fulfilled). Longer educated have a higher valuation over the health score for total fat, except older households in other regions where there is no significant difference between valuations for the longer and shorter educated. Younger households have a higher valuation than older households. Households in the capital have a higher valuation than households in other regions, except older shorter educated households where the difference is not significant.

Figure 3. 11: Marginal utility over the health score for fat, various types of households

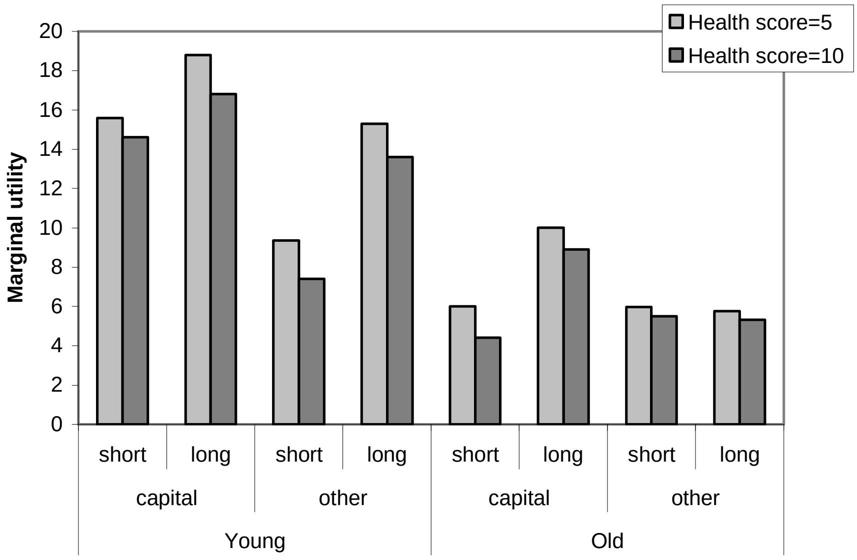


Figure 3.12 shows marginal utilities over the health score for sugar for different social and demographic groups. There are no particular age or educational differences, but there is a clear tendency that younger households value getting closer to the recommendation for sugar lower than older households. This indicates that especially younger households prefer to have a diet with a lot of sugar.

Figure 3. 12: Marginal utility over the health score for sugar, various types of households

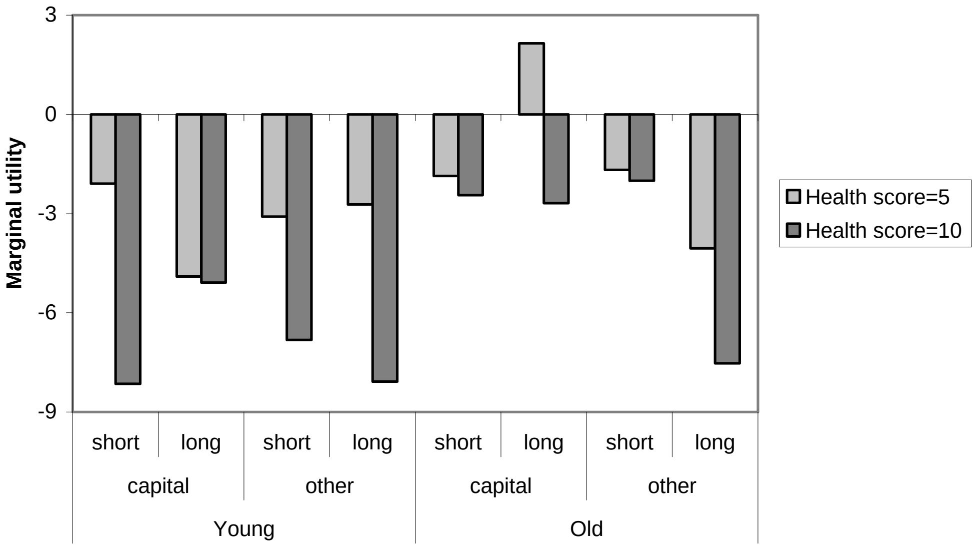
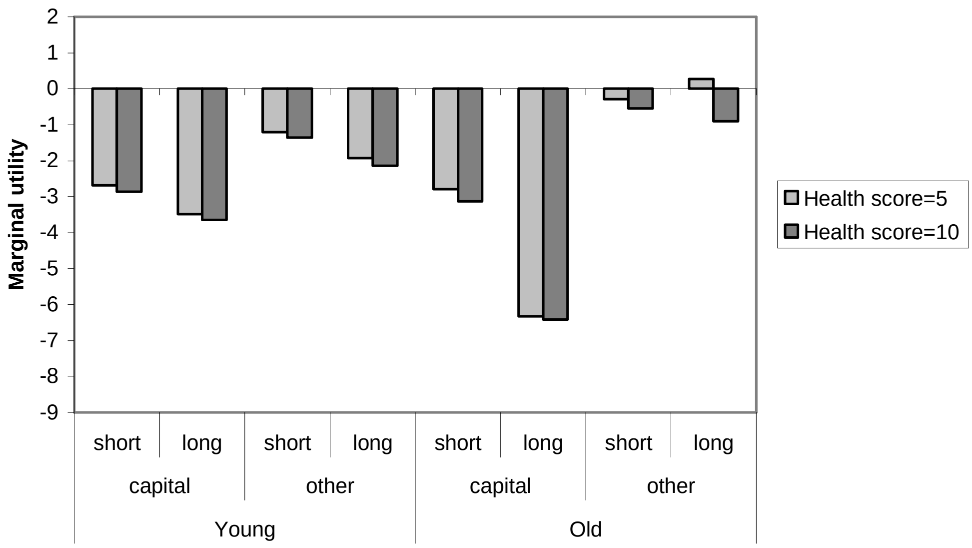


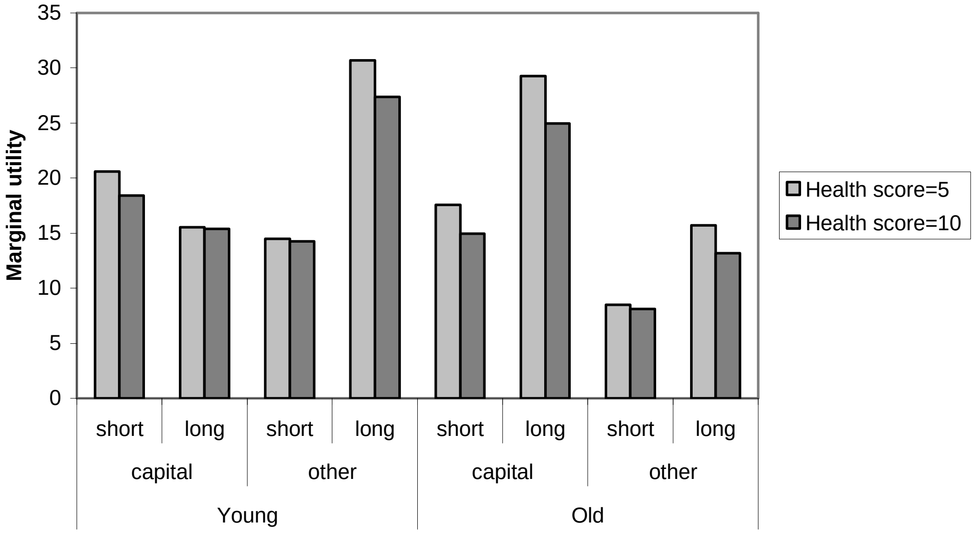
Figure 3.13 shows the marginal valuation over the health score for fish. For households located in other regions younger households have a lower valuation than older. Households in the capital seem to have a lower valuation of getting closer to the recommendations for fish than households living elsewhere. Both are reflected in actual consumption patterns (Smed, 2008). Longer educated seem to have a lower valuation than shorter educated households.

Figure 3. 13: Marginal utility over the health score for fish, various types of households



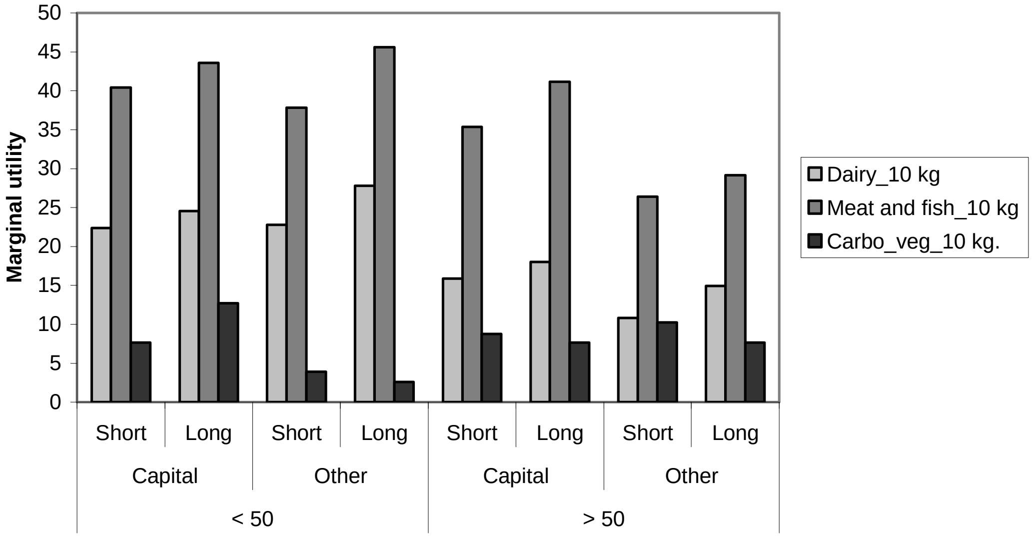
Finally Figure 3.14 shows the marginal valuation over the health score for fruit and vegetables. The figure shows that longer educated value health from fruit and vegetables more than shorter educated, and households in the capital have a larger valuation than households in other regions. An exception is younger households in the capital. These general patterns reflect actual consumption patterns pictured in Smed (2008). Younger value getting closer to the recommendations for fruit and vegetables more than older households, which is not in accordance with actual consumption patterns (Smed, 2008). The explanation for this might be due to either budget constraints or differences in the valuation of the non-nutritional characteristics. All types of households have a negative valuation of the health score for saturated fat (figure not shown). No regional or educational differences are found, but there is a clear tendency that younger value getting closer to the recommendation for fat more negatively than older households. The marginal valuation of the health score for fibres (figure not shown) shows clear educational and age patterns with the longer educated valuing health from fibres less negative than shorter educated and younger having a less negative valuation than older.

Figure 3. 14: Marginal utility over the health score for fruit and vegetables, various types of households



Whenever household valuation of health is not in accordance with actual consumption this might be due to substitution patterns between nutritional characteristics, budgets or that these particular households value the non-nutritional characteristics higher than the nutritional characteristics. Figure 3.15 shows the marginal utility of the non-nutritional characteristics valued at 10 kg. It is chosen to picture the marginal utility of the non-nutritional characteristics in a point only since it adds to the clarity of the picture and the quadratic term is rather small for the non-nutritional characteristic. The reason for the marginal utility being close to constant might be due to large substitution possibilities within each group of non-nutritional characteristics. The marginal utility of the non-nutritional characteristic from meat and fish lies between 25 DKK/kg and 45 DKK/kg. It is higher for longer educated households and households in the capital, apart from longer educated younger households, where households in other regions have the largest marginal utility. Younger households also have a higher marginal utility than older households. This applies to all kinds of households. The marginal utility of the non-nutritional characteristic from dairy takes values from 11 DKK/kg to 28 DKK/kg. Younger households have a larger valuation than older households and longer educated households have a larger valuation than shorter educated households. The marginal utility of the non-nutritional characteristic from carbohydrate containing foods is rather small and shows no specific patterns.

Figure 3. 15: Average marginal utility of non-nutritional characteristics, calculated at 10 kg.

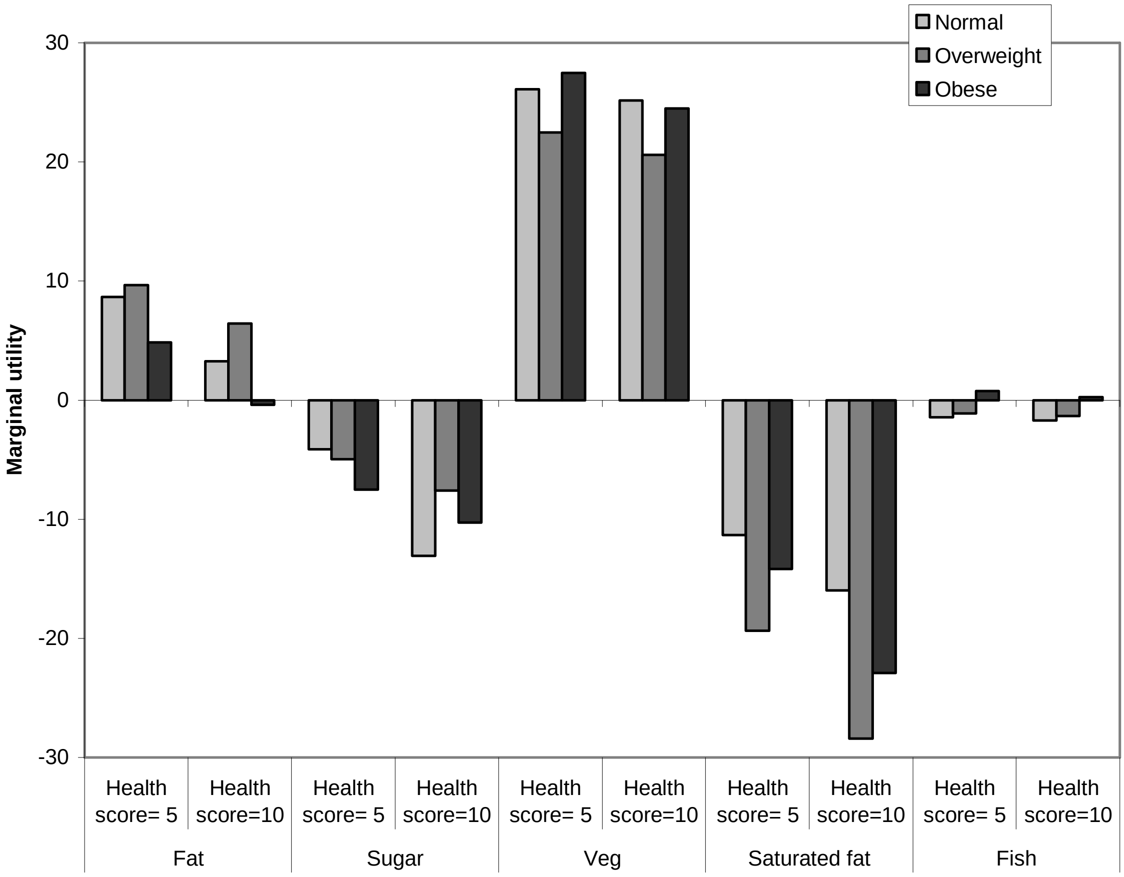


To finish this section, equation (3.27) is estimated for households where the diary keeper has a normal weight (BMI< 25) [34], are overweight (BMI 25-30) or obese (BMI 30 or above)[35]. Figures 3.16 shows the marginal valuation of the health score for selected recommendations evaluated at the health score equal to 5 and the health score equal to 10 (this is equal to that the recommendation is fulfilled). Obese has a smaller valuation of the health score than normal weight for the recommendations which implies a cutback on consumption (i.e. the recommendations for sugar, saturated fat and total fat). This is especially clear for saturated fat where also the overweight have a lower valuation than normal weight. For fish and vegetables obese have a larger valuation than normal weight. This might reflect that obese and overweight have preferences for eating more rather than for eating unhealthy when measured through the healthy eating index.

[34] BMI is calculated as $\frac{weight(kg)}{(height(m))^2}$

[35] The height and weight questions used to calculate the BMI are only posed in 2004, so the panel is divided according to their weight the final year.

Figure 3. 16: Marginal utility over the health score for fat, various BMI



3.5 Conclusion and discussion

The results of this paper underline the importance of taking individual heterogeneity into account when estimating implicit prices. Comparing results from a model where social and demographic variables are inserted to account for variation between households with a model in differences, removing all individual heterogeneity, shows that the model in differences is more convenient for the estimation of hedonic prices. Furthermore, the results show the importance of taking the valuation of taste into account. Using the model in differences and estimating the average consumers' valuation of health from foods show that the recommendations are not valued equally. Consumers are found to have a positive valuation of the recommendations for fat, fruit and vegetables and fibres. This does not seem in accordance with Smed (2008) finding that fruit and vegetables and fibres are the two areas where consumers are farthest away from fulfilling the dietary recommendations. For fruit and vegetables the lack of fulfilment might be due to the large cost involved in fulfilling the recommendation for fruit and vegetables. The costs involved with fulfilling the recommendations for fibres are not large, so the explanation for the lack of fulfilment here might not be due to the costs involved, but might be explained by a trade off between taste and health. This might be a route for further research. Consumers have a negative valuation of the recommendation for fish, sugar and saturated fat. From this we can conclude that consumers do not eat fish for their health value and they like to have a diet rich in sugar and saturated fat. This is in accordance with Drewnowski (1995) and Birch (1999), who suggest that consumers have preferences for energy dense foods. Under certain assumptions the implicit prices can be interpreted as exogenous. Using this, calculation of total expenditure on health shows a clear tendency that a healthier diet is more costly than a more unhealthy diet. Furthermore, the expenditure on health increases with the total expenditure on foods, suggesting that cost might be a barrier for some households in order to change towards a more healthy diet. This is in accordance with earlier research (see e.g. Darmon et al., 2002; Lennernas et al., 1997; Adelaja et al., 1997). But as food expenditures increases, health expenditure does not increase just as fast so the share of total food expenditure devoted to health is decreasing as food expenditure is increasing; health is a necessity. The share of total expenditure on health follows a non-linear path and there is, despite the correlation, a huge variation. This is due to that a healthy diet can be obtained in many different ways, i.e. it is costly to increase the healthiness of diets by increasing the consumption of fruit and vegetables, while money is saved if the improvement in health is obtained through cutting back on e.g. fat. Health expenditures are increasing with education everything else being equal, which is expected due to observed food composition patterns. But younger households also spend more on health than older and this is not expected since older households generally eat healthier than younger and health expenditure is positively correlated with the healthiness of diets. The explanation for this might be found in the non-uniqueness of having a healthy diet, i.e. a diet might be healthy in one direction and unhealthy in another.

The estimated valuation of health characteristics for various types of households reflects preferences since observed dietary patterns also reflect reactions to budgets and prices as well as substitution patterns. The results from the estimation of the valuation of healthiness for various types of households show that longer educated generally have a larger valuation of the dietary recommendations, with fish as the one exception. Households in the capital also value most dietary recommendations more than households in other regions. Besides fish, younger households value health more than older households. This is not reflected in the actual dietary patterns. Whenever household valuation of getting closer to the recommendations is not in accordance with the actual consumption this might be due to these particular households valuing the non-nutritional characteristics to a larger extent than the nutritional characteristics, substituting towards e.g. taste instead of health. The marginal utility of the non-nutritional characteristic from both dairy products and meat and fish is higher for younger households than older everything else equal. Younger households value health relatively high, but value also taste relatively high, which can explain why younger households can have a higher valuation of health compared to older households, but still eat unhealthier. Therefore, preferences for non-nutritional characteristics will have a large influence on healthiness of diets and whenever it is aimed to influence consumers' diets in a more healthy direction it is important to take this into consideration. Also households in the capital and households with a longer education have higher valuation of the non-nutritional characteristics for meat, fish and dairy products than other types of households. Therefore, there seems to be a positive correlation between the valuation of health and the valuation of the non-nutritional characteristics. This suggests that consumers with a general low valuation of all the characteristics from food eat less healthy than households that value food in general. When analysing the valuation of health for individuals being either normal weight, overweight or obese, the general results are that obese has a smaller valuation of the health score than normal weight for the recommendations which implies a cut back on consumption (i.e. the recommendations for sugar, saturated fat and total fat), but overweight and obese have a larger valuation of the health score than normal weight for the recommendations involving an increase in consumption (i.e. fish, fibres and fruit and vegetables). This might reflect that obese and overweight have preferences for eating more rather than for eating unhealthy when measured through the healthy eating index. This might be a route for further research.

References

Adelaja A.O., R.M. Nayga and T.C. Lauderbach (1997): Income and racial differentials in selected nutrient intakes. *American Journal of Agricultural Economics*, 79(5): 1452-1460

Astrup A., N.L. Andersen, S. Stender and E. Trolle (2005): *Kostrådene 2005*. Publication no. 36, Ernæringsrådet & Danmarks Fødevareforskning, Copenhagen

Atkinson A.B. (1995): *Income and the Welfare State, Essays on Britain and Europe.* Cambridge University Press, Cambridge, NY

Bartik T. (1987): The estimation of demand parameters in hedonic price models. *Journal of Political Economy*, 95(1), 81-88

Birch L.L. (1999): Development of food preferences. *Annual Review of Nutrition*, 19: 41–62

Cropper M.L., L.B. Leland and K.E. McConnel (1988): On the choice of functional form for hedonic price functions. *The Review of Economics and Statistics*, 70 (4): 668-675

Darmon N., Ferguson E.L. and Briend A. (2002): A cost constraint alone has adverse effects on food selection and nutrient density: An analysis of human diets by linear programming. *The Journal of Nutrition*, 132: 3764 –3771

Deaton A. and J. Muellbauer (1980): *Economics and consumer behaviour.* Cambridge University Press, Cambridge, UK

deJanvry A. (1966): *Measurement of demand parameters under separability.* Phd dissertation, Department of agricultural and economics, University of California- Berkeley

Drewnowski A. (1995): Energy intake and sensory properties of food. *American Journal of Clinical Nutrition*, 62(5 suppl.):1081–1085

Holm L., A. Astrup, A.W. Dynesen and J. Haraldsdóttir (2002): De store ernæringskampagner virker faktisk!. *Ugeskrift for læger*, 164(5): 649

Ladd G.W. and M. Zober (1977): Model of consumer reaction to product characteristics. *The Journal of Consumer Research*, 4(2): 89-101

Ladd G.W. and V. Suvannunt (1976): A Model of Consumer Goods Characteristics. *American Journal of Agricultural Economics*, 58: 504-510

Lancaster K. (1966): A new approach to consumer theory. *Journal of Political Economy*, 74(2): 132-157

Lennernas M., C.Fjellström, W. Becker, I. Giachetti, A. Schnitt, A.M. Remaut de Winter and M. Kearey (1997): Influences on food choice perceived to be important by nationally representative samples of adults in European Union. *European Journal of Clinical Nutrition*, 51: S8-S15

Lenz J.E., R.C. Mittelhammer and H. Shi (1994): Retail-level hedonics and the valuation of milk components. *American Journal of Agricultural Economics*, 76(3): 492-503

Pedersen S. and M. Kjær (2007): *Omnibus undersøgelse om frugt og grønt.* Tranberg marketing recommendations, Copenhagen, Denmark

Ranney C.K. and P.E. McNamara (2002): *Do healthier diets cost more?* Paper presented at the American Agricultural Economics Association's Annual meeting, July 28-31 2002, Long Beach, CA

Rosen S. (1974): Hedonic prices and implict markets: Product differentiation in pure competition. *Journal of Political Economy*, 74(1): 34-55

Shi H. and D.W. Price (1998): Impact of sociodemographic variables on the implicit values of breakfast cereal characteristics. *Journal of Agricultural and Resource Economics*, 23(1):126-139

Smed S., J.D. Jensen and S. Denver (2007): Socio-economic characteristics and the effect of taxation as a health policy instrument. *Food Policy*, 32 (5):.624-639

Smed (2008): Chapter 2, this thesis

Smith T.G. (2002a): *The McDonald's Equilibrium: Advertising, Empty Calories, and the Endogenous Determination of Dietary Preferences.* Departmental Working Paper no.17-02, University of California, Department of Economics, Santa Barbara

Smith T.G. (2002b): *Obesity and Nature's Thumbprint: How Modern Waistlines Can Inform Economic Theory.* Departmental Working Paper no. 18-02, University of California, Department of Economics, Santa Barbara

Strukturdirektoratet (2000): *Evaluering af foreningen fiskebranchens Markedsføringskampagne.* Strukturdirektoratet, Copenhagen

Yanovski S. (2003): Sugar and fat: Cravings and aversions. *Journal of Nutrition*, 133(3): 835S–837S

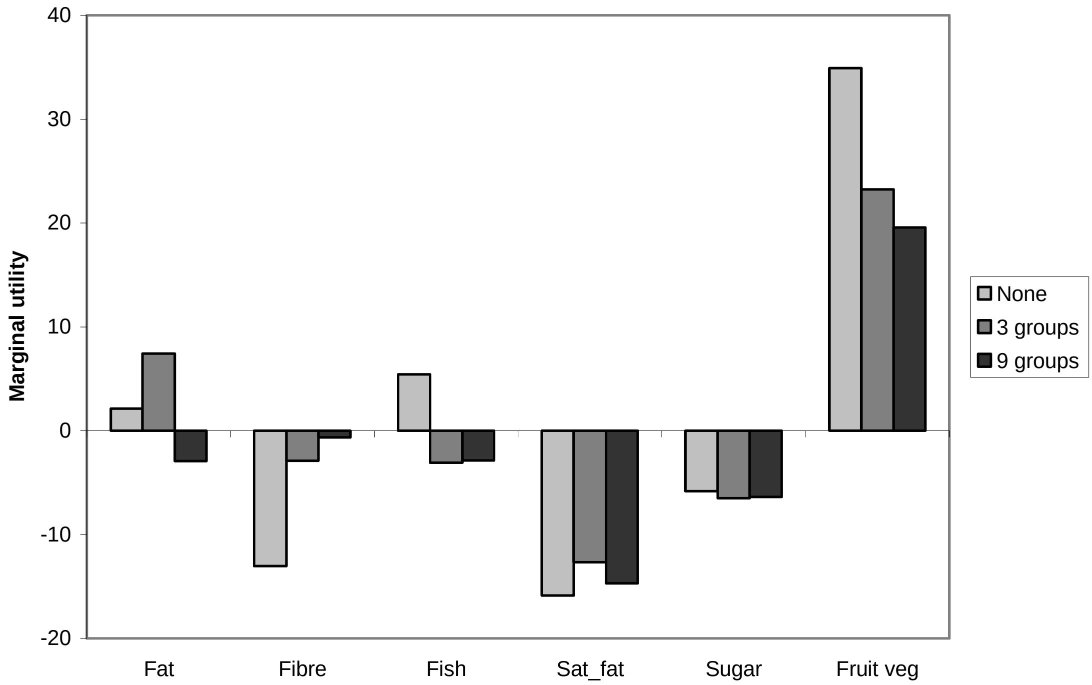
Appendix 3.A: The aggregation of foods into three main groups

Basic food group category	**Grouping in GfK data**
Meat, fish and fats	Other meat
	Bacon
	Fish
	Processed fish
	Poultry
	Rissole
	Processed meat for bread
	Liver pâté
	Beef
	Sausages
	Pork
	Brawn and pâté
	Butter
	Oil
	Margarine
Carbohydrate-containing foods	Fruit syrup
	Ice tea
	Juice
	Coffee
	Tea
	Fizzy drinks
	Bouillon and soups
	Dishes with rice and pasta
	Salad dressing etc.
	Sauce
	Pizza
	Baking mixture
	Chocolate (for bread)
	Ice cream
	Biscuits
	Macaroons
	Marmalade
	Cake
	Sugar
	Cookies
	Fruit
	Vegetables
	Frozen vegetables
	Potatoes
	Cereals
	White bread
	Brown bread
	Flour
	Crisp bread
	Rice
	Pasta
Dairy	Speciality cheese
	Ordinary cheese
	Dairy snack
	Milk
	Yoghurt
	Eggs

Appendix 3.B: Models with no, 3 or 9 non-nutritional characteristics

Figure 3.B.1 shows the differences in the valuation of health estimated with models containing either no, 3 or 9 non-nutritional characteristics. Note that the values for the model with no non-nutritional characteristics are slightly different from the numbers in Table 3.3 since the present model is estimated for the whole period 1999-2004 while the numbers in Table 3.3 are only based on 2004.

Figure 3.B.1: Comparison of models with no, 3 or 9 non-nutritional characteristics



Appendix 3.C: The quadratic model, parameters and standard deviation

Table 3.C.1: Parameters and std. errors for the quadratic model, average consumer

	Estimate	Std_error	t-value	Pr > |t|
Non_nut_carbo	0.0035	0.0000	128.4800	<.0001
Non_nut_carbo_sq	0.0000	0.0000	120.6600	<.0001
Non_nut_meatfish	0.0390	0.0003	139.0600	<.0001
Non_nut_meatfish_sq	0.0000	0.0000	40.7200	<.0001
Non_nut_dairy	0.0210	0.0002	124.1200	<.0001
Non_nut_dairy_sq	0.0000	0.0000	40.3800	<.0001
Dummy_jan_carbo	0.0025	0.0000	64.5300	<.0001
Dummy_jan_meatfish	-0.0019	0.0004	-4.7300	<.0001
Dummy_jan_dairy	-0.0028	0.0003	-10.4000	<.0001
Dummy_jan_fish	92.0921	63.3423	1.4500	0.1460
Dummy_jan_fruitveg	-407.6020	29.6651	-13.7400	<.0001
Dummy_jan_fat	74.0009	15.7520	4.7000	<.0001
Dummy_jan_satfat	84.2423	31.5793	2.6700	0.0076
Dummy_jan_sugar	-1.0519	15.9200	-0.0700	0.9473
Dummy_jan_fibre	458.7016	242.8000	1.8900	0.0588
Dummy_dec_carbo	0.0013	0.0000	48.6300	<.0001
Dummy_dec_meatfish	-0.0010	0.0004	-2.6200	0.0089
Dummy_dec_dairy	-0.0033	0.0003	-12.3100	<.0001
Dummy_dec_fish	38.6187	63.0190	0.6100	0.5400
Dummy_dec_fruitveg	-205.2370	38.5318	-5.3300	<.0001
Dummy_dec_fat	47.0205	15.1129	3.1100	0.0019
Dummy_dec_satfat	109.8163	30.3555	3.6200	0.0003
Dummy_dec_sugar	-13.4972	14.8749	-0.9100	0.3642
Dummy_dec_fibre	384.2897	237.4000	1.6200	0.1054
Fat	-116.2390	19.2573	-6.0400	<.0001
Fat_squares	131.0896	23.2259	5.6400	<.0001
Fibre	1654.1530	202.0000	8.1900	<.0001
Fibre_sq	55577.3200	4612.3000	12.0500	<.0001
Fish	-447.5300	37.7905	-11.8400	<.0001
Fish_sq	4132.7070	303.7000	13.6100	<.0001
Sat_fat	758.8347	35.0672	21.6400	<.0001
Sat_fat_sq	1219.3600	107.9000	11.3000	<.0001
Sugar	133.2814	8.1264	16.4000	<.0001
Sugar_sq	170.7301	21.3110	8.0100	<.0001
Veg	2069.2570	16.9995	121.7200	<.0001
Veg_sq	810.0181	11.2712	71.8700	<.0001
R^2 = 68.54.				

Appendix 3.D: Parameters, estimation for different households

Table 3.D 1: Parameters and std. errors for the quadratic model, differentiated consumers

	< 50 yrs, Capital, short		< 50 yrs, Capital, long		< 50 yrs, other, short		< 50 yrs, other, long	
Non_nut_carbo	0.0077	<.0001	0.0127	<.0001	0.0039	<.0001	0.0026	<.0001
Non_nut_carbo_sq	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001
Non_nut_meatfish	0.0404	<.0001	0.0436	<.0001	0.0379	<.0001	0.0456	<.0001
Non_nut_meatfish_sq	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001
Non_nut_dairy	0.0224	<.0001	0.0245	<.0001	0.0228	<.0001	0.0278	<.0001
Non_nut_dairy_sq	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001
Fat	-264.2	<.0001	-284.0	0.0077	-100.7	0.0018	-228.1	0.0021
Fat_sq	74.9	0.2485	152.3	0.2475	149.4	0.0001	129.6	0.1593
Fibre	96.7	0.8767	-2375.1	0.0999	488.3	0.1692	1143.4	0.1051
Fibre_sq	14066.9	0.3025	46877.6	0.1538	36220.4	<.0001	42286.3	0.0040
Fish	-806.6	<.0001	-1077.0	<.0001	-341.1	<.0001	-551.7	<.0001
Fish_sq	3812.3	<.0001	3168.1	0.0709	3092.6	<.0001	4398.6	<.0001
Sat_fat	1271.4	<.0001	1224.4	<.0001	732.3	<.0001	1378.7	<.0001
Sat_fat_sq	2355.5	<.0001	1523.5	0.0226	1076.6	<.0001	2900.7	<.0001
Sugar	165.2	<.0001	87.6	0.0920	132.5	<.0001	160.9	<.0001
Sugar_sq	353.0	0.0008	10.7	0.9410	217.1	<.0001	311.9	0.0001
Veg	2676.9	<.0001	1848.6	<.0001	1737.1	<.0001	4008.2	<.0001
Veg_sq	5994.9	<.0001	454.1	0.6544	698.7	<.0001	9273.9	<.0001
	> 50 yrs, Capital, short		**> 50 yrs, Capital, long**		**> 50 yrs, other, short**		**> 50 yrs, other, long**	
Non_nut_carbo	0.0088	<.0001	0.0077	<.0001	0.0103	<.0001	0.0077	<.0001
Non_nut_carbo_sq	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001
Non_nut_meatfish	0.0354	<.0001	0.0412	<.0001	0.0264	<.0001	0.0291	<.0001
Non_nut_meatfish_sq	0.0000	0.0015	0.0000	0.8224	0.0000	<.0001	0.0000	0.4808
Non_nut_dairy	0.0159	<.0001	0.0180	<.0001	0.0108	<.0001	0.0149	<.0001
Non_nut_dairy_sq	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001	0.0000	<.0001
Fat	-49.5	0.4707	-149.2	0.1781	-96.9	0.0093	-94.3	0.2324
Fat_sq	123.0	0.1420	84.7	0.5433	36.2	0.4256	33.6	0.7351
Fibre	-664.8	0.3156	648.4	0.5431	282.5	0.4646	275.7	0.7535
Fibre_sq	2734.1	0.8452	40807.3	0.0462	29110.2	0.0014	7686.7	0.7017
Fish	-791.6	0.0001	-2013.1	<.0001	-8.5	0.9170	468.2	0.0358
Fish_sq	7047.6	0.0009	1816.3	0.3934	5441.6	<.0001	24537.4	<.0001
Sat_fat	889.1	<.0001	520.8	0.0191	846.4	<.0001	928.4	<.0001
Sat_fat_sq	1580.5	<.0001	77.3	0.9182	1314.2	<.0001	1742.2	0.0002
Sugar	44.2	0.1963	66.6	0.3735	35.7	0.0133	143.3	0.0003
Sugar_sq	34.2	0.7219	281.4	0.3155	19.3	0.5461	202.0	0.0932
Veg	2375.3	<.0001	3946.1	<.0001	1043.6	<.0001	2147.5	<.0001
Veg_sq	7247.9	<.0001	11866.1	<.0001	1027.1	<.0001	7020.3	<.0001
Dummy_jan_carbo	0.001	<.0001						
Dummy_jan_meatfish	0.000	0.2547						
Dummy_jan_dairy	-0.001	<.0001						
Dummy_jan_fish	19.814	0.7488						
Dummy_jan_fruitveg	-93.120	0.0036						
Dummy_jan_fat	33.635	0.0282						
Dummy_jan_satfat	14.932	0.6269						
Dummy_jan_sugar	-4.579	0.7675						
Dummy_jan_fibre	279.013	0.2375						
Dummy_dec_carbo	0.000	<.0001						
Dummy_dec_meatfish	0.000	0.5755						
Dummy_dec_dairy	-0.002	<.0001						

Dummy_dec_fish	14.647	0.8114
Dummy_dec_fruitveg	-72.398	0.0586
Dummy_dec_fat	22.686	0.1220
Dummy_dec_satfat	64.229	0.0293
Dummy_dec_sugar	-22.436	0.1210
Dummy_dec_fibre	279.039	0.2271

Chapter 4

A censored structural characteristics model for milk

Laura M. Andersen and Sinne Smed

AKF
(Danish Institute of Governmental Research)
Copenhagen

Abstract

In this paper we investigate preferences for fat in milk through a structural characteristics model. Contrary to the usual hedonic model consumers' preferences over certain characteristics are allowed to vary non-systematically through an error term placed directly in the utility function. The functional form used is the quadratic form allowing the marginal utility of characteristics to become negative. In the empirical estimations we use a very comprehensive panel dataset spanning the period from 1997 to 2004. The data includes information about daily purchases and social and demographic characteristics of approximately 2500 households. These data are combined with information indices constructed from articles in newspapers mentioning a link between the consumption of fat and health. The panel structure of the data is exploited fully since the final two-sided censored Tobit model is estimated household by household allowing for the maximum degree of individual heterogeneity. We find that there has been a significant decrease in the consumption of fat from milk generated by systematic changes in preferences due to information and due to a general trend. In the discussion of whether to use prices or information as an instrument to decrease the consumption of fat from milk, prices seem the most effective. Consumers who prefer milk with a very high fat content can be reached both by information and prices, while consumers who prefer milk with a moderate to high fat share are not influenced by information, but are rather price sensitive. This is of great importance since the latter drink more milk and thereby consume most milk fat in grams per person per week.

Introduction

Health problems related to an excessive intake of saturated fat are among the major nutrition problems in most industrialised countries, as a high intake of saturated fat can lead to increased blood cholesterol levels and risk of various lifestyle-related illnesses. Since Denmark is a nation of milk drinkers with an annual consumption of about 100 kg per capita (Statistics Denmark, 2008) and saturated fat from milk constitutes on average 5.7 per cent of total consumption of saturated fat and 3.1 per cent of total fat consumption[36], milk may be an important source of fat. The consumption of saturated fat from milk has decreased during the last decade (Statistics Denmark, 2008), which in part might be a reaction upon massive campaigning from the Danish health authorities against an excessive intake of saturated fat, but also to a large extent due to the entrance of low fat varieties on the milk market (Smed and Jensen, 2004). These changes on the milk market give a good possibility to investigate preferences for saturated fat, how they can be expressed through demand and how they change over time and due to information. The demand for milk in Denmark has been analysed in a number of previous studies. Blow et al. (2005) develop a non-parametric revealed preference model for milk at household level and find that there are three types of consumers: those who have a high valuation of fat and a low valuation of the organic attribute in milk; those who have a moderate valuation of fat and a high valuation of the organic attribute and finally those who have a low valuation of fat and a high valuation of the organic attribute. From Smed and Jensen (2004) there is market evidence that there is a substantial trade-off between health concern and taste, since taste is valued higher than the fat content.

In this paper we investigate preferences for fat in milk in depth through a structural characteristics model, i.e. a model where consumers derive utility from the characteristics inherent in milk, not from milk itself (Lancaster, 1966; Gorman, 1956). This means that the demand for fat in milk has to be described as demand for a non-market good. Demand for non-market goods is often estimated through a hedonic model derived from the Gorman-Lancaster framework (for examples on the demand for nutrients in food, see e.g. Cook and Eastwood, 1992; Kim and Chern, 1995 or Eastwood et al., 1986). In the hedonic model it is usually assumed that consumers' preferences are stable over time and random noise is placed as an error term in the estimation equation, i.e. as random deviation from the true preferences. In this paper we test whether consumers' preferences over certain characteristics are stable or

[36] Own calculations based on the data from GfK Denmark used in this paper.

if they vary non-systematically through an error term placed directly in the utility function. Furthermore, we introduce systematic changes in preferences initiated by a trend and exogenous health information. The data used for the estimations are based on an extensive panel dataset at household level. This means that it is possible to estimate the models household by household allowing for the maximum degree of individual heterogeneity. There is a need to understand possible barriers for further reductions in the intake of saturated fat since this knowledge may be essential for the design of new actions aiming at reducing the intake of saturated fat. The derivation of a structural model for individual households brings us closer to separating preferences and changes in these due to e.g. information from reactions to prices and budget constraints and also to predict demand for none existing goods consisting of new combination of already existing characteristics on the market. In other words, it allows us to give a more interesting answer – not only to how much fat is consumed, but also why consumers choose to consume as they do.

The rest of this paper is organised as follows: Section 4.1 starts out with the basic theory of the characteristics model and then the data and the milk markets are described in Section 4.2. Section 4.3 is about empirical considerations and estimation issues, especially about the construction of prices in the characteristics model, the implications of choosing a quadratic model and the derivation of a model with an error term in the utility function. Section 4.4 summarises the results of the introductory model. Then in Section 4.5 the model is reformulated according to the best suited model to allow estimation of a Tobit model with two-sided censoring. Finally, Section 4.6 describes the final results, i.e. valuation of fat and reactions to prices and information for different types of households and predictions of demand. Section 4.7 is devoted to a discussion and conclusion.

4.1 The characteristics model

The characteristics model was first developed by Gorman (1980) and Lancaster (1966) and further developed by Muellbauer (1974) and Rosen (1974). Generally, we assume that the world consists of H individual households. The number of goods available in each period is I and the number of characteristics is J. The connection between goods q and characteristics z is described through the technology matrix A.

$$
\textit{Goods}\ \begin{cases} 1 \\ \vdots \\ i \\ \vdots \\ I \end{cases}
\overbrace{\begin{matrix} 1 & \cdots & j & \cdots & J \end{matrix}}^{\textit{Characteristics}}
\begin{pmatrix}
a_{11} & \cdots & a_{1j} & \cdots & a_{1J} \\
\vdots & \ddots & \vdots & \cdot^{\cdot^{\cdot}} & \vdots \\
a_{i1} & \cdots & a_{ij} & \cdots & a_{iJ} \\
\vdots & \cdot^{\cdot^{\cdot}} & \vdots & \ddots & \vdots \\
a_{I1} & \cdots & a_{Ij} & \cdots & a_{IJ}
\end{pmatrix} \equiv \mathrm{A} \tag{4.1}
$$

It is assumed that the amount of characteristics can be aggregated over goods (the utility of a characteristic does not depend on its origin) and the relationship is assumed to be linear which means that the relationship between goods purchased and characteristics obtained can be written as:

$$
z = \mathrm{A}'q \tag{4.2}
$$

The technology matrix *A* is constant over households which implies that all households meet the same *A* matrix and we assume it to be constant over the time span used in our model (in principle the *A* matrix can change over time as products with new and previously unknown characteristics get into the market). For each household we observe the quantity purchased of each good: $q_t^h = \left(q_{1t}^h, \ldots, q_{it}^h, \ldots, q_{It}^h\right)'$ and we also observe a unit price for each good in each period: $p_t^h = \left(p_{1t}^h, \ldots, p_{it}^h, \ldots, p_{It}^h\right)'$. The total expenditure by household *h* in period *t* is therefore $x_t^h \equiv \left(p_t^h\right)' q_t^h$. Knowing the technology matrix *A* and the amount of goods purchased we can calculate the amount of characteristics purchased.

Optimisation in general terms

The households have preferences over characteristics, and the purchased quantities of goods that we observe are a result of households maximising their utility given the technology, the prices and the budget. In each period the household therefore faces the problem:

$$
\begin{aligned}
&\underset{q_t^h}{Max} && U^h\left(z_t^h \middle| \Omega^h\right) \\
&s.t. && z_t^h = A'q_t^h \\
& && x_t^h \geq \left(p_t^h\right)' q_t^h \\
& && q_t^h \geq 0
\end{aligned} \tag{4.3}
$$

where Ω^h are socio-demographic characteristics and x_t^h is the total budget used by household *h* at time *t*. Note that the household optimises over goods *q*, but measures utility over characteristics *z*. This is because consumers purchase goods, but consume characteristics. The consumer's problem can be solved through Lagrange optimisation, assuming interior solutions and for a moment ignoring the socio-demographic characteristics. In a two good, two characteristic world this problem can be written as (the subscripts *h* and *t* are here suppressed due to ease of notation):

$$\max_{z,\lambda} L(z_1, z_2, \lambda) = u(z_1, z_2) - \lambda(p_1 q_1 + p_2 q_2 - x) \tag{4.4}$$

where λ is the utility value of increasing the binding constraint (the budget) dU/dx. If we, furthermore, substitute the technology into the budget restriction – which we assume to be binding – we get the following estimation problem:

$$\max_{z,\lambda} L(z_1, z_2, \lambda) = u(z_1, z_2) - \lambda(\pi_1 z_1 + \pi_2 z_2 - x) \tag{4.5}$$

where π_i is the implicit prices of the characteristics. The implicit prices π measure how much money the household is willing to pay for an extra unit of characteristic *j*, ($\pi = dx/dz$). If the *A* matrix is square and thereby invertible we can use the binding budget restriction to calculate the implicit prices of the characteristics directly by noting that the budget can be expressed both in actual prices of goods and implicit prices of characteristics:

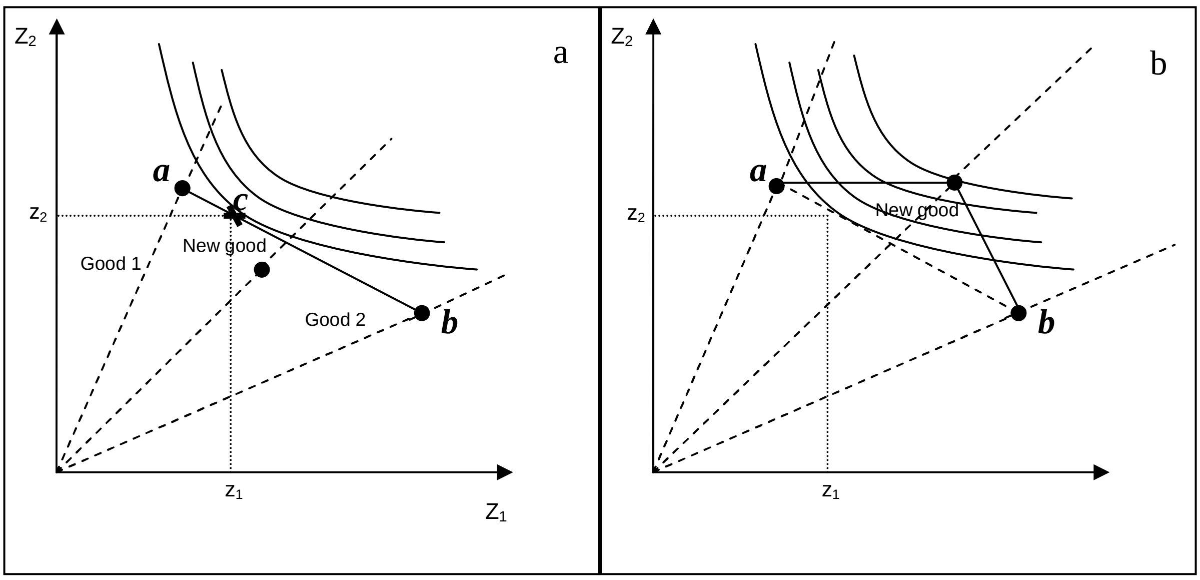
$$x = p'q = p'(\mathrm{A}')^{-1} z \tag{4.6}$$

$$x = \pi' z \Leftrightarrow \pi = \mathrm{A}^{-1} p \tag{4.7}$$

I.e. $\pi_1 = p_1\tilde{a}_{11} + p_2\tilde{a}_{12}$ and $\pi_2 = p_1\tilde{a}_{21} + p_2\tilde{a}_{22}$, where $\tilde{a}_{ij}$ are the elements in the inverse technology matrix. In this simple universe where the unit price is independent of the quantity, the implicit price of a characteristic is simply the monetary value of one unit of the characteristic. If there are more goods than characteristics the technology matrix is no longer invertible and the implicit prices have to be estimated through a hedonic price function.

In the world of two characteristics the consumers' problem can be shown visually. Knowing the prices *p* and the total amount spent[37] *x*, we can calculate the amount of each characteristic (z_1, z_2) that household *h* would obtain in period *t* spending all the money on good one (point *a* in Figure 4.1a below). If he spent all his money on good 2, he would obtain another amount of characteristics (point *b*). It is not possible to purchase characteristics outside the triangle $(a, b, 0)$ due to the technology restriction. On the Danish milk market it is not possible to purchase milk with less than 1 gram or more than 35 grams of fat per litre. We assume that all goods can be purchased in continuous quantities and the line between the highest obtainable level of characteristics (point *a* and point *b*) is the budget restriction. The continuous nature of the goods means that any linear combination of goods 1 and 2 is possible, e.g. point *c*. All three points lead to the same total cost. The consumers optimise where the marginal rate of substitution, MRS, is equal to the slope of the budget restriction, i.e. the point where the indifference curve for the highest attainable utility touches the boundary of the consumption set. When a new good, with known characteristics, but in new amounts, enters the market, the price of that good determines whether it will be purchased or not. In Figure 4.1a the price is too high (the consumer would get less of the characteristics z_1 and z_2 buying the new good) while in Figure 4.1b the price is so low that the budget constraint is pushed outwards and the consumers can obtain their preferred mix of characteristics in a cheaper way than by mixing good 1 and good 2.

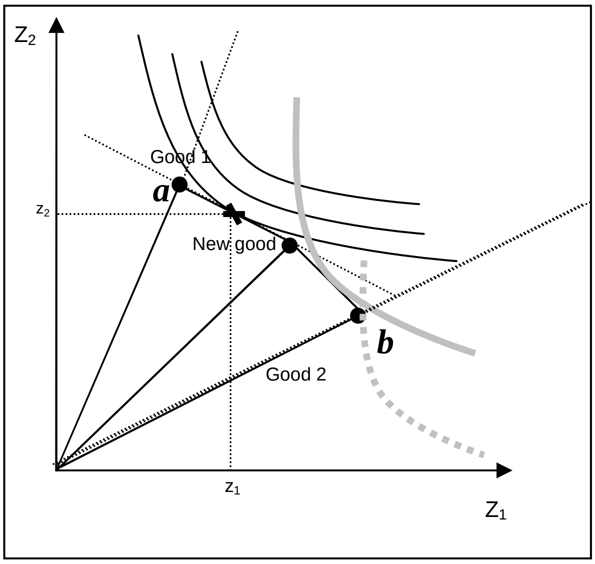
Figure 4.1: Consumers' optimisation problem in a two characteristics world



[37] In theory we need to know the amount available for consumption. However, this amount cannot be observed, so we have to assume that the budget constraint is binding and use the observed amount actually spent.

More goods exist in the world than are purchased by the individual household. For another household it might be more efficient to purchase a mix of the new good and good 2 as shown in Figure 4.2. It is not possible to buy goods outside the triangle consisting of zero and the lines running through *a* and *b* in Figure 4.2**.** This makes it difficult to point identify the parameters of the utility function for households who only purchase a good on the borderline, as e.g. the grey stipulated household in Figure 4.2. We will return to that later.

Figure 4.2: More consumers in a two goods, two characteristics world



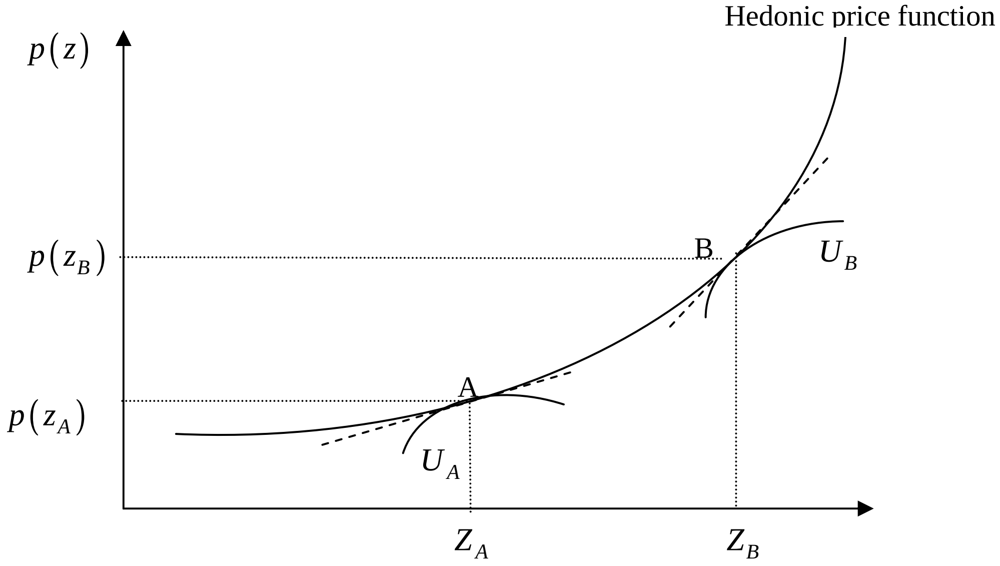
Estimation of implicit prices

Since we have more goods than characteristics we have to estimate the implicit prices using a hedonic price function, see e.g. Rosen (1974), Ladd and Zober (1977) or Ladd and Suvannunt (1976). In a world with *J* characteristics optimised over *I* goods, the Lagrange function (4.4) gives the following first-order conditions:

$$\frac{\partial L}{\partial q_i}: \quad \sum_{j=1}^{J} \frac{\partial u}{\partial z_j}\frac{\partial z_j}{\partial q_i} - \lambda p_i = 0 \Leftrightarrow \quad p_i = \sum_{j=1}^{J} \frac{\partial u}{\partial z_j} \cdot \frac{1}{\lambda} \tag{4.8}$$

The derivatives, $\frac{\partial z_j}{\partial q_i}$ are the elements in the technology matrix a_{ij}. The marginal utility of the budget $\lambda = \frac{\partial u}{\partial x}$ is assumed to be constant. This implies that we have to assume homothecity of the utility function. This assumption is not realistic for luxury goods or goods with a large share of total consumption, but more realistic for a normal good with a smaller share of total expenditure (like milk). Since $\frac{\partial U}{\partial z_j} \Big/ \lambda = \frac{\partial U}{\partial z_j} \Big/ \frac{\partial U}{\partial x}$ is equal to the marginal rate of substitution between the expenditure and the characteristics, this is equivalent to the marginal implicit price π_j of each of the characteristics $\left(\partial x / \partial z_j \right)$. This implies that the price of a good is a weighted sum of the implicit prices of its characteristics $p_i = \sum_j \pi_j \alpha_{ij}$, which is one of the most important features of the characteristics model. If $p_i \geq \pi_j \alpha_{ij}$ then good *i* is not bought as illustrated in Figure 4.1a. When implicit prices are used in a model estimating demand for characteristics there are several points to consider. Since one DKK spent on food will give you varying amounts of nutrients, dependent on which mixture of foods you choose to buy, the budget constraint in characteristics space is generally nonlinear. This leads to endogenous prices. However, at the optimal point where the indifference curve is a tangent to the budget constraint, the separating hyper-plane between these two loci is linear. In this optimal point and under the assumption of constant return to scale, prices can be assumed to be exogenous (Deaton and Muellbauer, 1980). Another problem is that consumers choose quantity and price simultaneously as illustrated in Figure 4.3. This means that the prices that equate the market depend on both the parameters that characterise demand and the distribution of the non-observable characteristics of demand (in the case where supply is not exogenous, as we assume here, the parameters characterising supply and the distribution of the non-observable characteristics of suppliers are also present in the hedonic price function). This means that the model is unidentified (Ekeland et al., 2004), the implicit prices provide no more information than the preferences originally used to estimate the implicit prices. Brown and Rosen (1982), Kahn and Lang (1988), Epple (1987) and Ekeland et al. (2004) suggest identification by allowing the price function to have higher powers of *z* (the characteristic) in the case of single market data or to use multi-market data to solve the identification problem. The main idea behind these identification strategies is that there must be additional parameters affecting the price functions that are not contained in the demand function. The multi-market identification approach, which is used here, builds on the assumption that the preference parameters and the distribution of tastes are identical across markets, but the price functions differ between markets, i.e. are affected by some additional variables not in the demand function. This implies different patterns of variance in different markets.

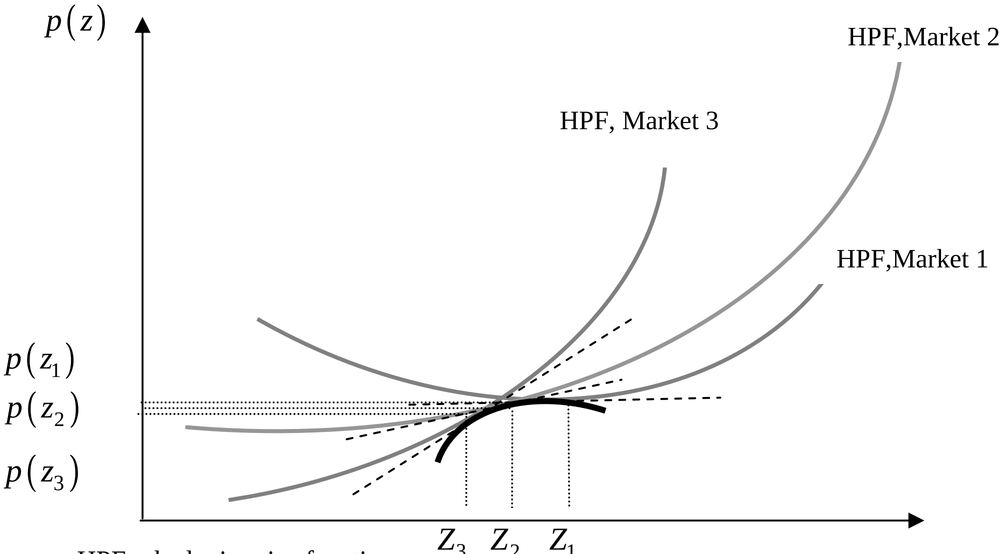
Figure 4.3: Simultaneous choice of price and quantity in the hedonic model*



* Adapted from Epple, 1987.

The identification of preferences from variation in the hedonic price functions are illustrated in Figure 4.4. Despite that the identification problems are solved in the multi-market case, a standard endogeneity problem persists, since the quantity and price of the characteristics are chosen simultaneously. This implies that the dependent variable (the chosen amount of the characteristic) and the implicit price are correlated through their dependence on the distribution of individual heterogeneity (Bartik, 1987; Kahn and Lang, 1988; Diamond and Smith, 1985).

Figure 4.4: Illustration of identification in the multi-market case



HPF = hedonic price function.

4.2 Data and the milk market

Purchase data and background data

In the empirical estimations we use a comprehensive panel dataset from GfK-Denmark (a marketing institute with branches all over the world). The data cover the period from 1997 to 2004 and include information about daily purchases for individual households. Additionally, a wide range of social and demographic questions about the households (income, location, media habits, favourite store etc.) and information about each individual in the household (BMI, exercise habits, education, age etc.) are posed annually. In principle, every time a household goes shopping the diary keeper reports the price and volume of each good and whether it is organic or conventional. For milk the data are reported at brand level. These purchase data are combined with nutrition data such as the content of fat, protein, calcium etc. for each type of milk. This means that whenever a household purchases milk, we know the equivalent bundle of nutrients purchased[38]. On average 2,500 households report their purchases on a daily basis which sums up to 10,500 weekly observations on purchases of milk. The milk purchase data are aggregated up to monthly observations in order to minimise the amount of zeros in the dataset. This also makes the inter-temporally separable model, which we use, more appropriate since milk is a non-durable good.[39] According to theory, a single consumer is only allowed to simultaneously purchase a number of goods corresponding to the number of characteristics. In a world with more goods than characteristics it becomes possible to violate this condition. If we observe households purchasing three types of milk at the same time, it means that there must be at least tree characteristics. If we aggregate data, we potentially violate this principle. It may be so that prices in one week make it optimal to combine skimmed milk with mini milk while the prices in another week make it optimal to combine mini and semi-skimmed milk. If these weeks are aggregated the result would suggest that the household purchased three types of milk simultaneously. The share of occasions where more than one type of milk is purchased increases significantly with the length of the aggregation period, but interestingly enough, the share of purchases of more than two types of milk remains relatively low (less than 5 per cent), so we choose to ignore the problem in this paper. Households that only buy one type of milk constitute another problem in the data since that gives little or no information about preferences. Less than 2 per cent always buy only one

[38] For a throughout description of the data see Smed (2008)

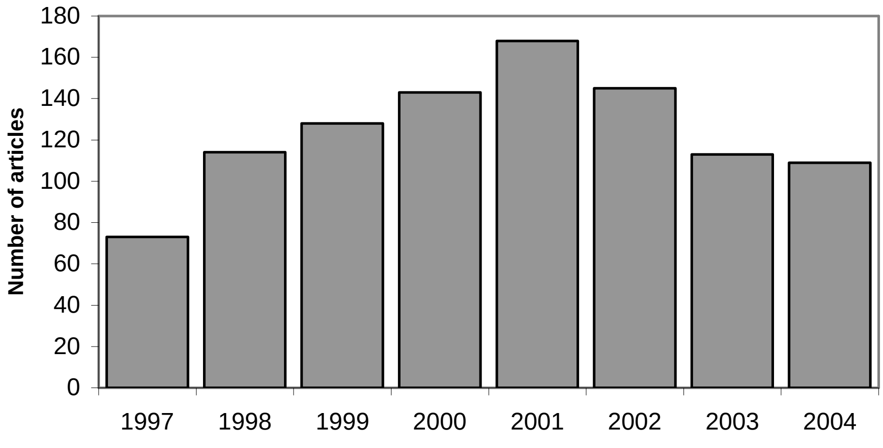
[39] Milk will only keep fresh for a little longer than a week The market for UHT milk is minimal in Denmark and almost all households buy and consume fresh milk.

type of milk per month, while 61 per cent mix different types of milk in more than 30 per cent of the months we observe we observe the household

Information data

Consumers receive information about the connection between health and the intake of fat through various channels. This includes the internet, face-to-face conversations, television and newspapers. As it is not possible to capture all these diverse types of information most studies incorporating the effect of health information on food demand use proxies to account for the amount of information that consumers receive. Some studies use the number of published medical articles mentioning a link between intake of a special nutrient and health (e.g. Brown and Schrader, 1990; Kinnucan et al., 1997; Chang and Kinnucan, 1991; Chern and Zuo, 1995; Kim and Chern, 1997, 1999). The basic assumption behind these indices is that the information in these articles is transmitted down to the consumer through various means, e.g. newspapers and television. A more direct approach uses the number of relevant newspaper articles and/or the number of television transmissions (e.g. Piggott and Marsh, 2004; McGuirk et al., 1995; Schmidt and Kaiser, 2004; Verbeke and Ward, 2001; Smith et al., 1988). The direct approach is used here as the number of articles mentioning a link between the intake of fat and health are collected from Danish newspapers. The search is done in Infomedia.[40] The basic search words are fat/fat-rich/low fat in connection with health, slim, overweight, obesity resulting in 12 different combinations of searches. Figure 4.5 shows the number of hits for fat. The number of articles is steadily increasing until 2001 and then the number of articles decrease.

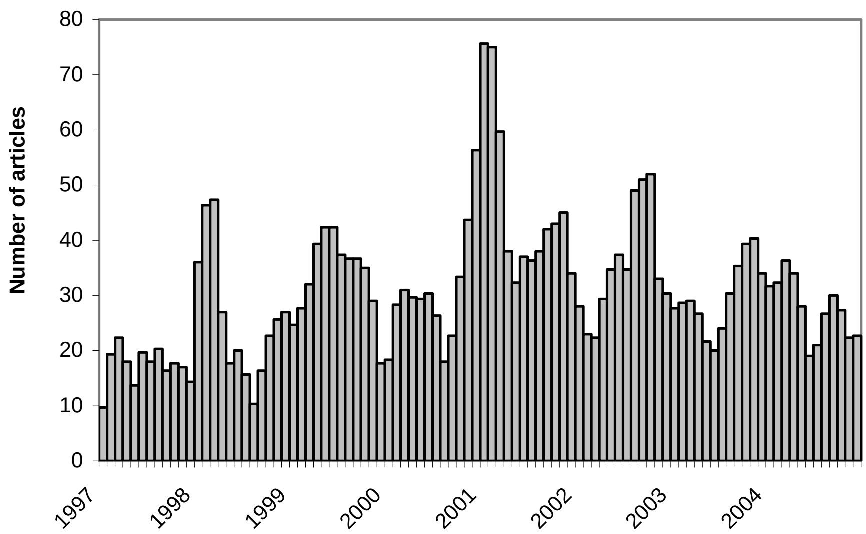
Figure 4.5: Absolute number of hits in newspapers about the link between consumption of fat and health



[40] Infomedia is a database collecting articles from all Danish newspapers.

The articles are aggregated over newspapers independently of the size or location of the article. Several of the indices introduced in the literature use a lag structure, as they find that press coverage have a cumulative effect. This includes simple cumulative indices as in McGuirk et al. (1995) and Schmit and Kaiser (2004), declining shares to lagged index values as in Rickertsen et al. (1995) or more sophisticated structures as in Verbeke and Ward (2001). Based on the literature we choose to let the information last for a three-month period.[41] As we have aggregated the data to monthly observations the information that arrives at the end of the month will have a larger influence in the next month than the current month. Therefore, we construct a floating index from the original newspaper articles where each article is allowed to last for three months. This gives the information loads in each month presented in Figure 4.6.

Figure 4.6: Number of hits in newspapers per month, three months floating index



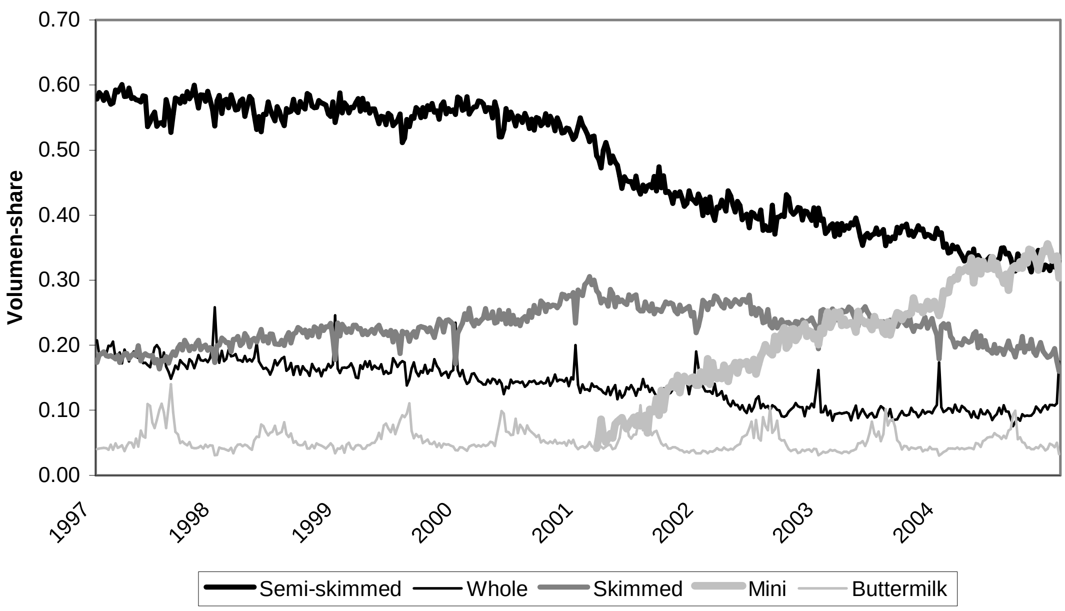
The milk market

Until February 2001, there were four major types of milk on the Danish market: Whole milk, semi-skimmed milk, skimmed milk and buttermilk. Whole milk has a fat content of 3.5 per cent, semi-skimmed milk of 1.5 per cent, skimmed milk and butter milk has a fat content of 0.1 per cent. Furthermore, buttermilk is soured. There has been a steady decrease in the consumption of whole milk since the introduction of semi-skimmed milk in 1972. This decrease has been accompanied by an increase in the consumption of semi-skimmed milk until the early 1990s (Statistics Denmark, 2008), where the Danish authorities' general

[41] We have also tried a cummulative structure with no decay and a current index with no lags and the three-month structure shows the best result. More sophisticated analyses of the lag structure will be a route of further research.

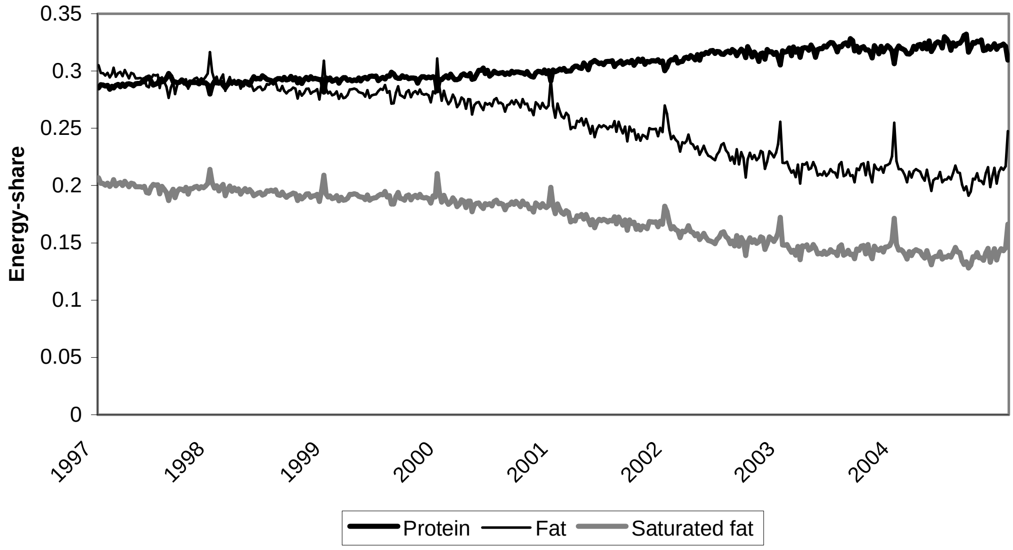
campaigns concerning fat intake were initiated. These campaigns affected the milk market by increasing demand for skimmed milk and decreasing the demand for semi-skimmed milk, as illustrated in Figure 4.7. On the other hand, the increased demand for low-fat food inspired development of new low-fat varieties of milk. In February 2001, a new type of milk (mini milk) was introduced on the Danish market. This new type of milk targets consumers, who wants a product that tastes like semi-skimmed milk, yet has almost the low fat content of skimmed milk. Mini milk has a fat content of 0.5 per cent compared to the 1.5 per cent in semi-skimmed milk. This new type of milk took over part of the market for semi-skimmed milk and reversed the increasing trend for skimmed milk, while the trends for whole milk and buttermilk were almost unaffected as it is evident from Figure 4.7. The December peaks for whole milk is due to traditional eating during Christmas, while the summer peaks for buttermilk is due to another traditional dish called "Koldskål" eaten on (especially warm) summer days.

Figure 4.7: The Danish milk market, January 1997 to December 2004



During the rest of this analysis we will not take the consumption of buttermilk into account, mainly because it is soured and therefore the use of this type of milk is rather different from the use of the other types of milk. The total volume of milk purchased in the same period has been more or less stable. As explained above the purchase data are combined with nutrition data making it possible to follow the consumption of different nutrients over time. Figure 4.8 shows the development in the energy share of total fat, saturated fat and protein from milk from January 1997 to December 2004. The share of fat consumed in milk has been declining, especially after the introduction of mini milk in February 2001. The systematic peaks in December each year is due to the increased consumption of whole milk around Christmas.

Figure 4.8: The purchase of nutrients



In Smed (2005) and Smed and Jensen (2004) price elasticities for milk were estimated at an aggregate level both before and after the introduction of the new low fat type of milk. These elasticities show that before the introduction of the new type of milk whole milk and semi-skimmed milk was substitutes, which was also the case for semi-skimmed and skimmed milk. After the introduction of the new low fat milk there is no longer any substitution between semi-skimmed milk and skimmed milk, while semi-skimmed is a substitute to mini milk.

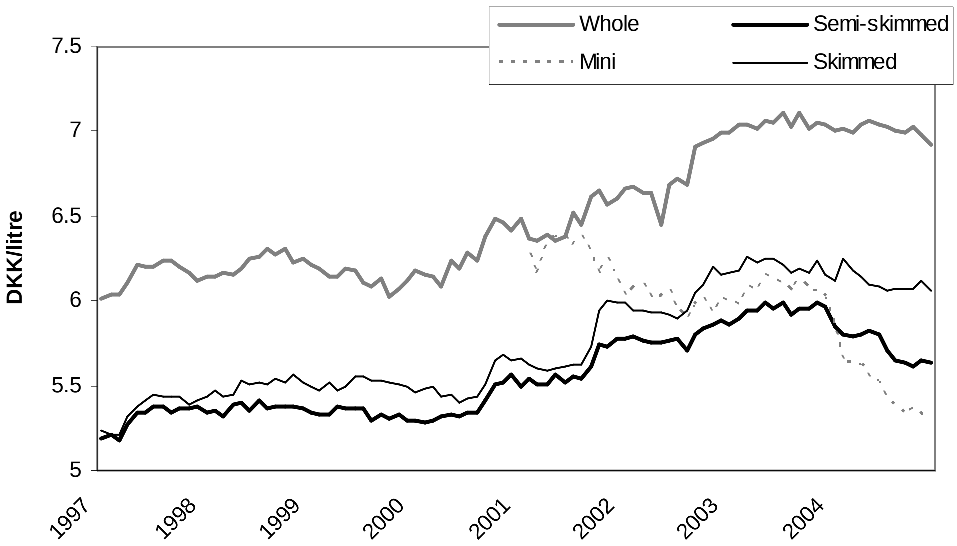
Table 4.1: Price elasticities before and after the introduction of mini milk

	Whole milk	Semi-skimmed milk	Skimmed milk	Mini milk
January 1997 to February 2001				
Whole milk	-1.45	0.12	0.00	-
Semi-skimmed milk	0.30	-1.16	0.36	-
Skimmed milk	0.00	0.16	-1.00	-
September 2001 to September 2002				
Whole milk	-1.44	0.32	0.06	0.06
Semi-skimmed milk	0.78	-1.68	0.03	0.74
Skimmed milk	0.00	0.00	-1.00	0.00
Mini milk	-0.01	0.30	0.00	-2.06

Source: Smed (2005) and Smed and Jensen (2004).

According to the characteristics model consumers mix their consumption of different types of milk to gain the optimal amount of fat. Before mini milk a fat content between 0.1 per cent and 1.5 per cent could only be obtained by consuming both skimmed and semi-skimmed milk. After the introduction of mini milk consumers who follow the characteristics model will either mix skimmed and mini milk, or mini and semi-skimmed, which is exactly what happened. The estimated change in elasticities indicates that the market for milk is probably correctly described by a characteristics model. Figure 4.9 shows average prices from January 1997 to December 2004. Until just before the introduction of mini milk prices have been rather stable with an average price of whole milk well above the other and semi-skimmed milk as the cheapest. The prices of the "old" milk types increased just before the introduction of mini milk in 2001 and this continued until the end of 2003, meanwhile the price of mini milk decreased. In 2004 all prices declined which might be due to a price war on milk initiated by one of the larger retail chains and the introduction of discount milk. This milk does not exist in a whole milk version which might be the reason why the price of whole milk did not decline along with the price of the other types of milk. The introduction of German milk in the supermarkets also forced prices down.

Figure 4.9: Development in average milk prices



In the following figures the consumption of fat in milk for different types of households is described. Figure 4.10 illustrates the development in average grams of fat per litre of milk for households where the head has different level of education.[42] In 1997 two types of households distinguish themselves by consuming milk with a large fat content. These are households where the head has no further education or has vocational education. Households where the head has a longer education consume milk with a lower fat content. This has changed, in 2004 it is those with no further education and the longest educated who consume milk with a high fat content.

Figure 4.10: Fat per litre of milk for households with different education

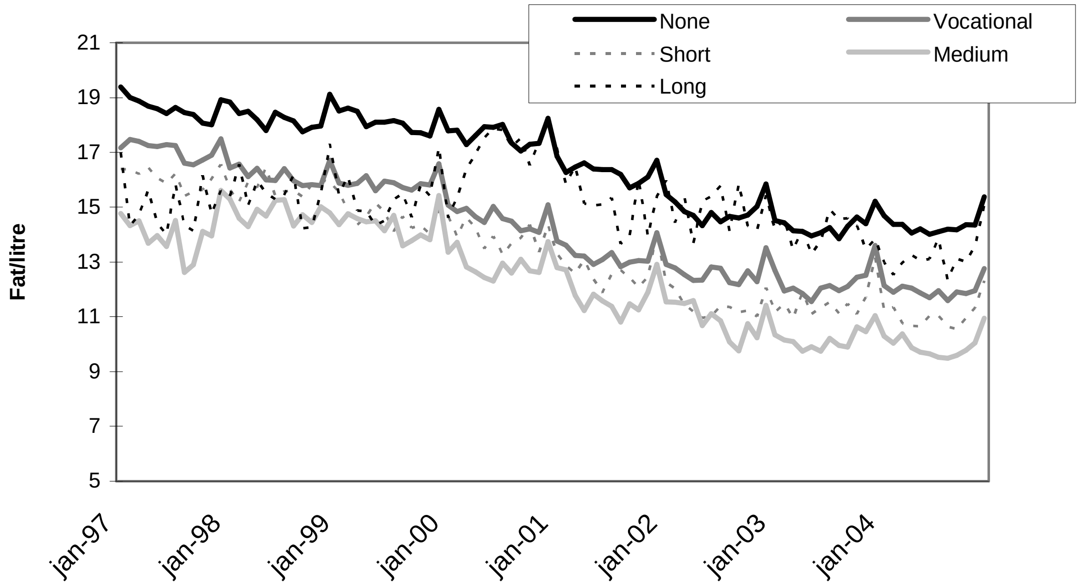


Figure 4.11 illustrates the development in average grams of fat per litre of milk for different family types. In 1997 households with children between the age of 0 and 3 distinguish themselves by consuming milk with more fat per litre than other households. Families with older children seem to prefer a more moderate amount of fat per litre. In 2004 this picture has changed since households with small children no longer distinguish themselves. This might be because small children in 2004 no longer are recommended to drink whole milk, but instead are encouraged to drink semi-skimmed milk. In 2004 households with no children consume the fattiest type of milk. Even though households with no children consume the fattiest type of milk, they consume less milk so households with children 0-3 years of age still consume most fat in grams per person per week. The peaks around Christmas are clearer for households with no children than for other types of households.

---

[42] Vocational oriented education is e.g. carpenter, nursing aide; short further education is e.g. policeman; technical education; medium further education is e.g. nurse, school teacher, while long further education is e.g. a university degree.

Figure 4.11: Fat per litre of milk for different family types

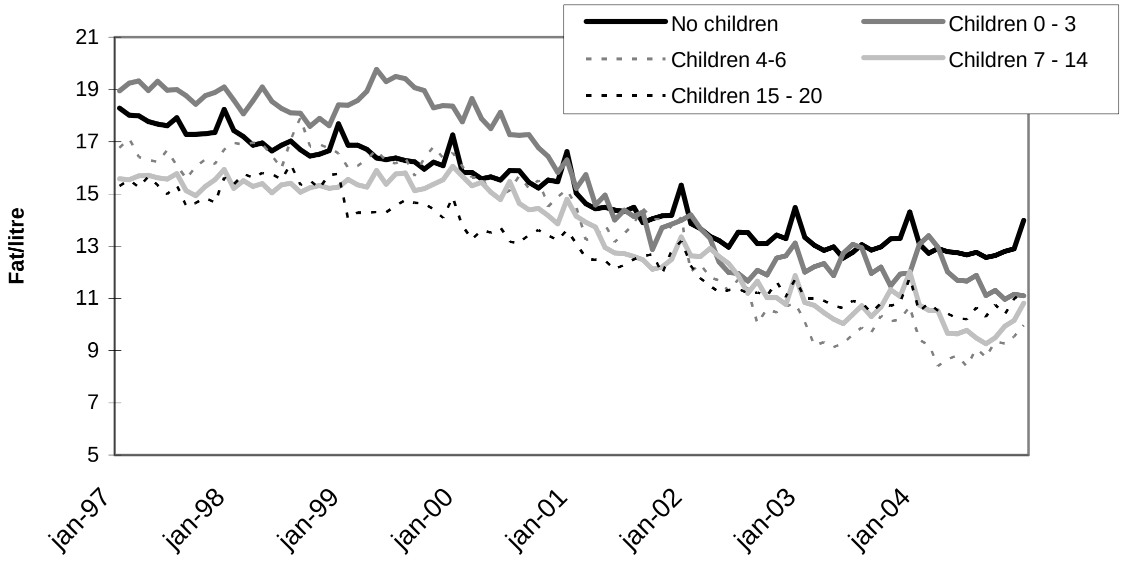
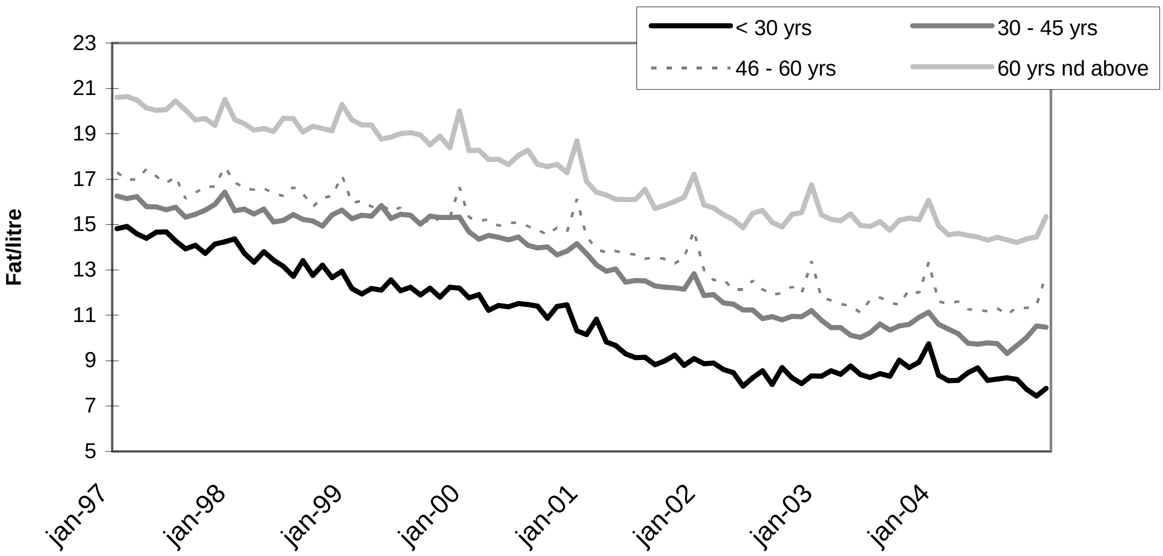


Figure 4.12 illustrates the development in average grams of fat per litre of milk for households in different age groups. In general, older households consume more fatty milk than other households. Younger people below the age of 30 consume milk with the lowest fat content. As they have a moderate consumption of milk this implies that they get the smallest amount of fat in grams per person per week compared to other age groups. The Christmas peaks are most clear among households above 45 and are almost non-existing for households below 30.

Figure 4.12: Fat per litre of milk for households in different age groups



4.3 Empirical considerations and estimation

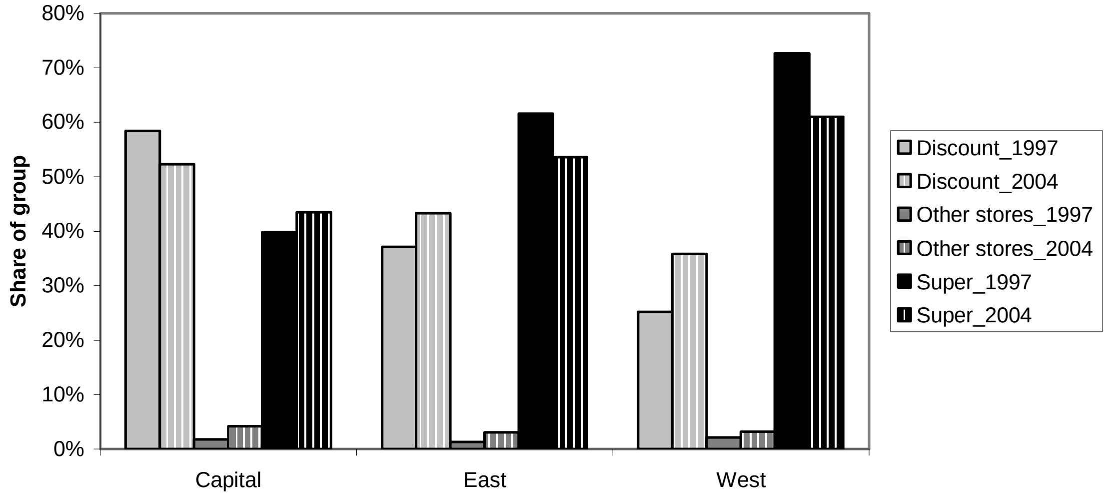
We take prices as given for the individual households, and thereby focus on the demand side. This is equivalent to the approach in Muellbauer (1974) and Blow et al. (2005) and opposite Rosen (1974) who focuses on both the demand and supply side. The comprehensive dataset that we use allows us to follow individual households over a very long time (up to eight years) so we can deal with individual heterogeneity in the most extreme way by estimating the model individually for each household. We concentrate on the four main types of milk, whole milk, semi-skimmed milk, mini milk and skimmed milk. All these types of milk exist in both a conventional and an organic version. Milk is assumed to consist of two characteristics: milkiness and fat. Milkiness is best explained as the characteristic that distinguishes milk from a mixture of calcium and water, i.e. the fact that you can use it in your coffee, use it in pastry or on your cereals etc. One unit of milk contains one unit of milkiness independently of the type of milk, i.e. milkiness is measured in litres.

Estimation of prices

We estimate a hedonic price function for several markets (different stores and different modes of produce) using observed purchases from all consumers. Demand is then estimated for the households individually assuming that the household visits several markets i.e. go into different kinds of stores and buy both conventional and organic milk. This ensures identification, since parameters that do not influence the demand function for the individual consumer, namely other consumers' preferences, influence the hedonic price function. As our consumer only to a minor degree contributes to each particular hedonic price function, prices can be assumed to be exogenous. Furthermore, the usual problem of endogeneity does not apply since each consumer's demand function is estimated individually. We assume that supply is given exogenously, which is reasonable in the market for foods since the individual consumer's decision cannot affect suppliers in the hedonic model for milk. It is assumed that there are three types of stores: discount stores, supermarkets and other shops.[43] Furthermore, the country is divided into three regions: capital area, east and west since it is assumed that the price of milk depends on which part of the country it is bought in. Figure 4.13 shows the share of milk bought in each kind of store in different regions. In the capital the share of milk bought in discount stores has been declining while the share bought in supermarkets has increased a little. It is the opposite in east and west Denmark.

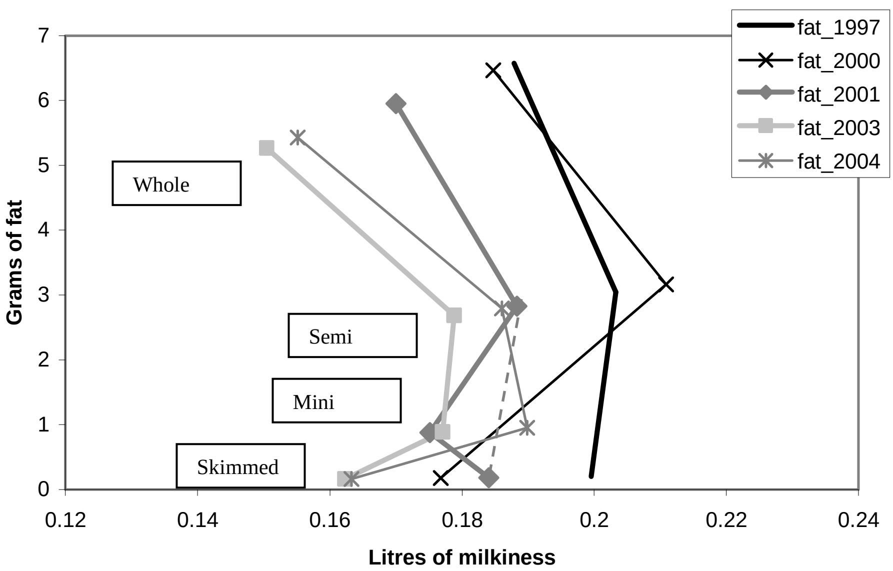
[43] Other stores are bakeries, gas stations etc.

Figure 4.13: Share of milk bought in each kind of store in each region



The following figures show how much of each of the characteristics fat and milkiness you get if you use one DKK on a particular type of milk, i.e. this is the empirical version of the theoretical Figure 4.1. In 1997 one DKK used on skimmed milk provided 0.2 units of milkiness and 0.2 units of fat, while one DKK used on whole milk provided only 0.19 units of milkiness, but 6.6 grams of fat. In 1997 and 2000 (1998 and 1999 are removed due to the clarity of the figure) the consumption set consists of only three points (skimmed, semi and whole milk), while the consumption sets in the other years have four points due to the entrance of mini milk on the market. (2002 is removed due to the clarity of the figure)

Figure 4.14: The empirical consumption set, capital, discount, conventional, standard dairy



In 2001 conventional mini milk is too expensive (the efficient consumption set is indicated by the stipulated grey line) and the consumers should not actually be buying it. That they do it anyway might be due to that the product is new on the market and has been marketed rather heavily. Similar consumption sets can be constructed for the other markets.
Figure 4.15 shows the average price for different types of organic and conventional milk produced at a standard dairy and bought in different regions in either supermarkets or discount stores in 2003.[44] From the figure it is clear that there are nonlinear relations between the price and the fat content. This nonlinear connections seem to be different dependent on whether the milk is conventionally or organically produced hence the price function for fat is different for organic and conventional milk, respectively. Together this implies that we have 18 different markets (3 types of stores, 3 regions and two modes of produce).

Figure 4.15: The average price for milk in various stores and regions, standard dairy

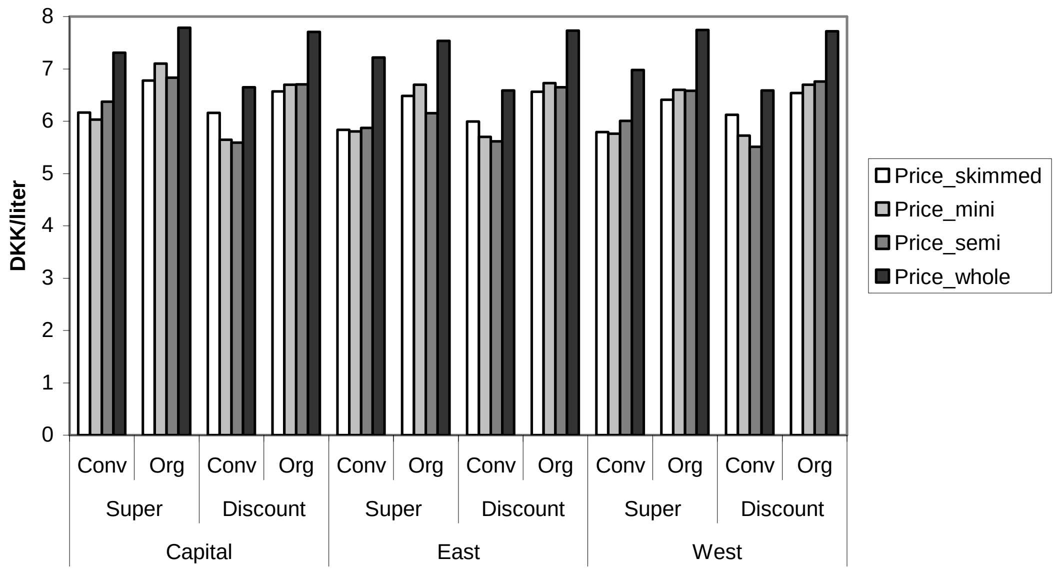
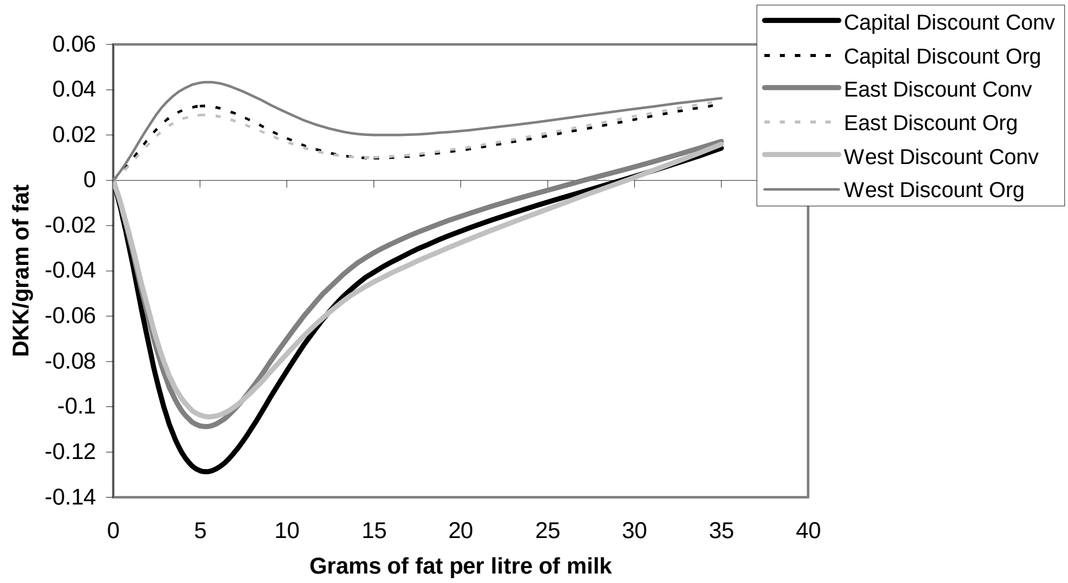


Figure 4.16 is a crude illustration of the hedonic price function for fat illustrated for selected markets. The figure is used to illustrate the motive behind choosing a quadratic form for the hedonic price function and separate markets for organic and conventional. The figure is crude in the sense that the average price of milk is used so the figure does not take into account the distribution of consumer preferences. Skimmed milk is the basis and the price of skimmed milk is assumed to reflect the price of milkiness (i.e. the amount of fat in skimmed milk is set to 0 in these figures, which also is a simplification, in the estimations skimmed milk contains

[44] Other stores are left out of the figure. They have a rather small share of the market, in 2004 less than 3% in each region.

1 gram of fat per litre of milk). The price of fat is then calculated as the difference between the price of the milk in question and the price of skimmed milk since all milk is assumed to contain the same amount of milkiness.

Figure 4.16: A crude empirical hedonic price function for fat, year 2003, standard dairy



In the demand model we treat preferences for milk as separable from all other food, which of course is questionable as is all separability assumptions. Furthermore, we treat preferences for milkiness and fat as separable from the mode of produce (organic or conventional) and dairy (standard, discount or luxury dairy). As it appears from Figure 4.16 the hedonic price function for organic and conventional milk differs, but the hedonic price function for fat is unaffected by the dairy (not shown in the figure). This implies that mode of produce is treated as a separate market, while dairy appear as a dummy within the hedonic price equation. This means that 18 different versions of the hedonic price equation (4.9) are estimated, one for each market.

$$p_{i,t} = \beta_{milkiness,t} + \beta_{small\_dairy,t} \cdot D_s + \beta_{discount\_dairy,t} \cdot D_d + \beta_{fat,t} \cdot z_{fat,t} + \beta_{fat\_sq,t} \cdot \left(z_{fat,t}\right)^2 + \varepsilon_{it} \quad (4.9)$$

The constant accounts for the price of one litre of "milkiness", D_s and D_d, are dummies accounting for a luxury and discount dairy, respectively,[45] z_{fat} accounts for the content of fat in grams. The polynomial of second order implies that the price of fat varies with the type of

[45] The base is here a standard dairy. Discount dairies are mainly milk from foreign dairies, store brands etc. The lukury dairies are local or speciality dairies.

milk; as illustrated in Figure 4.16 it is more expensive to get your fat from whole milk than from semi-skimmed milk. The parameters from this estimation result in a monthly implicit prices of characteristics, one for each market, equivalent to the two shown in Figures 4.17 and 4.18. As an example for the organic market in supermarkets in the capital in January 1997 the price of whole milk is equivalent to the price of milkiness, 6.82 DKK plus $0.0265 \cdot 35 \approx 0.93$ DKK for fat in whole milk, i.e. in total 7.75 DKK. To compare, the same milk can be purchased for 5.13 DKK plus $0.0295 \cdot 35 \approx 1.03$ equal to 6.16 DKK at the conventional market.

Figure 4.17: Hedonic prices for organic milkiness and fat, supermarkets in the capital

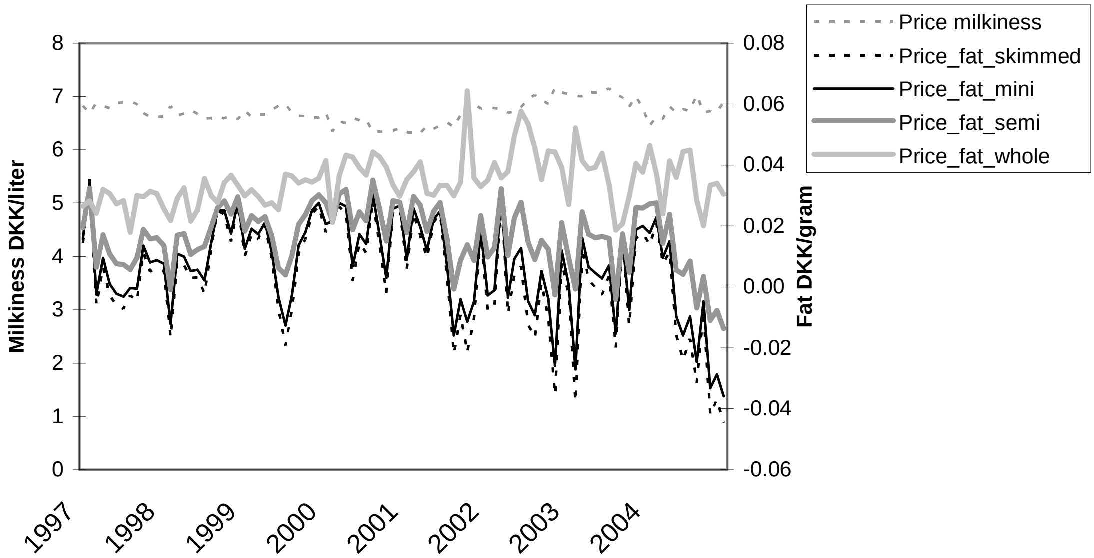
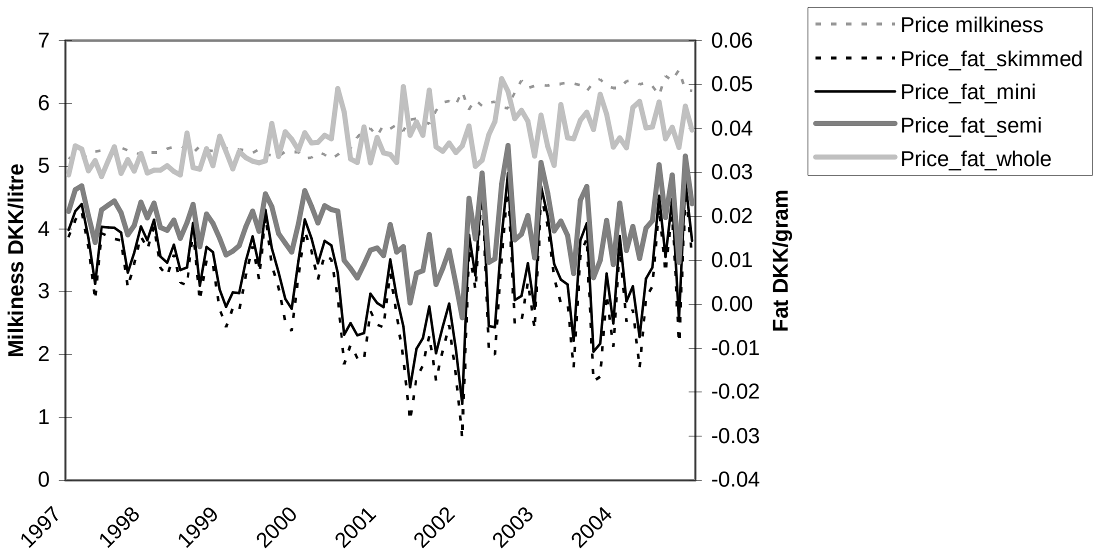


Figure 4.18: Hedonic prices for conventional milkiness and fat, supermarket in the capital

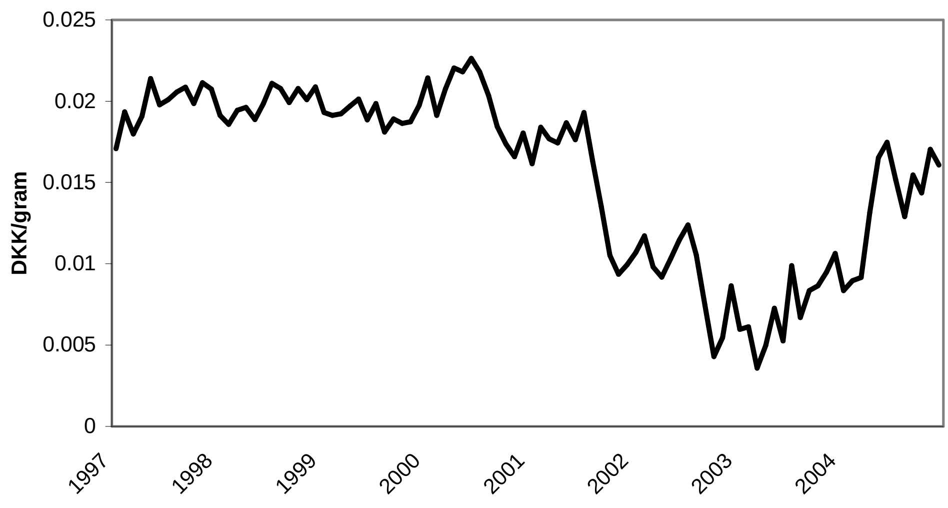


To construct prices for each household the estimated implicit prices at each market are weighted according to actual purchase patterns at either the organic or the conventional market and in the three different stores.[46] For example, imagine a household living in east Denmark who only consumes whole and semi-skimmed milk buys in a particular month some of their organic semi-skimmed milk in supermarkets and some in discount stores. All their conventional whole milk is bought in other shops. The price per gram of fat will then be calculated as:

$$p_{fat=} \left(fat_{semi,\sup er} \cdot p_{semi\_fat,org,\sup er} + fat_{semi,disc} \cdot p_{semi\_fat,org,discount} + fat_{whole,conv} \cdot p_{whole\_fat,other} \right) \Big/ totfat \quad (4.10)$$

The weighted price paid for fat over time averaged over households is shown in Figure 4.19 (the price multiplied with the amount of fat bought per litre of milk). The fall in the value of fat from the end of 2001 to the middle of 2002 might be initiated by a fall in the price of mini milk relative to the price of skimmed milk as shown in Figure 4.9. The changes in prices have been accompanied by a general movement towards leaner types of milk (see Figure 4.7) which also adds to the lower price paid for fat. The price of fat is much higher for fat from whole milk (Figures 4.17 and 4.18) and the movement away from whole milk therefore decreases the price paid for fat per litre of milk. In the middle of 2002 the share of whole milk stabilised, and the price of fat from conventional whole milk started to increase. The combination of these two factors may be the reason for the increase in the mean price paid for fat per litre of milk from 2003 and onwards.

Figure 4.19: Mean price paid for fat per gram



[46] We assume that the consumer only buys milk in his own region.

The choice of the quadratic utility function and where to put the error term

The quadratic utility model is characterised by having a point with maximum utility and the possibility of negative marginal as well as absolute utility of characteristics. This makes sense when estimating a model for characteristics. Free disposal is usually possible for goods, but not always for characteristics. It is not possible to dispose of fat without disposing of milkiness, and a positive utility of milkiness may outweigh a negative absolute utility of fat. In one version of the model we assume that we possibly do not observe everything perfectly; a household may in some periods like a characteristic more than in others due to influence from non-systematic (or non-observable reasons). We therefore include a time-specific random error with mean 0 for each characteristic.

$$U(z) = z'(\alpha + \varepsilon) - 0.5 \cdot z'\beta z\,, \quad \varepsilon \sim N(0, \Sigma) \tag{4.11}$$

The derivative of the utility in (4.11) with respect to characteristics is then:

$$\frac{\partial U}{\partial z} = (\alpha + \varepsilon) - \beta z \tag{4.12}$$

Disregarding technology and goods, the first order conditions from the Lagrange equation, leads to the following demand function (see appendix 4A for derivation).[47]

$$z = \beta^{-1}(\alpha + \varepsilon) - \left(\beta^{-1}\pi\left(\pi'\beta^{-1}\pi\right)^{-1}\right)\left(\pi'\beta^{-1}(\alpha + \varepsilon) - x\right) \tag{4.13}$$

This result has a fine intuitive interpretation. Note that:

$$\frac{\partial U}{\partial z} = (\alpha + \varepsilon) - \beta z = 0 \Leftrightarrow z = \beta^{-1}(\alpha + \varepsilon) \tag{4.14}$$

the first part of (4.13) is therefore the consumption that would be chosen if there was no budget restriction. The last part of (4.13) is:

$$\pi'\beta^{-1}(\alpha + \varepsilon) - x \tag{4.15}$$

This is the difference in price between the optimal consumption from (4.14) and the actual budget *x*. If the budget is binding the price of the optimal consumption is higher than the budget, which means that the consumption is lower than the optimal level in a world without budget constraint. This can be seen directly from (4.13) (as long as prices are positive).

[47] Theil (1971) optimises the utility function without the error term in the utility function.

The middle term in (4.13) is

$$\beta^{-1}\pi\left(\pi'\beta^{-1}\pi\right)^{-1} \tag{4.16}$$

This term creates the link between the budget, the prices and the actual consumption. This is an interior solution, which means that we ignore the fact that characteristics cannot always be combined just as the consumer would prefer. A brief look at this demand function demonstrates the problems that are involved in obtaining independent estimates of β and α. The usual way of approaching the problem is to acknowledge that the world offers other types of goods than the goods in question (here milk) and a simple way of including other goods is to include a linear term which represents all other goods (or all other types of food). With a linear term the quadratic utility function becomes quasi linear which results in linear demand curves (Gravelle and Rees, 1992). This gives some restrictions in relation to the optimal consumption of milk since the optimal consumption is where the marginal utility of milk equals the marginal utility of other goods (or foods) which is assumed constant. This also implies that there is no income effect for milk and the marginal utility of money is constant. As we assume this to be unrealistic we use another approach exemplified in the equations below. A trend is introduced in the model in order to catch up with changes in preferences over time. β is assumed to be a diagonal matrix (a matter of convenience). The trend is made exponential (a matter of empirical evidence) and added to the alpha parameter, but is not assumed to be a part of the normalisation of the alphas (the alphas is assumed to sum to one). These decisions are based on empirical evidence through repeated reformulations of the model. In a two characteristics world equation (4.11) looks like:

$$\begin{aligned} U(z) &= z'\left(\alpha+\varepsilon+\tau\ln(t)\right)-0.5z'\beta z \\ &= \left(\alpha_1+\varepsilon_1+\tau_1\ln(t)\right)z_1+\left(\alpha_2+\varepsilon_2+\tau_2\ln(t)\right)z_2-0.5\left(\beta_1 z_1^2+\beta_2 z_2^2\right) \end{aligned} \tag{4.17}$$

which means that in optimum we have:

$$\frac{\partial U/\partial z_1}{\partial U/\partial z_2}=\frac{\left(\alpha_1+\varepsilon_1+\tau_1\ln(t)\right)-\beta_1 z_1}{\left(\alpha_2+\varepsilon_2+\tau_2\ln(t)\right)-\beta_2 z_2}=\frac{\pi_1}{\pi_2} \tag{4.18}$$

rearranging leads to:

$$\left(\alpha_1+\varepsilon_1+\tau_1\ln(t)\right)\pi_2-\beta_1 z_1\pi_2=\left(\alpha_2+\varepsilon_2+\tau_2\ln(t)\right)\pi_1-\beta_2 z_2\pi_1 \tag{4.19}$$

which can be further reduced to:

$$\varepsilon_1 + \varepsilon_2 \frac{\pi_1}{\pi_2} = \left(\alpha_2 + \tau_2 \ln(t)\right)\frac{\pi_1}{\pi_2} - \left(\alpha_1 + \tau_1 \ln(t)\right) + \beta_1 z_1 - \beta_2 \frac{\pi_1}{\pi_2} z_2 \tag{4.20}$$

If we use the fact that

$$x = \pi_1 z_1 + \pi_2 z_2 \Leftrightarrow z_1 = \frac{x - \pi_2 z_2}{\pi_1} \tag{4.21}$$

and substitute z_1 into equation (4.20) it becomes:

$$\varepsilon_1 + \varepsilon_2 \frac{\pi_1}{\pi_2} = \left(\alpha_2 + \tau_2 \ln(t)\right)\frac{\pi_1}{\pi_2} - \left(\alpha_1 + \tau_1 \ln(t)\right) + \beta_1 \frac{x}{\pi_1} - \beta_1 \frac{\pi_2}{\pi_1} z_2 - \beta_2 \frac{\pi_1}{\pi_2} z_2 \tag{4.22}$$

If we normalise the alphas to sum to one in each period $\alpha_1 + \alpha_2 = 1$ and $\varepsilon_2 = 0$ and re-introduce the household specific notation we get:

$$\varepsilon_{1t}^h = \left(\alpha_2^h + \tau_2^h \ln(t)\right)\frac{\pi_{1t}^h}{\pi_{2t}^h} - \left(1 - \alpha_{2h} + \tau_1^h \ln(t)\right) + \beta_1^h \frac{x_t^h}{p_{1t}^h} - \beta_1^h \frac{\pi_{2t}^h}{\pi_{1t}^h} z_{2t}^h - \beta_2^h \frac{\pi_{1t}^h}{\pi_{2t}^h} z_{2t}^h \tag{4.23}$$

Demand can also be expressed much simpler in an *m*-demand version (Browning, 1999), which implies that demand for one good is expressed as a function of demand of a reference good, here milkiness. As long as the reference good is normal this is a satisfactory measure of utility conditional on prices. This means that with the same restrictions as above (4.23) can be expressed as:

$$\varepsilon_{1t}^h = \left(\alpha_2^h + \tau_2^h \ln(t)\right)\frac{\pi_{1t}^h}{\pi_{2t}^h} - \left(1 - \alpha_2^h + \tau_1^h \ln(t)\right) + \beta_1^h z_{1t}^h - \beta_2^h \frac{\pi_{1t}^h}{\pi_{2t}^h} z_{2t}^h \tag{4.24}$$

Above, we have assumed that consumers experience random shifts in preferences. If we instead assume that changes in preferences are systematic, the random part of alpha disappears and instead we assume that we do not measure consumption perfectly, a random term is added to the *z*'s. The random terms on the *z*'s are connected by the budget:

$$x = \pi_1\left(z_1 + \xi_1\right) + \pi_2\left(z_2 + \xi_2\right) \Leftrightarrow \xi_1 = \frac{x - \pi_1 z_1 - \pi_2\left(z_2 + \xi_2\right)}{\pi_1} \tag{4.25}$$

and we can therefore only identify one error term. We choose to assume that milkiness is observed perfectly, but fat is observed with uncertainty. Then (4.23) becomes:

$$0=\left(\alpha_2^h+\tau_2^h\ln(t)\right)\frac{\pi_{1t}^h}{\pi_{2t}^h}-\left(1-\alpha_2^h+\tau_1^h\ln(t)\right)+\beta_1^h\frac{x_t^h}{\pi_{1t}^h}-\beta_1^h\frac{\pi_{2t}^h}{\pi_{1t}^h}\left(z_{2t}^h+\xi_{2t}^h\right)-\beta_2^h\frac{\pi_{1t}^h}{\pi_{2t}^h}\left(z_{2t}^h+\xi_{2t}^h\right)$$

(4.26)

and the *m*-demand version (4.24) becomes:

$$0=\left(\alpha_2^h+\tau_2^h\ln(t)\right)\frac{\pi_{1t}^h}{\pi_{2t}^h}-\left(1-\alpha_2^h+\tau_1^h\ln(t)\right)+\beta_1^h z_{1t}^h-\beta_2^h\frac{\pi_{1t}^h}{\pi_{2t}^h}\left(z_{2t}^h+\xi_{2t}^h\right) \quad (4.27)$$

In the classical demand functions we know that the budget is endogenous and in the *m*-demand versions that z_1 is endogenous due to the correlation between milkiness and fat through the budget. In the budget version the budget version is instrumented by the total budget for milk (i.e. milk including buttermilk, chocolate milk, milk with taste etc.) and in the *m*-demand versions we choose to instrument by the lagged value of milkiness and the total budget for milk. The instrumentation is done for each household individually:

$$z_{1t}^h=\eta_1^h z_{1t-1}^h+\eta_2^h\tilde{x}_t^h+\zeta_{1t}^h \Rightarrow \hat{z}_{1t}^h=\eta_1^h z_{1t-1}^h+\eta_2^h\tilde{x}_t^h \quad (4.28)$$

where z_{1t-1}^h is the lagged value of z_{1t}^h and $\tilde{x}_t^h$ is the budget for purchases of all types of milk. We include both the estimated value $\hat{z}_{1t}^h$ and the residual in the estimations; this is called the control function approach (Blundell and Powel, 2003). Equation (4.24) then changes to:

$$\varepsilon_{1t}^h=\left(\alpha_2^h+\tau_2^h\ln(t)\right)\frac{\pi_{1t}^h}{\pi_{2t}^h}-\left(1-\alpha_2^h+\tau_1^h\ln(t)\right)+\beta_1^h\hat{z}_{1t}^h+\beta_2^h\left(z_{1t}^h-\hat{z}_{1t}^h\right)-\beta_3^h\frac{\pi_{1t}^h}{\pi_{2t}^h}z_{2t}^h \quad (4.29)$$

The other versions of the demand functions change in the same way due to instrumentation.

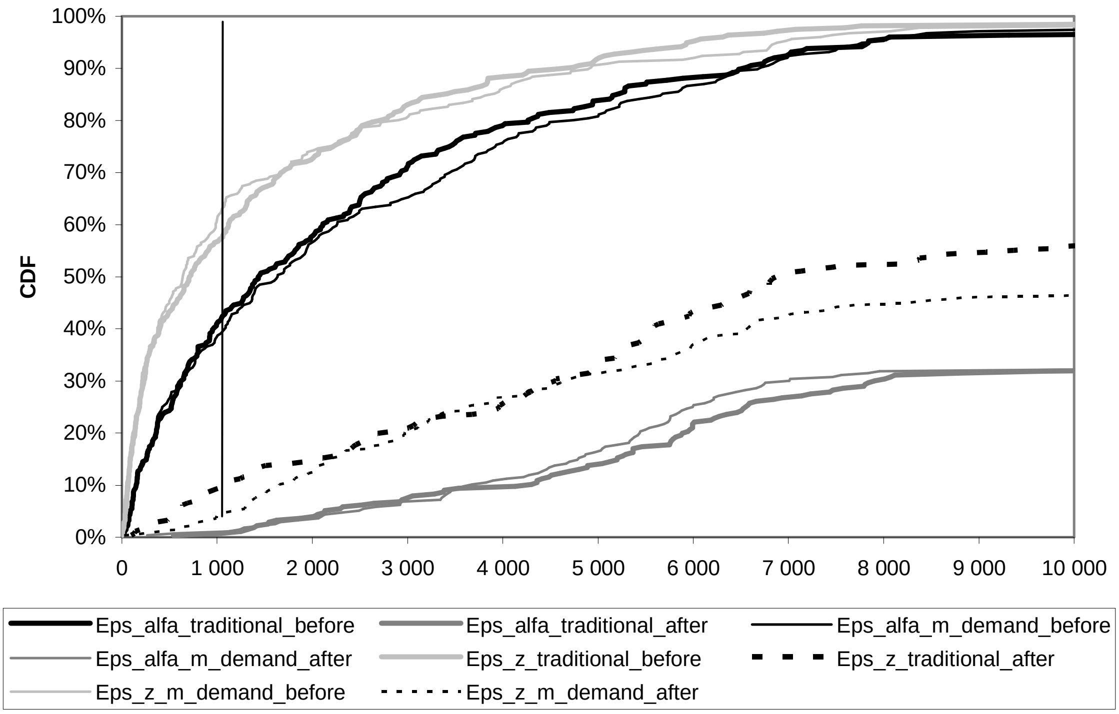
4.4 Results: Where to put the error-term?

As econometricians we never observe everything perfectly, and it is therefore important to be aware of the assumptions we make about what is observed and what is not. In this paper we choose to investigate whether preferences are (un-observably) volatile, or whether we do not observe the optimal consumption perfectly (measurement error). The question is whether the error terms in the utility function (equation 4.17) should be placed on the structural parameters or on the consumption.[48] If the error terms are placed on the parameters, it means that preferences change from period to period, in a way we cannot predict. If the error terms

[48] We have not been able to estimate a model with error terms on both parameters and consumption.

are placed on the consumption, it means that preferences are stable over time, but we do not observe the optimal consumption perfectly. When choosing between models, we ignore the censoring problem (illustrated in Figure 4.2) and only estimate on households that are not censored (n=275). We estimate the different models household by household, using GMM. We estimate both the traditional demand equation and the *m*-demand with error terms on alpha and with error terms on consumption. This leads to four different models; (4.23), (4.24), (4.26) and (4.27). The results from these estimations are compared in order to find the best model and decide whether preferences change over time (random utility error model) or whether we observe consumption imperfectly (measurement error model). The models are estimated in the period before the introduction of mini milk and predictions are calculated both in the period before and in the period after. For each model and each household we calculate the mean of the squared difference between actual consumption and predicted consumption. In Figure 4.20 the Cumulative Density Function (CDF) of these mean squared errors is pictured. The line at 1,000 indicates a mean error of approximately 31.6 per cent. In the model with random utility more than 60 per cent have more than 31.6 per cent error while only 40 per cent in the model with measurement error. In the prediction period the model with measurement error also performs better than the model with random utility.

Figure 4.20: Mean squared percentage error on fat, random utility model and measurement error model

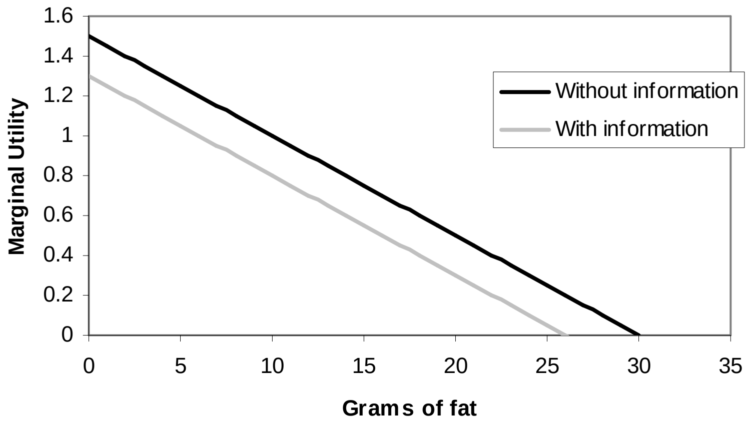


Based on the above realisation of the model we choose to estimate a classical model with measurement errors. This allows us to estimate the linear *m*-demand with measurement error as a two-side censored Tobit model. Furthermore, we include exogenous information to account for changes in preferences over time.

4.5 Final model formulation: Tobit estimation, censoring and information

We model the influence from information as additive on the alpha parameter, which implies that information decreases the marginal utility of fat with the same amount independently of how much fat is consumed. This is illustrated in Figure 4.21.

Figure 4.21: The way information influences the marginal utility for fat



This means that we get at utility function of the form:[49]

$$U(z_1, z_2) = \alpha_1 z_1 + \left(\alpha_2 + \tau_2 \ln(t) + \gamma_2 I\right) z_2 - 0.5\left(\beta_1 z_1^2 + \beta_2 z_2^2\right) \tag{4.30}$$

We do not include the trend and the information in the normalisation $(\alpha_1 + \alpha_2 = 1)$.
The *m*-demand from (4.27) becomes:

$$0 = \left(\alpha_2 + \tau_2 \ln(t) + \gamma_2 I\right)\frac{\pi_1}{\pi_2} - (1-\alpha_2) + \beta_1 z_1 - \beta_2 \frac{\pi_1}{\pi_2}(z_2 + \xi_2) \tag{4.31}$$

which can be rearranged to:

$$z_2 = \omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} + \omega_5 \frac{\pi_2}{\pi_1} z_1 + \xi_2 \tag{4.32}$$

where $$\omega_1 = \frac{\alpha_2}{\beta_2}, \quad \omega_2 = \frac{\tau_2}{\beta_2}, \quad \omega_3 = \frac{\gamma_2}{\beta_2}, \quad \omega_4 = -\frac{(1-\alpha_2)}{\beta_2}, \quad \omega_5 = \frac{\beta_1}{\beta_2} \tag{4.33}$$

[49] Due to the stability of total consumption of milk and to save on degrees of freedom we choose here to formulate the model with only a trend on fat.

Note that $\omega_4 = -\frac{(1-\alpha_2)}{\beta_2} = \omega_1 - \frac{1}{\beta_2} \Leftrightarrow \frac{1}{\beta_2} = \omega_1 - \omega_4$, which means that the relationships are:

$$\alpha_1 = \frac{\omega_4}{\omega_1 - \omega_4}, \quad \alpha_2 = \frac{\omega_1}{\omega_1 - \omega_4}, \quad \beta_1 = \frac{\omega_5}{\omega_1 - \omega_4}, \quad \beta_2 = \frac{1}{\omega_1 - \omega_4}$$
$$\tau_2 = \frac{\omega_2}{\omega_1 - \omega_4}, \quad \gamma_2 = \frac{\omega_3}{\omega_1 - \omega_4} \tag{4.34}$$

The equation can of course also be estimated with z_1 as the dependent variable. The identification issues are equivalent.

Estimation of final model

It is not possible to buy a litre of milkiness without buying at least one gram of fat (skimmed milk), and it is not possible to purchase more than 35 grams of fat per litre of milkiness (whole milk). These restrictions mean that the analytical solution in (4.13) cannot always be obtained. Households that have preferences for milk with less fat than skimmed milk and households that have preferences for milk with more fat than whole milk are censored. This problem is solved by estimating a Tobit model with two-sided censoring (Amemiya, 1984; Tobin, 1958). As the model is estimated for each household individually the actual equation to estimate with instruments (see 4.28) becomes:

$$z_{2t}^h = \omega_1^h + \omega_2^h \ln(t) + \omega_3^h I_t + \omega_4^h \frac{\pi_{2t}^h}{\pi_{1t}^h} + \omega_5^h \frac{\pi_{2t}^h}{\pi_{1t}^h} \hat{z}_{1t}^h + \omega_6^h \frac{\pi_{2t}^h}{\pi_{1t}^h}\left(z_{1t}^h - \hat{z}_{1t}^h\right) + \xi_{2t}^h, \quad z_{1t}^h \leq z_{2t}^h \leq 35\, z_{1t}^h \tag{4.35}$$

After estimating the parameters we then predict consumption of fat both in the estimation period and in the prediction period, ignoring the effect of the residual and using the true value of z_{1t}^h instead of the instrumented variable:

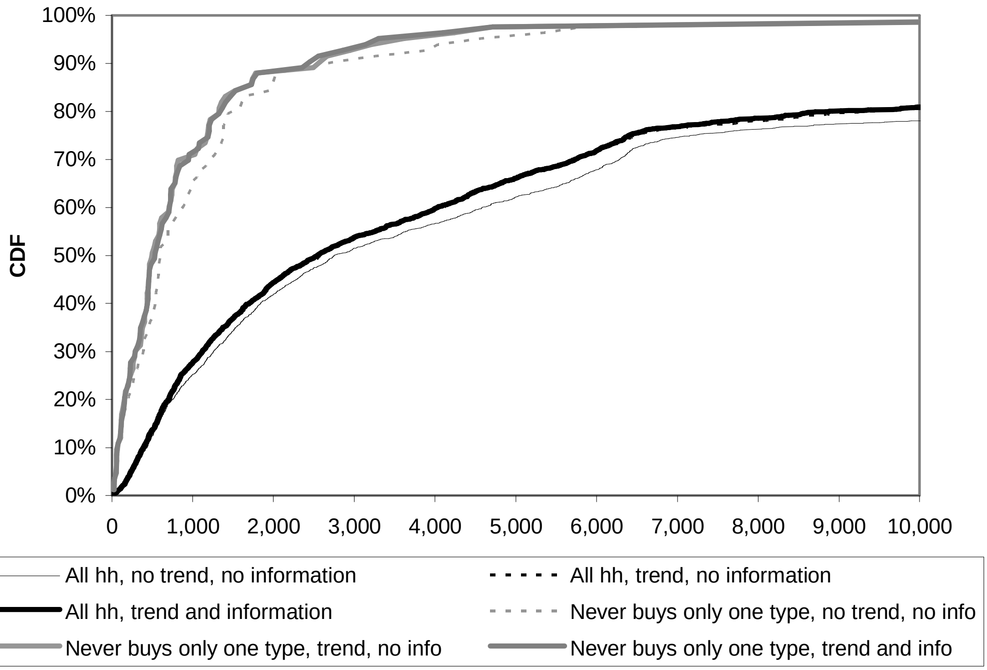
$$\hat{z}_{2t}^h = \omega_1^h + \omega_2^h \ln(t) + \omega_3^h I_t + \omega_4^h \frac{\pi_{2t}^h}{\pi_{1t}^h} + \omega_5^h \frac{\pi_{2t}^h}{\pi_{1t}^h} z_{1t}^h \tag{4.36}$$

We then calculate the predicted milkiness from this and the budget and prices:

$$\hat{z}_{1t}^h = \frac{x_t^h - \pi_{2t}^h \hat{z}_{2t}^h}{\pi_{1t}^h} \tag{4.37}$$

Figure 4.22 shows the distribution of the mean squared percentage error on fat in the final estimation of the Tobit with two-sided censoring with instrumentation. The model is estimated over the whole period with and without trend and information. It is evident that the model which includes a trend to account for changing preferences for fat does better than a model without a trend. Including information along with the trend improves the model slightly.

Figure 4.22: Mean squared percentage error on fat, in an instrumented Tobit model with and without trend and information



Note that the distribution of squared percentage errors in figure 4.20 only includes households who never buy only one type of milk (n=275). Figure 4.22 contains both a curve for households that never buy only one type of milk and curves for all types of households. The households that never buy only one type of milk provide the highest level of information about preferences and therefore lead to much better fits than the average household in the sample.

4.6 Results: Final model formulation

The estimated parameters give a range of possibilities to investigate household preferences for fat. One of the features of a quadratic utility function is that it is possible to calculate a bliss point for fat and for milkiness for each household, i.e. the preferred amount of fat and milkiness bought if there were no prices. If β is diagonal the bliss points can be calculated from the utility function (4.30) as:

$$z_{1t}^{h*} = \frac{\alpha_1^h}{\beta_1^h} \quad \text{and} \quad z_{2t}^{h*} = \frac{\alpha_2^h + \tau_2^h \ln(t) + \gamma_2^h I_t}{\beta_2^h} \tag{4.38}$$

Where z_{1t}^h is milkiness and z_{2t}^h is fat. The optimal fat per cent can then be calculated from (4.38):

$$\frac{z_{2t}^{h*}}{z_{1t}^{h*}} = \frac{\left(\alpha_2^h + \tau_2^h \ln(t) + \gamma_2^h I_t\right)\beta_1^h}{\alpha_1^h \beta_2^h} \tag{4.39}$$

Both the optimal fat and the optimal fat per cent are changing over time due to the influence from the trend and information. Apart from the bliss point and the optimal fat share of fat in milk we also look at the own- and cross price elasticities. The derivation of the own price elasticities and cross-price elasticities between milkiness and fat are shown in Appendix 4B. The rest of this section is divided into subsections each concentrating on one type of results. The first section analyses whether we are able to predict who is buying which types of milk within and out of the estimation period. The second subsection concentrates on describing optimal fat shares for different types of households, while the last section focuses on policy issues, how to regulate consumption of fat from milk. To get more reliable results only households which buy more than one type of milk more than 30 per cent of the time are used in the figures below.

Are we able to predict who will actually choose to buy mini milk?

If the characteristics model is appropriate we ought to be able to predict who will buy mini milk based on parameters estimated in the period before the entrance of mini milk. We do not expect to be able to predict in all possible future due to exogenous shocks, but only within a reasonable time-span from the estimation period. Figure 4.23 shows the share of different types of milk bought in October 2000, a few months before the entrance of mini milk, separated by predicted optimal fat shares based on estimated parameters in the period before the entrance of mini milk. Note that the optimal fat share is the amount of fat per litre of milk the household would prefer if there was no budget constraint and no prices. The fat-haters (optimal fat share <1) have a volume share for skimmed milk close to 80 per cent. The share of skimmed milk is declining with the optimal fat share. The opposite is the case for the volume share for whole milk. The fat-lovers (optimal fat share > 35) have an almost equal share of whole milk and semi-skimmed milk. This might be due to prices since this group of households are found to be rather price elastic.

Figure 4.23: Predicted optimal fat share compared with actual purchases of milk in October 2000

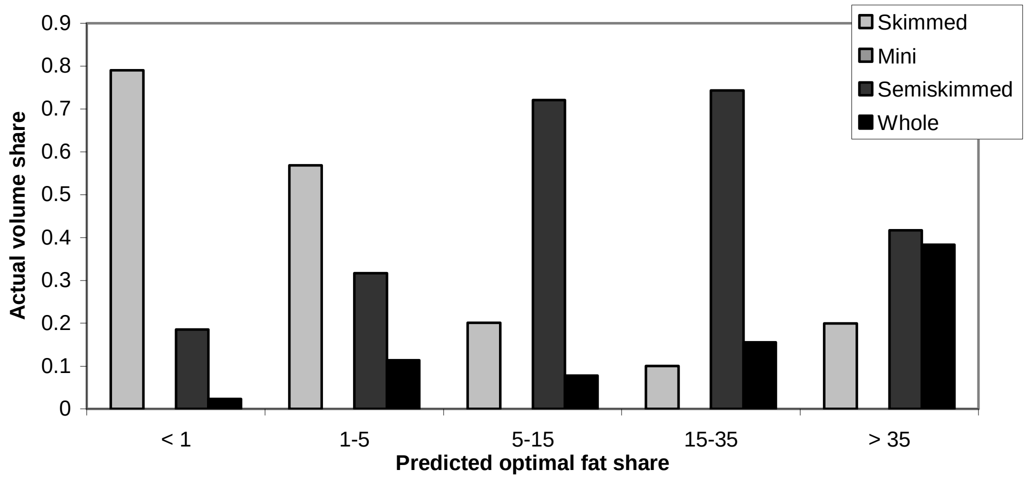


Figure 4.24 shows actual volume shares of different types of milk in October 2001 ten months after the entrance of mini milk separated by predicted optimal fat share based on parameters estimated in the period before the entrance of mini milk i.e. predicted optimal fat share is based on estimations in the period before, while actual consumption is calculated in the period after. Generally, the volume share for mini milk lies between 10-20 per cent for all consumers. This indicates a period where most households try the new type of milk, perhaps initiated by heavy marketing strategies. Mini milk is still rather expensive compared to other types of milk. Apart from the small share of mini milk among all types of consumers the consumption is not very different from consumption illustrated in Figure 4.23.

Figure 4.24: Predicted optimal fat share compared with actual purchases of milk in October 2001

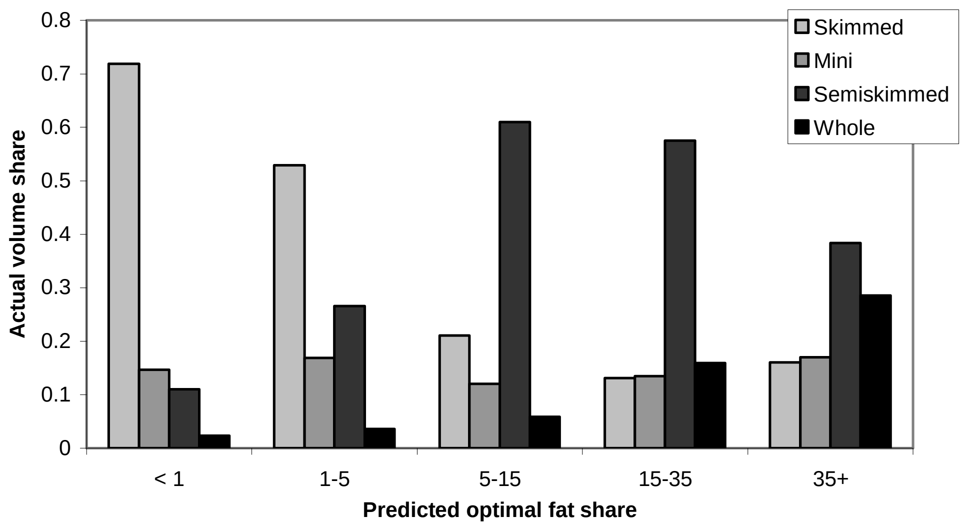
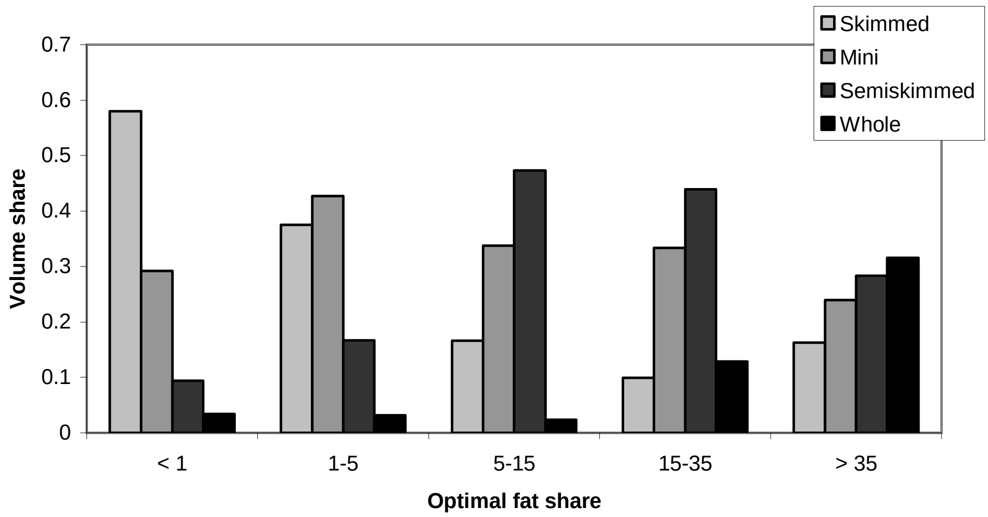


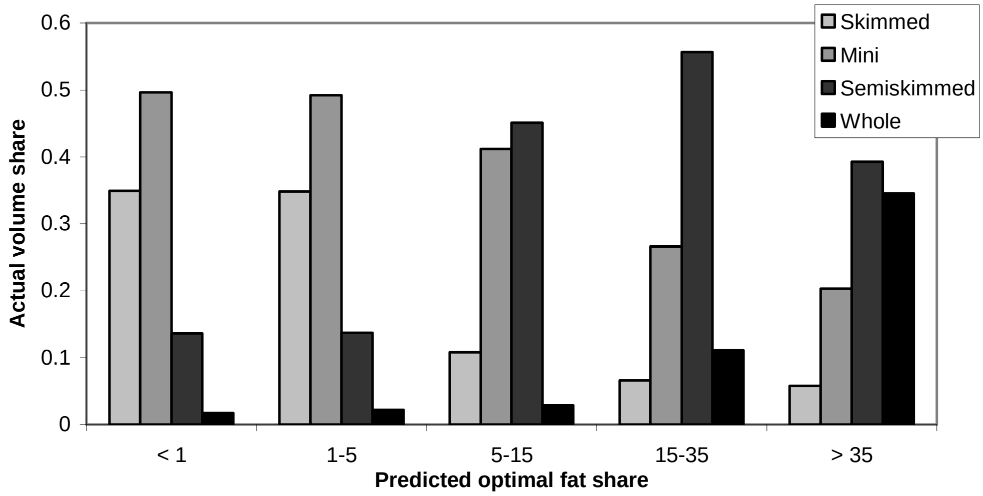
Figure 4.25 shows predicted optimal fat share based on estimated parameters in the period before the entrance of mini milk and actual purchase of milk in October 2004. This means that we are four years out of the estimation period. At this point mini milk has gained an almost stable volume share and prices have declined to a reasonable level. We expect mini milk to increase its volume share especially for those with an optimal share of fat between 1 and 15 grams per litre. This is also what happens, but the volume share is also increasing for the fat-haters (optimal fat share > 35). But generally predictions are not out of proportions compared to the estimated optimal fat share in October 2000, i.e. the characteristics model appears to be suitable to describe the milk market.

Figure 4.25: Predicted optimal fat share compared with actual purchases of milk in October 2004



But predictions get worse as the prediction period gets further away from the estimation period, due to exogenous shocks. The last Figure 4.26 therefore shows actual purchased volume shares in October 2004 separated by predicted optimal fat shares based on parameters estimated on data from the whole period both before and after the entrance of mini milk. This picture is more in accordance with expectations since the largest share of mini milk is consumed among the low to moderate fat consumers (1-15 grams of fat per litre) and have gained some market share from the households with a high optimal fat share. It is interesting that the share of mini milk is so high in the group of very low fat consumers (those that prefer a fat share <1 gram per litre of milk). This might be caused by the extremely low relative price of mini milk as compared to skimmed milk, as it is seen from Figure 4.14.

Figure 4.26: Predicted optimal fat share in October 2004 compared with actual purchases of milk



From this we conclude that the structural characteristics model does a fair job of predicting who will buy the new mini milk.

Valuation of fat for various social and demographic groups and over time

The optimal fat share shows the type of milk that the households would buy if there were no prices and no budget. Especially in marketing strategies, but also in the design of public campaigns with the aim of decreasing the intake of saturated fat it is useful to know the socio-demographic characteristics of the target groups. This subsection shows differences in optimal fat share for different types of households and changes over time. Table 4.2 shows the percentage of households with various combinations of optimal fat and optimal milkiness values. Households with a negative optimal fat value and a negative optimal milkiness value ought not to be buying milk. There are only a few of these (between 2.4 and 3.7 per cent of the panel). They are deleted from the figures below. A little more than four fifths of the panel have a positive optimal value of both fat and milkiness. Most households have a positive optimal fat share. A negative optimal fat share implies that the households would prefer milk with no fat and they think of the fat that comes along with the milkiness in a litre of milk as a nuisance. Those with a positive optimal fat share regard fat as a good to some extent.

Table 4.2: Percentage of the households with different combinations of optimal fat, milkiness and fat share

	Optimal fat < 0		Optimal fat> 0		Optimal fat share *	
Optimal milk	Negative	Positive	Negative	Positive	Negative	Positive
1997	3.7%	8.1%	5.2%	83.0%	13.8%	86.2%
1998	3.2%	7.7%	6.2%	83.0%	14.3%	85.7%
1999	3.0%	6.1%	6.3%	84.6%	12.8%	87.2%
2000	2.6%	6.5%	6.8%	84.1%	13.7%	86.3%
2001	3.7%	14.5%	6.3%	75.5%	21.6%	78.4%
2002	2.6%	7.9%	7.5%	81.9%	15.9%	84.1%
2003	2.4%	9.1%	8.5%	80.1%	17.9%	82.1%
2004	2.5%	9.7%	7.6%	80.3%	17.7%	82.3%

* The optimal fat share (optimal fat share = optimal fat/optimal milkiness) is only calculated for households with a positive valuation of milkiness

Figure 4.27 shows the change over time for the density function over optimal fat shares for households that are in the panel the whole period from 1998 to 2003 (this gives in total 447 households). The distribution is calculated as a kernel regression with Gaussian kernel (see e.g. Blundell and Duncan, 1998). The figures show clearly how the optimal fat share declines over time. The stipulated areas in the figures show the area where it is not possible to reveal preferences i.e. households will have to buy milk with a smaller or larger fat content than actually preferred.

Figure 4.27: The density function for the optimal fat share over time

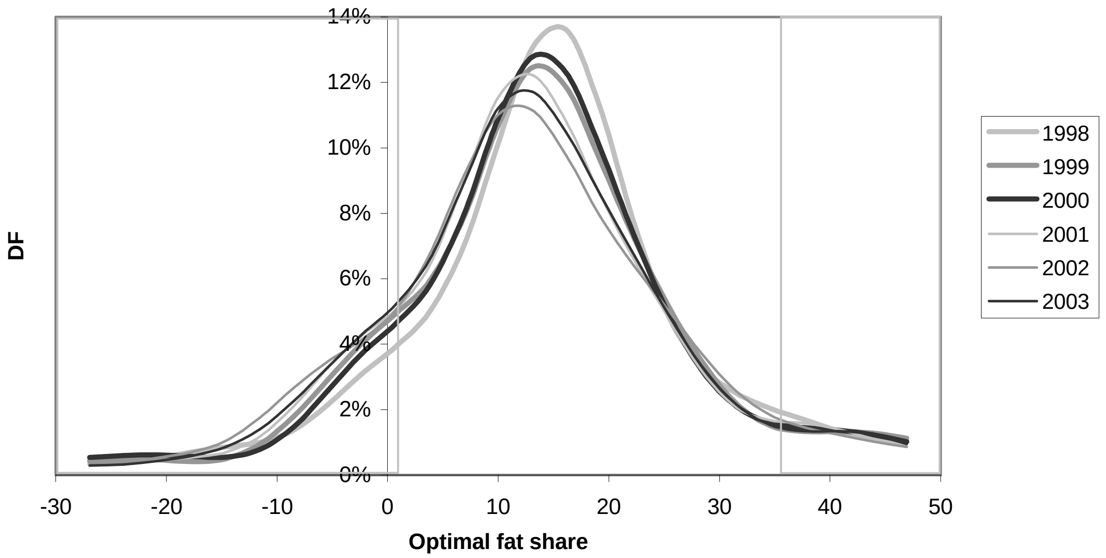


Figure 4.28 and Figure 4.29 show the optimal milkiness consumption together with optimal fat share. The milkiness haters are left out of the figures due to the definition of the optimal fat share. All columns in the figure sum to one. Many households, 40 per cent of the panel,

have a moderate optimal milkiness consumption and a moderate to high optimal fat share (optimal fat between 5 to 35 grams of fat per litre) in 2004. The fat-haters (optimal fat share less than 1) are represented in each group of milkiness attitudes while the fat-lovers (optimal fat share 35 or above) are concentrated among those who prefer a low milkiness consumption. There are no fat-lovers who prefer a high weekly consumption of milkiness. The change in preferences towards milk with lower fat share is clear when comparing the combinations of optimal milkiness consumption and optimal fat share in 1997 (Figure 4.28) with 2004 (Figure 4.29).

Figure 4.28: Distribution of the panel over different optimal fat share and milkiness in 1997

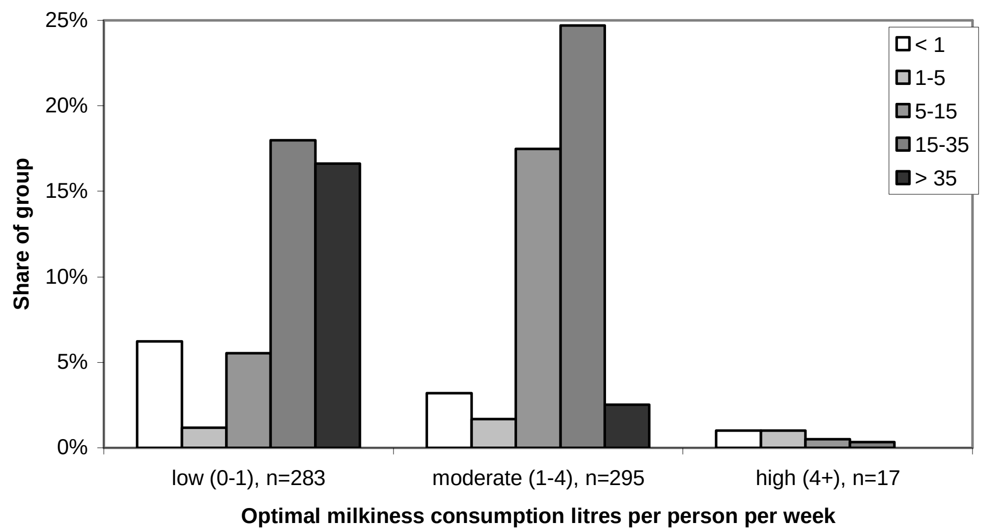


Figure 4.29: Distribution of the panel over different optimal fat share and milkiness in 2004

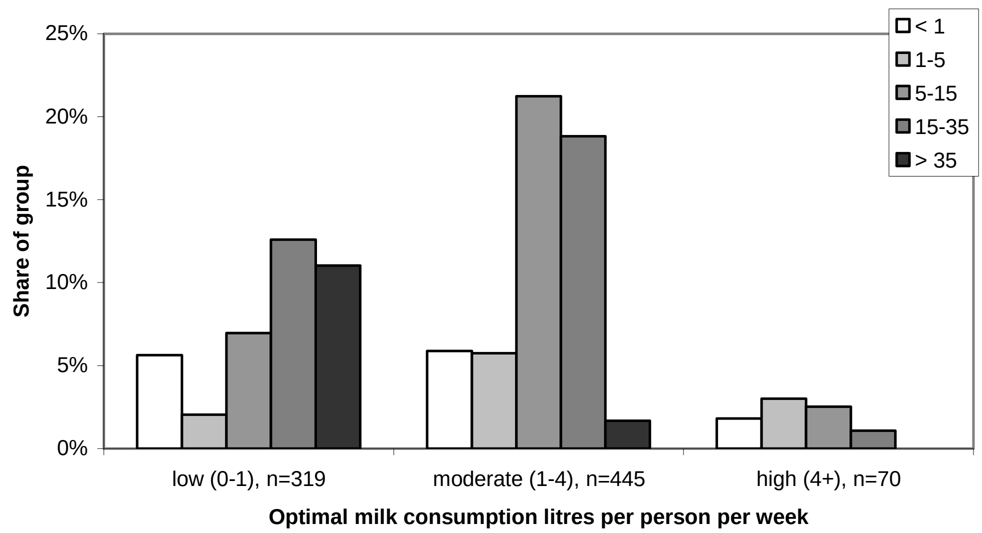


Figure 4.30 shows the optimal fat share for households with different education. There is not much difference between households with no or vocational education, while households with a longer or medium further education[50] prefer a lower fat content. Households with a short education show a distribution with two bulks, one around 12 and another around 32 grams of fat per litre of milk.

Figure 4.30: Optimal fat per cent for households in different educational groups, 1997

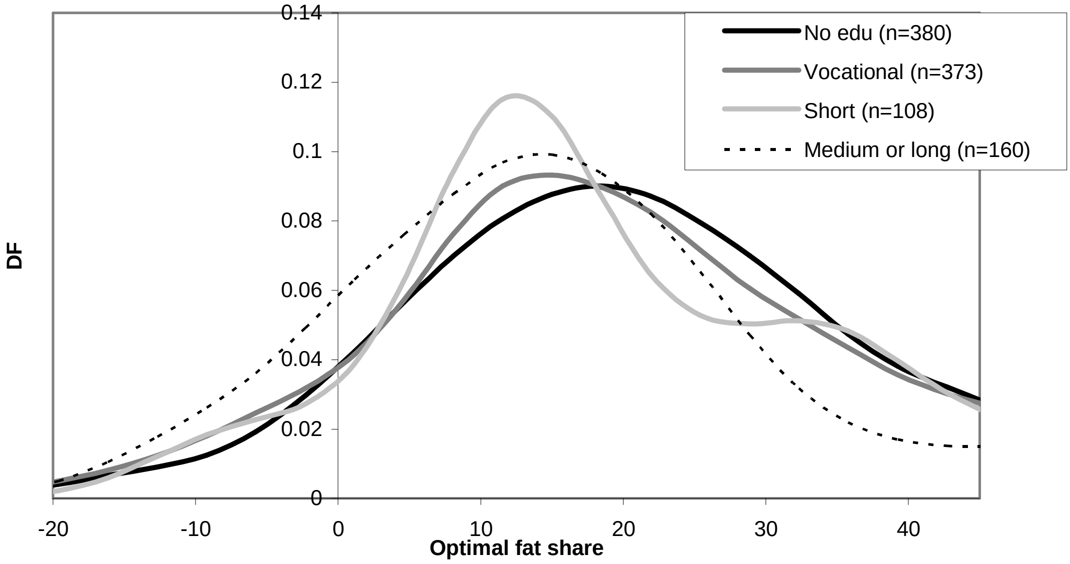


Figure 4.31 shows the distribution over fat share for a combination of education and age, note that the educational definitions here are slightly different, namely divided into practical and no education versus theoretical education. For each of the age groups the theoretical educated prefer milk with lower fat content.

[50] For a detailed description of the educational groups see Smed (2008)

Figure 4.31: Distribution function over fat share for a combination of education and age, 1997

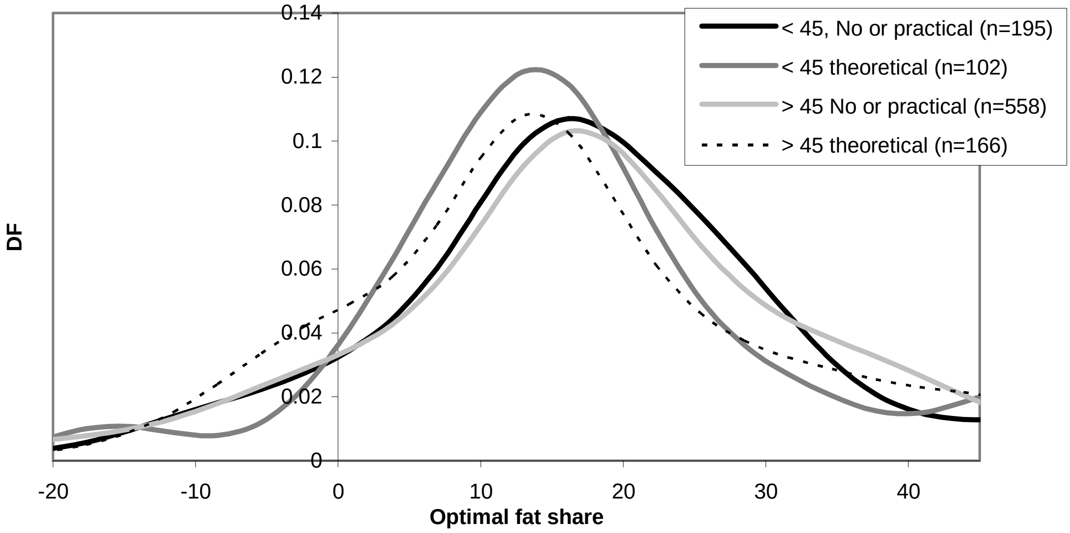
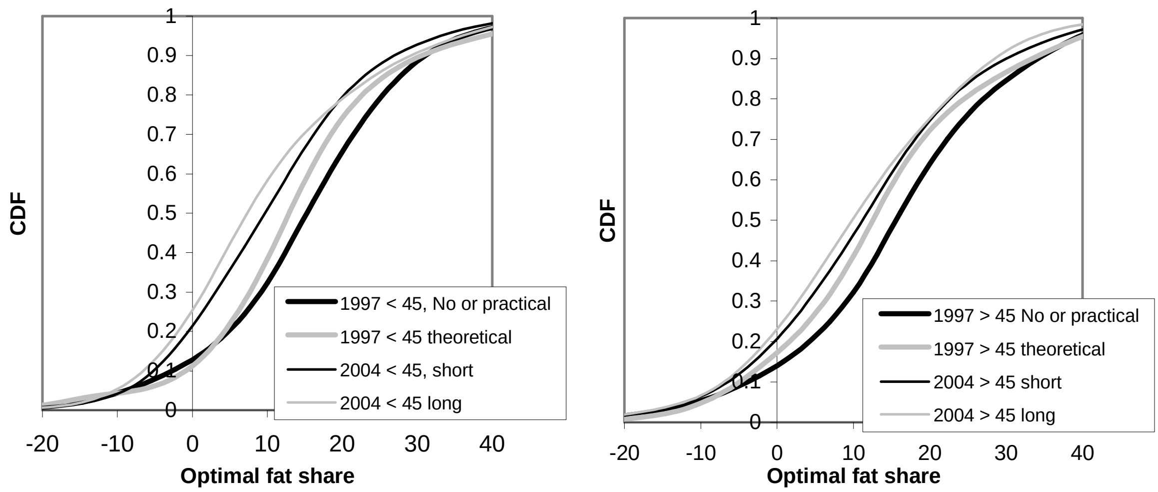


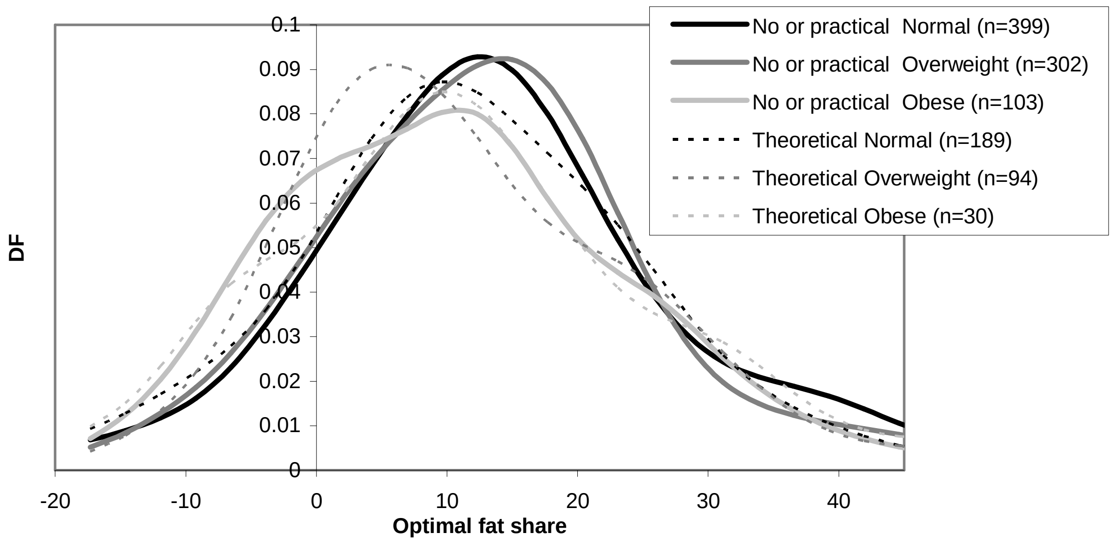
Figure 4.32 shows the change in the cumulative distribution over optimal fat share for combinations of age and educational groups. For older households (45 years or above) there is a larger difference between educational groups than for younger (below 45). The change from 1997 to 2004 seems to be equally large for practical or theoretical educated younger households while the practical or no educated older decrease their optimal fat share more than the theoretical educated older.

Figure 4.32: Change in CDF of optimal fat share for combinations of age and education



Finally, Figure 4.33 shows the distribution over optimal fat share in 2004 for combinations of BMI[51] and education. Again, the theoretically educated households have a lower optimal fat share than households with no or practical education, but interestingly it seems like obese individuals prefer a lower optimal fat share than those with normal weight. This might indicate that the consumption of milk is an area where it is rather convenient to save calories.

Figure 4.33: Distribution of optimal fat share for combinations of BMI and education



Political implications – who can be affected by prices and information

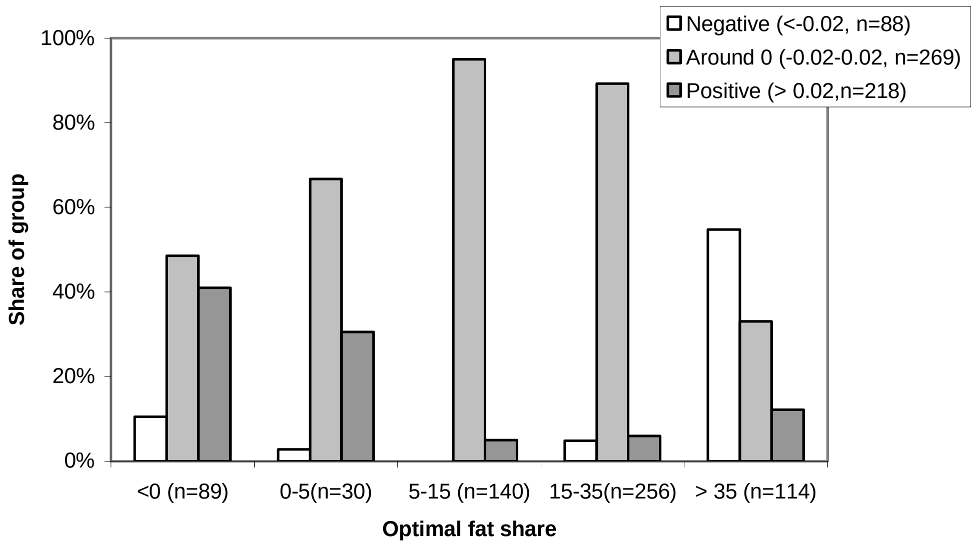
It is of great interest to investigate the size and sign of the price elasticity, the trend and information parameter for households with different optimal fat share. Is it the fat-lovers who decrease their consumption of fat according to information or over time or are they more sensitive to price changes or both? In the following figures the panel is divided into groups according to their optimal fat share and their trend and information parameters are compared together with own price elasticities for fat. A negative trend parameter indicates that the optimal amount of fat in grams per week per person or the optimal fat share decline over time, while a negative information parameter indicates that households decrease their optimal fat share according to the incoming information about the relation between fat consumption and

---

51 Questions of height and weight for each individual in the household are only posed in 2004. BMI is calculated as: $BMI = \frac{weight(kg)}{height(m)*height(m)}$. Overweight is then defined as a BMI above 25, but below 30, while obesity is defined as having a BMI above 30.

health. On average, 57 per cent of the households have a negative trend parameter. Figure 4.34 shows the share of households with negative and positive trends, respectively, separated by optimal fat share (the columns within each group sum to 1). In general, households that like fat (the fat lovers who prefer an optimal fat share > 35) have a larger tendency to have a negative trend for fat, while households that do not like fat (optimal fat share < 5) have a larger tendency to have a positive trend than the average. Most households with a moderate fat share do not change consumption (the trend parameter is around zero).

Figure 4.34: Optimal fat share and the trend parameter in 1997



Most households have an information parameter just around zero. A positive and significant reaction to information gives no meaning in the current model. Of great interest is the 11 per cent of the panel having a large reaction to information (an information parameter below -0.0005). One fourth of these are fat-haters (optimal fat share < 0 grams per litre) while one third are high fat consumers (optimal fat share =15-35 grams per litre) and another fourth are fat-lovers (> 35 grams of fat per litre). Figure 4.35 shows the sign of the information parameter separated by optimal fat share (columns within each group sum to 1). The figure shows clearly that those who react to information are either the fat-lovers or fat-haters. Those who reacts the least are moderate to high fat consumers.

Figure 4.35: Optimal fat share and the sign of the information parameter in 1997

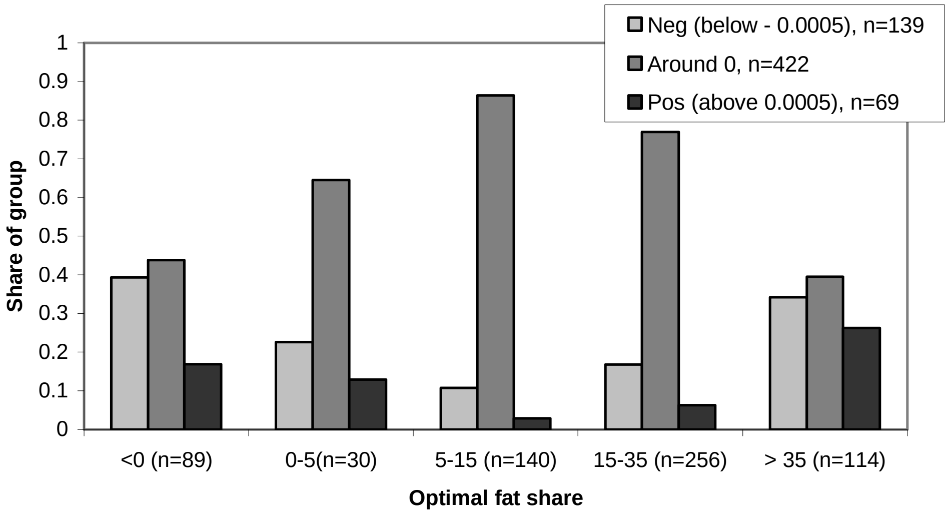
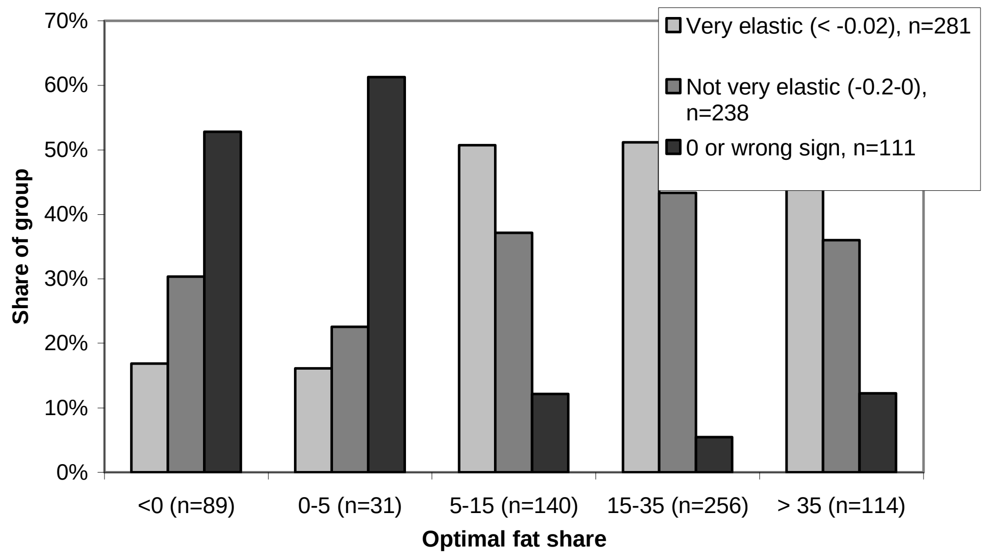


Figure 4.36 shows the price elasticity separated by optimal fat share (columns in each group sum to one). Most households have a negative own price elasticity for fat (17 per cent have an own price elasticity of 0 or with wrong sign). As much as 45 per cent are rather price elastic with an own price elasticity below -0.2. This figure clearly shows that fat-haters (optimal fat share below 0) and low fat consumers (optimal fat share between 0 and 5) are not very price elastic, while the fat-lovers (optimal fat share at 35 or above) and the moderate to high fat consumers (optimal fat share at 5-35 grams per litre) are rather price elastic. That the fat-haters are price inelastic comes naturally from these households being on the edge and the closest they are to having their preferences fulfilled are by consuming skimmed milk. The prices of the other types of milk would have to change radically to make these types of milk attractive to the fat-haters. More interestingly is it that the fat-lovers, who are also on the edge, but in the other end of the possible consumption set, are rather influenced by prices.

Figure 4.36: Optimal fat share and mean own price elasticity in 1997



4.7 Conclusion and discussion

The market for milk is suitable for economic analysis since almost all Danish households purchase milk and the characteristics inherent in milk are well defined. During our data period there has been a significant decrease in the consumption of fat from milk without any particular decrease in the total consumption of milk. This decrease has been due to both changing preferences for fat and the entrance of a new low-fat variety of milk. In this paper, the demand for fat in milk has been analysed in a structural characteristics model for milk. Estimating a structural model makes it possible to separate preference for milk from the influence of prices, trends and information. The analysis state that a model with measurement errors performs better than a model with random parameters in the utility function. The entrance of a new type of milk with the same characteristics as existing products on the market, but in new proportions, makes us capable of testing whether the characteristics model is appropriate to analyse the market for milk. If the model is correct the households with an optimal fat share between 1 and 15 are those that will be the target groups for this type of milk since mini milk has a fat content at 5 grams per litre. This is true to a large extent. Those with the largest volume share of mini milk are the low to moderate fat consumers. This implies that the characteristics model is considered to be appropriate to describe the market for milk.

Over time consumers seem to prefer milk with less fat. This change seems to be due to both a general trend and for some consumers also the influence of information. In 1997 households with small children preferred milk with a higher fat share than other types of households, in 2004 this had changed, presumably because children below the age of 3 now were recommended to drink semi-skimmed milk instead of whole milk. Higher educated households prefer milk with a lower fat content than lower educated, but for households where the head of the family is above 45 this difference seems to disappear over time. Interestingly, there are no large differences between weight groups and preferences for milk. It even seems like obese and overweight have preferences for milk with a lower fat content than normal weight individuals. Both among those who consume milk in moderate and in low amounts there has been a decrease in the preferred optimal fat share. The majority of the fat-haters (those with an optimal fat share below 0) have a positive trend in the optimal fat consumption while most fat-lovers (optimal fat share above 35) have a large negative trend for fat. This indicates that households that prefer milk with a high fat content decrease their consumption of fat more than other types of households. Most households that prefer milk with a high fat content are moderate milk consumers (i.e. prefer less than 1 litre a week). It is therefore important to take the amount of milk consumed into account when predicting the changes in total amount of fat consumed, not only the share of fat.

In order to plan, design and implement political interventions with the aim of changing consumers' preferences for fat it is of major importance to know how to reach the target groups. Most households do not react to information, but among those who do, there is an over-representation of fat-lovers and the fat-haters. Information might therefore be one way to reach households that prefer milk with a high fat content. However, using information to change consumption might also influence the fat-haters. It is therefore important to consider what happens if the fat-haters get lower preferences for fat. Price policy might be a more effective way to reach high fat consumers since most households have a negative own price elasticity for fat. Households that prefer milk with a fat content lower than 5 grams per litre are mostly price inelastic so the price instrument will not influence the fat-haters to the same extent as will information. The price instrument will reach a broader group of households since also moderate fat consumers are rather price sensitive. This is of great importance since there is a larger share of high milk consumers to be found among the moderate fat consumers. Introducing new products on the market might also be a route to having consumers decrease their consumption of fat from milk. This might be important since on average 5.7 per cent of total saturated fat consumption comes from milk. If this is decreased by two thirds due to a change from semi-skimmed milk to mini milk this will have significant influence on total fat consumption. Another consequence of new products on the market might be that often new products are accompanied by a huge amount of advertising. This was also the case when the mini milk entered the market. How this advertising influences preferences might be a route for further research.

References:

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

Bartik T. (1987): The estimation of demand parameters in hedonic price models. *Journal of Political Economy*, 95(1):81-88

Blundell R.W. and J.L. Powell. (2003): *Endogeneity in Nonparametric and Semiparametric Regression Models.* In M. Dewatripont, L.P. Hansen and S.J. Turnovsky (eds.): *Advances in Economics and Econometrics,* Chapter 8: 312-357 Cambridge University Press, Cambridge

Blundell R.W. and A. Duncan (1998): Kernel regression in empirical microeconomics. *Journal of Human Resources*, 33(1): 62-87

Brown J.N. and H.S. Rosen (1982): On the estimation of structural hedonic price model. *Econometrica*, 50(3): 765-768

Brown D.J. and L.F. Schrader (1990): Cholesterol information and shell egg consumption. *American Journal of Agricultural Economics*, 72(3): 548-555

Browning M. (1999): *Modelling commodity demands and labour supply with m-demands*. Discussion Papers 99-08, Department of Economics, University of Copenhagen, Denmark

Blow L., M. Browning and I. Crawford (2005): *Revealed preference analysis of characteristics models.* CAM working paper no 4, CAM, University of Copenhagen

Chang H. and H.W. Kinnucan (1991): Advertising, information, and product quality: the case of Butter. *American Journal of Agricultural Economics*, 73(4): 1195-1203

Chern W.S. and J. Zuo (1995): *Alternative measures of changing consumer information on fat and cholesterol*. Paper presented in the annual meeting of the American Agricultural Economics Association, Indianapolis, Indiana, 6-9 August

Cook C.M. and D.B. Eastwood (1992): Incorporating subsistence into hedonic price and nutrient demand equations. *Journal of Consumers Affairs*, 26(2): 288-304

Deaton A. and J. Muellbauer (1980): *Economics and consumer behaviour*. Cambridge University Press, Cambridge, UK

Diamond D.B. and B.A. Smith (1985): Simultaneity in the market for housing characteristics. *Journal of Urban Economics*, 17: 280-292

Eastwood D.B., J.E. Brooker and D.E Terry (1986): Household nutrient demand: Use of characteristics theory and a common attribute model. *Southern Journal of Agricultural Economic*, 18(2): 235-246

Ekeland I., J. J. Heckman and L. Nesheim (2004): Identification and Estimation of hedonic models. *Journal of Political Economy*, 112(S1): 60-109

Epple D. (1987): Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy*, 95(1): 59-80

Gorman W.M. (1980): A possible procedure for analysing quality differentials in the egg market. *Review of Economic Studies*, 47(5): 843-856

Gravelle H. and R. Rees (1992): *Microeconomics*. Longmand Group UK, Harlow, England

Kahn S. and K. Lang (1988): Efficient Estimation of Structural Hedonic Systems. *International Economic Review*, 29(1): 157-166

Kim W. and W.S. Chern (1995): Health risk concern of households vs. processors: Estimation of hedonic prices in fats and oils. In Caswell, J.A. (ed.): *Valuing food safety and nutrition*. Westview Press, Boulder, Colorado

Kim W. and W.S. Chern (1997): Indicative measures of health-risk information on fat and cholesterol for US and Japanese consumers. *Consumer Interest Annual*, 43 (2): 84-9

Kim W. and W.S. Chern (1999): Alternative measures of health information and demands for fats and oils in Japan. *Journal of Consumer Affairs*, 33: 92-109

Kinnucan H.W., H. Xiao, C-J. Hsia and J.D. Jackson (1997): Effects of Health Information and Generic Advertising on U.S Meat Demand. *American Journal of Agricultural Economics*, 79(2): 13-23

Ladd G.W. and M. Zober (1977): Model of consumer reaction to product characteristics. *The Journal of Consumer Research*, 4(2): 89-101

Ladd G.W. and V. Suvannunt (1976): A Model of Consumer Goods Characteristics. *American Journal of Agricultural Economics*, 58: 504-510

Lancaster K. (1966): A new approach to consumer theory. *Journal of Political Economy*, 74(2): 132-157

McGuirk A., P. Driscoll, J. Alwang and L. Huang (1995): System misspecification testing and structural change in the demand for meat. *Journal of Agricultural and Resource Economics*, 20(1): 1-21

Muellbauer J. (1974): Household production theory, quality and the hedonic technique. *American Economic Review*, 64(6): 977-994

Piggott N.E. and T.L. Marsh (2004): Does food safety information impact U.S. food demand? *American Journal of Agricultural Economics*, 86 (1): 154–174

Rickertsen K., J.A. Chalfant and M. Steen (1995): The Effects of Advertising on the Demand for Vegetables. *European Review of Agricultural Economics*, 22(4): 481-494

Rosen S. (1974): Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1): 34-55

Schmit T.M. and H.M. Kaiser (2004): Decomposing the variation in generic advertising response over time. *American Journal of Agricultural Economic*, 86(1): 139-153

Smith M.E., E.O. van Ravenswaay and S.R. Thompson (1988): Sales loss determination in food contamination incidents: An application to milk bans in Hawaii. *American Journal of Agricultural Economics*, 70(3): 513-520

Smed S. (2005): Demand structure and willingness to pay for organic dairy products. *Cahiers Options Méditerraneées*, 64:33-44

Smed (2008): Chapter 2 this thesis

Smed S. and J.D. Jensen (2004): Demand for low fat dairy product – demand for healthiness or taste? In Russel C.S. and S. Krarup (eds.): *Environment, Information and Consumer Behaviour*. AKF Forlaget, Copenhagen

Statistics Denmark (2008): http://www.statistikbanken.dk/statbank5a/default.asp?w=1024

Theil H. (1971): *Introduction to demand and index number theory*. Centre for mathematical studies in business and economics, Report 7126, University of Chicago

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

Verbeke W. and R.W. Ward (2001): A fresh meat almost ideal demand system incorporating negative TV press and advertising impact. *Agricultural Economics*, 25(2-3): 359-374

Appendix 4A: Derivation of demand function given quadratic utility

Assume the utility function:

$$U(z_t) = (\alpha + \varepsilon_t)' z_t - 0.5 z_t' B z_t, \ \ \varepsilon \sim N(0, \Sigma) \tag{A.1}$$

Where *z* is quantities of characteristics purchased. The number of characteristics is *J*, so the dimension of *z* is $J \times 1$. Let π be the price of the characteristics. This leads to the maximisation problem:

$$\begin{aligned} &\max_z U(z) \\ &\text{s.t. } x = \pi' z \end{aligned} \tag{A.2}$$

The Lagrange equation becomes:

$$\begin{aligned} L(z, \lambda) &= U(z) - \lambda(\pi' z - x) \\ &= -0.5 z' \beta z + (\alpha + \varepsilon)' z - \lambda(\pi' z - x) \end{aligned} \tag{A.3}$$

and the first-order conditions become:

$$\frac{\nabla L}{\nabla z} = -\beta z + (\alpha + \varepsilon) - \lambda \pi = 0 \tag{A.4}$$

$$\frac{\nabla L}{\nabla \lambda} = \pi' z - x = 0 \tag{A.5}$$

We would like to find the demand function, so we isolate z (A.4.), which leads to:

$$z = \beta^{-1}\left((\alpha + \varepsilon) - \lambda \pi\right) \tag{A.6}$$

combining this with the budget restriction in (A.5) leads to:

$$\begin{aligned} 0 &= z'\left(\beta^{-1}\left((\alpha + \varepsilon) - \lambda \pi\right)\right) - x \\ &= z' \beta^{-1}(\alpha + \varepsilon) - \lambda \pi' \beta^{-1} \pi - x \\ &\Updownarrow \\ \lambda &= \left(\pi' \beta^{-1} \pi\right)^{-1}\left(\pi' \beta^{-1}(\alpha + \varepsilon) - x\right) \\ &= \left(\underbrace{\underbrace{\pi'}_{1\times J} \underbrace{\beta^{-1}}_{J\times J} \underbrace{\pi}_{J\times 1}}_{1\times 1}\right)^{-1} \left(\underbrace{\underbrace{\pi'}_{1\times J} \underbrace{\beta^{-1}}_{J\times J} \underbrace{(\alpha + \varepsilon)}_{J\times 1}}_{1\times 1} - \underbrace{x}_{1\times 1}\right) \end{aligned} \tag{A.7}$$

Inserting this in the first order conditions in (A.6)leads to:

$$\begin{aligned} z &= \beta^{-1}\left(\left(\alpha+\varepsilon\right)-\lambda\pi\right) \\ &= \beta^{-1}\left(\left(\alpha+\varepsilon\right)-\left(\pi'\beta^{-1}\pi\right)^{-1}\left(\pi'\beta^{-1}\left(\alpha+\varepsilon\right)-x\right)\pi\right) \end{aligned} \tag{A.8}$$

Rearranging (A.8) leads to:

$$z = \beta^{-1}\left(\alpha+\varepsilon\right)-\left(\beta^{-1}\pi\left(\pi'\beta^{-1}\pi\right)^{-1}\right)\left(\pi'\beta^{-1}\left(\alpha+\varepsilon\right)-x\right) \tag{A.9}$$

with the dimensions:

$$\underbrace{z}_{J\times1} = \underbrace{\underbrace{\beta}_{J\times J}{}^{-1}\underbrace{\left(\underbrace{\alpha}_{J\times1}+\underbrace{\varepsilon}_{J\times1}\right)}_{J\times1}}_{J\times1}-\underbrace{\left(\underbrace{\beta}_{J\times J}{}^{-1}\underbrace{\pi}_{J\times1}\underbrace{\left(\underbrace{\pi'}_{1\times J}\underbrace{\beta}_{J\times J}{}^{-1}\underbrace{\pi}_{J\times1}\right)}_{1\times1}{}^{-1}\right)}_{J\times1}\underbrace{\left(\underbrace{\pi'}_{1\times J}\underbrace{\beta}_{J\times J}{}^{-1}\underbrace{\left(\underbrace{\alpha}_{J\times1}+\underbrace{\varepsilon}_{J\times1}\right)}_{J\times1}-\underbrace{x}_{1\times1}\right)}_{1\times1} \tag{A.10}$$

Appendix 4B: Derivation of elasticities in the Tobit model

Fat:

The predicted demand for fat is given by:

$$z_2 = \omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} + \omega_5 \frac{\pi_2}{\pi_1} z_1 \qquad \text{(B.1)}$$

If we remember that the relationship between milkiness, fat and the budget is:

$$z_1 = \frac{x - \pi_2 z_2}{\pi_1}, z_2 = \frac{x - \pi_1 z_1}{\pi_2} \qquad \text{(B.2)}$$

we can calculate the demand for fat as a function of the budget instead of the milkiness:

$$z_2 = \left(\frac{\pi_1^2}{\pi_1^2 + \omega_5 \pi_2^2} \right) \left(\omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} + \omega_5 \frac{\pi_2}{\pi_1} \frac{x}{\pi_1} \right) \qquad \text{(B.3)}$$

this can be translated into:

$$\begin{aligned} & z_2 \equiv \frac{f_2}{g_2} h_2 \\ & f_2 \equiv \pi_1^2, \qquad g_2 \equiv \pi_1^2 + \omega_5 \pi_2^2, \quad h_2 \equiv \omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} + \omega_5 \frac{\pi_2}{\pi_1} \frac{x}{\pi_1} \end{aligned} \qquad \text{(B.4)}$$

In general, the derivative of a function like z_2 is:

$$\frac{\partial \left(\frac{f}{g} h \right)}{\partial y} = \frac{g \frac{\partial f}{\partial y} - f \frac{\partial g}{\partial y}}{g^2} h + \frac{f}{g} \frac{\partial h}{\partial y} \qquad \text{(B.5)}$$

In order to calculate price and income elasticities we need the derivatives:

$$\begin{aligned} & \frac{\partial f_2}{\partial \pi_1} = 2\pi_1, \quad \frac{\partial f_2}{\partial \pi_2} = 0, \quad \frac{\partial f_2}{\partial x} = 0, \quad \frac{\partial g_2}{\partial \pi_1} = 2\pi_1, \quad \frac{\partial g_2}{\partial \pi_2} = 2\omega_5 \pi_2, \quad \frac{\partial g_2}{\partial x} = 0 \\ & \frac{\partial h_2}{\partial \pi_1} = -\omega_4 \frac{\pi_2}{\pi_1^2} - 2\omega_5 \frac{\pi_2 x}{\pi_1^3}, \quad \frac{\partial h_2}{\partial \pi_2} = \frac{\omega_4}{\pi_1} + \omega_5 \frac{x}{\pi_1^2}, \quad \frac{\partial h_2}{\partial x} = \omega_5 \frac{\pi_2}{\pi_1^2} \end{aligned} \qquad \text{(B.6)}$$

Define

$$\begin{aligned} & D \equiv \pi_1^2 + \omega_5 \pi_2^2 \\ & C \equiv \omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} \end{aligned} \qquad \text{(B.7)}$$

then the elasticities become:

$$\frac{\partial z_2}{\partial \pi_1}\frac{\pi_1}{z_2}=\frac{\partial\left(\frac{f_2}{g_2}h_2\right)}{\partial \pi_1}\frac{\pi_1}{z_2}=\left(\frac{2\omega_5\pi_1\pi_2^2}{D^2}\left(C+\omega_5\frac{\pi_2}{\pi_1}\frac{x}{\pi_1}\right)+\frac{\pi_2}{D}\left(-\omega_4-2\omega_5\frac{x}{\pi_1}\right)\right)\frac{\pi_1}{z_2} \qquad \text{(B.8)}$$

$$\frac{\partial z_2}{\partial \pi_2}\frac{\pi_2}{z_2}=\frac{\partial\left(\frac{f_2}{g_2}h_2\right)}{\partial \pi_2}\frac{\pi_2}{z_2}=\left(\frac{-2\omega_5\pi_1^2\pi_2}{D^2}\left(C+\omega_5\frac{\pi_2}{\pi_1}\frac{x}{\pi_1}\right)+\frac{\pi_1}{D}\left(\omega_4+\omega_5\frac{x}{\pi_1}\right)\right)\frac{\pi_2}{z_2} \qquad \text{(B.9)}$$

$$\frac{\partial z_2}{\partial x}\frac{x}{z_2}=\frac{\partial\left(\frac{f_2}{g_2}h_2\right)}{\partial x}\frac{x}{z_2}=\frac{\omega_5\pi_2}{D}\frac{x}{z_2} \qquad \text{(B.10)}$$

Milk:

From the equations (B.1) and (B.2) it is also possible to calculate z_1 as a function of the budget:

$$z_1=\left(\frac{\pi_1\pi_2}{\omega_5\pi_2^2+\pi_1^2}\right)\left(-\omega_1-\omega_2\ln(t)-\omega_3 I-\omega_4\frac{\pi_2}{\pi_1}+\frac{x}{\pi_2}\right) \qquad \text{(B.11)}$$

Just checking:

Remember that: $C\equiv\omega_1+\omega_2\ln(t)+\omega_3 I+\omega_4\frac{\pi_2}{\pi_1}$ and $D\equiv p_1^2+\pi_5 p_2^2$. Then (from (B.3) and (B.10):

$$z_1=\left(\frac{\pi_1\pi_2}{D}\right)\left(\frac{x}{\pi_2}-C\right),\quad z_2=\left(\frac{\pi_1^2}{D}\right)\left(C+\omega_5\frac{\pi_2}{\pi_1}\frac{x}{\pi_1}\right) \qquad \text{(B.12)}$$

and the price of the choices is

$$\pi_1 z_1+\pi_2 z_2=\pi_1\left(\frac{\pi_1\pi_2}{D}\right)\left(\frac{x}{\pi_2}-C\right)+\pi_2\left(\frac{\pi_1^2}{D}\right)\left(C+\omega_5\frac{\pi_2}{\pi_1}\frac{x}{\pi_1}\right)=x \qquad \text{(B.13)}$$

as desired.

In order to calculate the elasticities we reformulate z_1 in the same way as we reformulated z_2 in (B.4):

$$z_1 \equiv \frac{f_1}{g_1} h_1$$
$$f_1 \equiv \pi_1 \pi_2, \quad g_1 \equiv \pi_1^2 + \omega_5 \pi_2^2, \quad h_1 \equiv -\omega_1 - \omega_2 \ln(t) - \omega_3 I - \omega_4 \frac{\pi_2}{\pi_1} + \frac{x}{\pi_2} \tag{B.14}$$

again we calculate the derivatives:

$$\frac{\partial f_1}{\partial \pi_1} = \pi_2, \quad \frac{\partial f_1}{\partial \pi_2} = \pi_1, \quad \frac{\partial f_1}{\partial x} = 0, \quad \frac{\partial g_1}{\partial \pi_1} = 2\pi_1, \quad \frac{\partial g_1}{\partial \pi_2} = 2\omega_5 \pi_2, \quad \frac{\partial g_1}{\partial x} = 0$$
$$\frac{\partial h_1}{\partial \pi_1} = \omega_4 \frac{\pi_2}{\pi_1^2}, \quad \frac{\partial h_1}{\partial \pi_2} = -\frac{\omega_4}{\pi_1} - \frac{x}{\pi_2^2}, \quad \frac{\partial h_1}{\partial x} = \frac{1}{\pi_2} \tag{B.15}$$

and again this leads to a set of elasticities:

$$\frac{\partial z_1}{\partial \pi_1}\frac{\pi_1}{z_1} = \frac{\partial\left(\frac{f_1}{g_1}h_1\right)}{\partial \pi_1}\frac{\pi_1}{z_1} = \left(\frac{\omega_5 \pi_2^3 - \pi_1^2 \pi_2}{D^2}\left(-C + \frac{x}{\pi_2}\right) + \frac{\pi_1 \pi_2}{D}\left(\omega_4 \frac{\pi_2}{\pi_1^2}\right)\right)\frac{\pi_1}{z_1} \tag{B.16}$$

$$\frac{\partial z_1}{\partial \pi_2}\frac{\pi_2}{z_1} = \frac{\partial\left(\frac{f_1}{g_1}h_1\right)}{\partial \pi_2}\frac{\pi_2}{z_1} = \left(\frac{\pi_1\left(\pi_1^2 - \omega_5 \pi_2^2\right)}{D^2}\left(-C + \frac{x}{\pi_2}\right) + \frac{\pi_1 \pi_2}{D}\left(-\frac{\omega_4}{\pi_1} - \frac{x}{\pi_2^2}\right)\right)\frac{\pi_2}{z_1} \tag{B.17}$$

$$\frac{\partial z_1}{\partial x}\frac{x}{z_1} = \frac{\partial\left(\frac{f_1}{g_1}h_1\right)}{\partial x}\frac{x}{z_1} = \frac{\pi_1}{D}\frac{x}{z_1} \tag{B.18}$$

Chapter 5

Information processing strategies and health information

Sinne Smed

Lars Gårn Hansen

AKF
(Danish Institute of Governmental Research)
Copenhagen

Martin Browning

(Nuffield College)
University of Oxford & Fellow

Abstract

This paper addresses how consumers process information about health characteristics in fish and transfer into changed consumption patterns. Information about the positive nutritional effects of fish consumption applies generally to both fatty and lean fish, while information about dioxin is especially related to fatty fish. Some consumers may know that it is possible to avoid the risk of dioxin poisoning by substituting away from fatty fish and toward lean fish (a sophisticated reaction) while other consumers may not be able to make the distinction (unsophisticated reaction). A third way of reaction may be to ignore the information (ignorance). Consumers' choice of strategy is determined by estimating a two-stage demand system (AIDS) for each household, with new information treated as an adjustment to the prices. We find that approximately half of the consumers ignore the negative information, while two third choose not to react to the positive information. Conditional of reacting to the negative information half of the consumers choose a sophisticated strategy. Based on the initial estimations a Probit is estimated to identify consumer characteristics that seem important for the choice of strategy. We find that especially age and education are positively correlated with the probability of choosing a sophisticated strategy while the volume share for fish heavily influences the probability of reacting to both negative and positive information. To our knowledge this is the only micro-data study analysing differences in consumers' information processing strategies.

Introduction

Consumption of food typically satisfies needs such as hunger or thirst and generates pleasures of taste that are directly observed by the consumer. In addition food contains healthy nutrients and in some cases also detrimental compounds with a negative effect on health. These are not observed directly at the time of consumption and therefore consumers may be uncertain about the underlying cause-effect relationships between the food they consume and the health effects. Their beliefs about these relationships may therefore be affected by information provided through news media, advertising, labelling schemes, information campaigns, etc. Often both positive and negative health effects of the same food are communicated to consumers at the same time through various media and with various means. Food safety scares in connection with the revelation of high risk detrimental compounds in foods like e.g. BSE and salmonella have in some cases had dramatic effects on food consumption (Verbeke et al., 2000; Verbeke and Ward, 2001; Piggott and Marsh, 2004; Smed and Jensen, 2005). One might therefore suspect that also information about nutrients and other compounds with long-term cumulative positive and negative health effects may influence consumer behaviour, but that the reaction patterns are different from information about compounds with short term health effects. At any rate, the practice of actively informing consumers about the long-term cumulative positive and negative health affects of the dietary choices is widespread and thus understanding when and how different types of information affect consumers seems increasingly relevant for policy-makers as well as for producers, retailers and marketing strategists.

The idea that information and knowledge play a crucial role in food demand is not new in economics. There is a substantial literature investigating how various types of exogenous information about long-term cumulative positive and negative health effects influence the demand for foods. A number of studies using aggregate time-series data find significant evidence that public information campaigns, the spill-over effect from scientific articles and mass media stories about long-term cumulative health effects influence food demand (Kim and Chern, 1999; Brown and Schrader, 1990; Chang and Kinnucan, 1991; Rickertsen et al., 2003). Generic advertising seems to have little or no effect (Rickertsen, 1998; Piggott et al., 1996; Kinnucan et al., 2003) while brand advertising has a large effect both alone and in combination with generic advertising (Chang and Kinnucan, 1991; Tellis, 1988). The idea that social and demographic characteristics of the individual affect the acquisition of information and its effects on behaviour is put forward in theoretical articles, (for example, Becker, 1965; Grossman, 1972) and supported in a number of empirical studies based on cross-sectional data. These find substantial heterogeneity across consumers both in self-reported use of health information labels (Guthrie et al., 1995; Nayga, 1996, 2001) and in reaction to information (Variyam et al., 1996; Ippolito and Mathios, 1995; Chern and Zuo, 1995). Several studies also find that negative information typically influences demand more than positive information (Kinnucan et al., 2003; Fox et al., 2002; Smith et al., 1988; Mizerski, 1982). To our knowledge there is only one other study using micro-panel data (Verbeke and Ward, 2001).

A framework for understanding why the effects of information are correlated with consumer heterogeneity and may vary with positive and negative information has been suggested by Verbeke (2005a). The basic idea is that consumers weigh the costs of acquiring and processing information against the expected gains from optimising their food consumption in accordance with this information. This can be interpreted as allowing consumers to be 'rationally' ignorant as suggested by Swinnen et al. (2005). Most of the theoretically and empirically observed heterogeneities in reaction to information cited above seem in line with the basics of this framework. For example, better educated or more health conscious consumers more often use nutrition labels and this may be because of lower costs of processing information or greater expected health benefits upon reacting. One of the more intriguing implications of this framework, drawing upon the psychological literature, suggests that consumers in daily routine purchasing situations are typically guided by a heuristic information processing strategy ("a rule-of-thumb" strategy) as opposed to the systematic information processing strategy used in situations of great relevance to the consumers.

Our point of departure is how news in the media about the positive and negative health effects of fish consumption influences demand for fish. Information about the positive nutritional effects of fish applies generally to both fatty and lean fish. Information about the negative health effects of fish consumption comes down to information about dioxin in fish. Fatty fish are substantially more susceptible to dioxin poisoning than lean fish since dioxin accumulates in fatty tissue. Some consumers may know the difference between fatty and lean fish types and understand that it is possible to avoid the risk of dioxin poisoning by substituting away from fatty fish and toward lean fish (sophisticated reaction). Other consumers may not be able to make the distinction between fatty and lean fish and may, if they choose to react to dioxin information, instead substitute away from all types of fish (unsophisticated reaction). A third way of reaction may be to ignore the information (ignorance). Our evidence is based on a household level dataset that combines data on food purchases with indices of news media stories. Specifically, we follow 2500 households for six years and analyse their weekly fish and meat purchases over this period and how it is influenced by new information. To do this we derive and estimate two equations for each household: the budget share of fatty fish in total fish and the share of all fish in total meat and fish. New information is treated as an adjustment to the prices. In addition we are able to identify a number of consumer characteristics that seem important for choice of reaction pattern. To our knowledge this is the first empirical evidence from a micro-econometric study of differences in sophistication levels of consumers' information processing. Our empirical model fits within a theoretical framework of information processing strategies and therefore it may well apply more generally than just in relation to fish consumption and as such have broader relevance. It may in particular be relevant to consider information processing strategies when designing and implementing health campaigns and other attempts to regulate food consumption.

The rest of this paper is organised as follows: Section 5.1 contains the theory model, 5.2 the empirical specification. Section 5.3 is a description of the data and the fish market, while section 5.4 is an estimation section. Section 5.5 contains the results while section 5.6 is devoted to a discussion and conclusion.

5.1 The theory model

We assume that it is meaningful to think of fish consumption in terms of two aggregated types: lean and fatty fish. The consumption of fish presumably satisfies hunger and generates pleasures of taste etc.; attributes that we assume are captured by a basic fish characteristic called *taste*.[52] We assume that consumers can ascertain the content and experience the full utility value of this characteristic in connection with consumption and use the quantity of consumed fish (q_f and q_l, respectively) as proxies for the consumed amounts of taste. Since lean and fatty fish differ in many ways we assume that the taste characteristics connected to each type of fish are imperfect substitutes. In addition, fish provide nutrients with cumulative long-term health benefits. The health benefits of fish and seafood are well documented and widely promoted by e.g. nutrition experts in recent years. Fish provide the body with essential

[52] Taste is just an appellation and the characteristic captures much more than just mere taste. We assume that the characteristic contains all the utility derived directly from fish consumption in connection with satisfaction of hunger pleasures of taste, preparing the meal etc.

vitamins and minerals including, vitamins A, B, and D; iodine; selenium and of course protein (Bender, 2002; Astrup et al., 2005). Omega-3 fatty acids found in fish are also beneficial, particular in terms of cardiovascular health (Domingo, 2007; Sidhu, 2003). Furthermore, the consumption of fish is often encouraged in an obesity preventing diet, since fish is relatively low in saturated fats and is a healthier alternative to meat. This is captured by a *nutrition* characteristic. Consumers do not observe or experience utility of this characteristic in connection with consumption, but estimate the content of the characteristic and its ultimate utility value based on, among other things, information from television and newspapers. We assume that consumers expect both types of fish to contain the same quantity of the nutritional attribute (i.e. nq_f and nq_l, respectively, where n is the perceived utility value of nutritional health per volume of fish)[53] and the nutritional health advantage from each type of fish are assumed to be close substitutes. Finally, fish often live, feed, and breed in environments polluted by toxic compounds such as mercury, polychlorinated biphenyls (PCB's) and dioxin. These compounds accumulate in the food chain (mercury) or are bound in fatty tissue (PCB's and dioxins) (Pompa et al., 2003; Sidhu, 2003; Astrup et al., 2005). This we assume is captured in a *dioxin* characteristic associated with long-term health disadvantages that, in the same way as nutrition, are unobserved at the time of consumption. Even though the detrimental effect of dioxin consumed from fatty and lean fish is the same, the dioxin content is much lower in lean than in fatty fish (i.e. $d_f q_f$ and $d_l q_l$, respectively, where d_f and d_l is the perceived utility value of dioxin in a volume of fatty and lean fish respectively).[54] When thinking of how these different characteristics affect consumers' utility and ultimately their behaviour it seems natural that the *taste* characteristics in their broad sense are closely related, and that these again are more closely related to the *taste* characteristics of meats than to other types of consumer goods[55]. Finally, it seems that the immediate pleasures of taste are probably not closely related to the perceived long-term utility derived from the nutritional health and dioxin characteristics. We therefore assume that the following basic utility structure applies for consumers of fish.

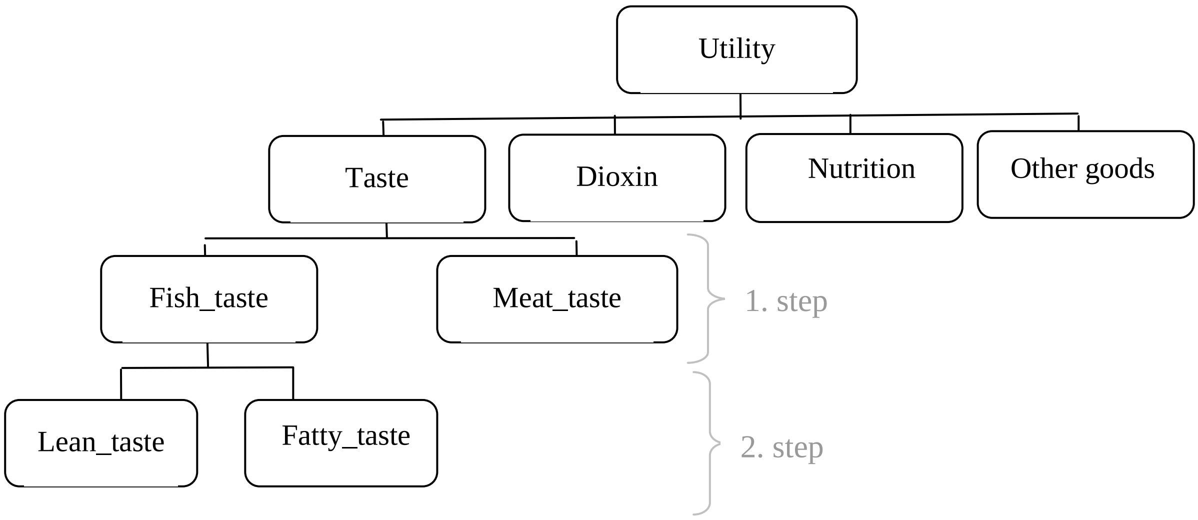
---

[53] There may be differences in the health content of fish. Fatty fish contain more of the healthy Omega3 fatty acids. But fatty fish also contain more energy than lean fish so in an obesity preventing perspective lean fish is more healthy than fatty fish. To consider the nutrition characteristic as different for fatty and lean fish will be a route of further research

[54] There is a corresponding difference in mercury content between predatory and non-predatory fish (Pompa et al., 2003; Sidhu, 2003) when there is information about the accumulation of mercury in predatory fish. We disregard this twist since predatory fish only account for a small proportion of fish consumption (under 5 per cent) and there is only scattered media information about mercury in our data period.

[55] The meat-fish structure could easily be enlarged to comprise other types of food.

Figure 5.1: Basic utility structure for consumption of health and taste characteristics



To capture this formally let the consumer be characterised by a utility function (*U*) defined on the two types of taste characteristics (where q_f and q_l are the quantities of fish consumed), the two types of health (nutrition and dioxin*)*, an aggregate of meats (q_M) that is only endowed with a taste characteristic and finally an aggregate of other goods (q_{OG}). We assume that the meat and fish taste characteristics are separable from the health characteristics and other goods and that the fish taste characteristics are separable from the meat taste characteristics. Thus the budget-constrained consumer problem becomes[56]:

$$\begin{aligned} &\underset{q}{Max}\, E\left[U(u_T\left(u_F(q_f,q_l),q_M\right),u_n\left(nq_f+nq_l),u_d(d_fq_f+d_lq_l\right),q_{OG})\right] \\ &S.T. \qquad p_l\cdot q_l+p_f\cdot q_f+p_M\cdot q_M+p_{OG}\cdot q_{OG}\leq x \end{aligned} \tag{5.1}$$

p_l is the price of lean fish, p_f is the price of fatty fish, p_M is a price index of the meat aggregate and p_{OG} some price index of the other goods aggregate. The consumers first order conditions are:

[56] Note here that consumers get utility from characteristics, but maximise over goods.

$$
\begin{aligned}
&\frac{\partial U}{\partial u_T}\frac{\partial u_T}{\partial u_F}\frac{\partial u_F}{\partial q_f}+\frac{\partial U}{\partial u_n}n-\frac{\partial U}{\partial u_d}d_f=\lambda p_f\\
&\frac{\partial U}{\partial u_T}\frac{\partial u_T}{\partial u_F}\frac{\partial u_F}{\partial q_l}+\frac{\partial U}{\partial u_n}n-\frac{\partial U}{\partial u_d}d_l=\lambda p_l\\
&\frac{\partial U}{\partial u_T}\frac{\partial u_T}{\partial q_M}=\lambda p_M\\
&\frac{\partial U}{\partial q_{OG}}=\lambda p_{OG}
\end{aligned}
\tag{5.2}
$$

which gives the following first order conditions for the demand for food taste characteristics:

$$
\begin{aligned}
&\frac{\partial U}{\partial u_T}\frac{\partial u_T}{\partial u_F}\frac{\partial u_F}{\partial q_f}=\lambda(p_f-\tilde{n}+\tilde{d}_f)\\
&\frac{\partial U}{\partial u_T}\frac{\partial u_T}{\partial u_F}\frac{\partial u_F}{\partial q_l}=\lambda(p_l-\tilde{n}+\tilde{d}_l)\\
&\frac{\partial U}{\partial u_T}\frac{\partial u_T}{\partial q_M}=\lambda p_{OF}
\end{aligned}
\tag{5.3}
$$

where $\lambda=\frac{\partial U}{\partial q_{OG}}/p_{OG}$, $\quad \tilde{n}=\frac{\partial U}{\partial u_n}n/\lambda,\quad \tilde{d}_f=\frac{\partial U}{\partial u_d}d_f/\lambda,\quad \tilde{d}_l=\frac{\partial U}{\partial u_d}d_l/\lambda$

Our focus here is on the consumption of fish, which for the typical Danish household only accounts for a small part of the total budget for foods (below 5 per cent). The effect of changes in fish demand on the shadow price of funds (λ) is therefore presumably small. Further, nutrition and dioxin are long-run health effect so even though fish consumption in the current period determines the flow of the long-run health effects in the current period, marginal health values $\left(\frac{\partial U}{\partial u_n},\frac{\partial U}{\partial u_d}\right)$ presumably depend on the accumulated stocks that are not sensitive to current flows. Since we must introduce functional structure at some point it does not seem blatantly unreasonable to assume that $\tilde{n},\tilde{d}_f,\tilde{d}_l$ are unaffected by variations in fish demand within the span of our data. By assuming that the marginal utility of health is independent of consumed fish quantities and expenditure these effectively enter as an exogenous correction to the good’s price in the resulting demand function. This allows us to utilise standard separability results when deriving demand functions (without making other functional form assumptions at this point) while retaining an explicit and consistent representation of the health attributes that are of primary interest. With the assumed separability of fish taste characteristics we therefore have the following Marshalian demands for fish taste characteristics in the second step:

$$\begin{aligned} q_f &= q_f(\tilde{p}_f, \tilde{p}_l, \tilde{x}_F) \\ q_l &= q_l(\tilde{p}_f, \tilde{p}_l, \tilde{x}_F) \end{aligned} \tag{5.4}$$

where $\tilde{x}_F$ is expenditure on fish taste characteristics and $\tilde{p}_f = p_f - \tilde{n} + \tilde{d}_f$ and $\tilde{p}_l = p_l - \tilde{n} + \tilde{d}_l$ are the net price of fish after adjustment for the utility value of the consumers perceived dioxin and nutritional health content of fish. It is important to remember here that $\tilde{x}_F$ is *not* equal to the total expenditure on fish – but is the part of expenditure on fish that is allocated to purchase of the taste characteristic. This carries over to the first step where demand for the fish taste aggregate becomes a function of the aggregated health corrected price $P_F(\tilde{p}_f, \tilde{p}_l)$. This gives the following first step Marshalian demand functions:

$$\begin{aligned} q_F &= q_F(P_F(\tilde{p}_f, \tilde{p}_l), p_M, \tilde{x}_{MF}) \\ q_M &= q_M(P_F(\tilde{p}_f, \tilde{p}_l), p_M, \tilde{x}_{MF}) \end{aligned} \tag{5.5}$$

where $\tilde{x}_{MF}$ is expenditure on all meat-fish taste characteristics, i.e. net of expenditure on fish health characteristics.

5.2 Empirical specification

The demand model

We assume the AIDS specification $A(\cdot)$ for taste characteristics demand in (5.4) and (5.5) where adding up constraints reduces the system to two equations:

$$q_f = A_f\left(\tilde{p}_f, \tilde{p}_l, \frac{\tilde{x}_F}{\tilde{P}_F^A(\cdot)}\right)\frac{\tilde{x}_F}{\tilde{p}_f} \tag{5.6}$$

$$q_F = A_F\left(\tilde{P}_F^C(\cdot), p_M, \frac{\tilde{x}_{MF}}{\tilde{P}_{MF}^A(\cdot)}\right)\frac{\tilde{x}_{MF}}{\tilde{P}_F^C(\cdot)}$$

where $\tilde{P}_F^C(.)$ is a consumer price index for the fish taste aggregate and $\tilde{P}_F^A(\cdot)$ and $\tilde{P}_{MF}^A(\cdot)$ are the AIDS price indices for fish taste and the meat-fish taste aggregates, respectively. This would generate equations in taste characteristic budget shares $\tilde{w}_f = \frac{q_f \tilde{p}_f}{\tilde{x}_F}$ and $\tilde{w}_F = \frac{q_F \tilde{P}_F^C(\cdot)}{\tilde{x}_{MF}}$.

Only observing uncorrected budgets, prices and budget shares we approximate

$\tilde{x}_F = x_F \frac{\tilde{P}_F^C(\cdot)}{P_F^C(\cdot)}$ and $\tilde{x}_{MF} = x_{MF} \frac{\tilde{P}_{MF}^C(\cdot)}{P_{MF}^C(\cdot)}$ where $P_F^C(\cdot)$ is a price index for the fish good aggregate, $P_{MF}^C(\cdot)$ and $\tilde{P}_{MF}^C(\cdot)$ are consumer price indices for the meat-fish good and taste aggregates, respectively. Multiplying by uncorrected prices (p_f and $P_F^C(\cdot)$ respectively) this gives us demand equations in observed uncorrected budget shares:

$$w_f = A_f\left(\tilde{p}_f, \tilde{p}_l, \frac{x_F}{\tilde{P}_F^A(\cdot)} \frac{\tilde{P}_F^C(\cdot)}{P_F^C(\cdot)}\right) \cdot \frac{p_f}{\tilde{p}_f} \cdot \frac{\tilde{P}_F^C(\cdot)}{P_F^C(\cdot)} \tag{5.7}$$

$$w_F = A_F\left(\tilde{P}_F^C(\cdot), p_M, \frac{x_{MF}}{\tilde{P}_{MF}^A(\cdot)} \frac{\tilde{P}_{MF}^C(\cdot)}{P_{MF}^C(\cdot)}\right) \cdot \frac{P_F^C(\cdot)}{\tilde{P}_F^C(\cdot)} \cdot \frac{\tilde{P}_{MF}^C(\cdot)}{P_{MF}^C(\cdot)}$$

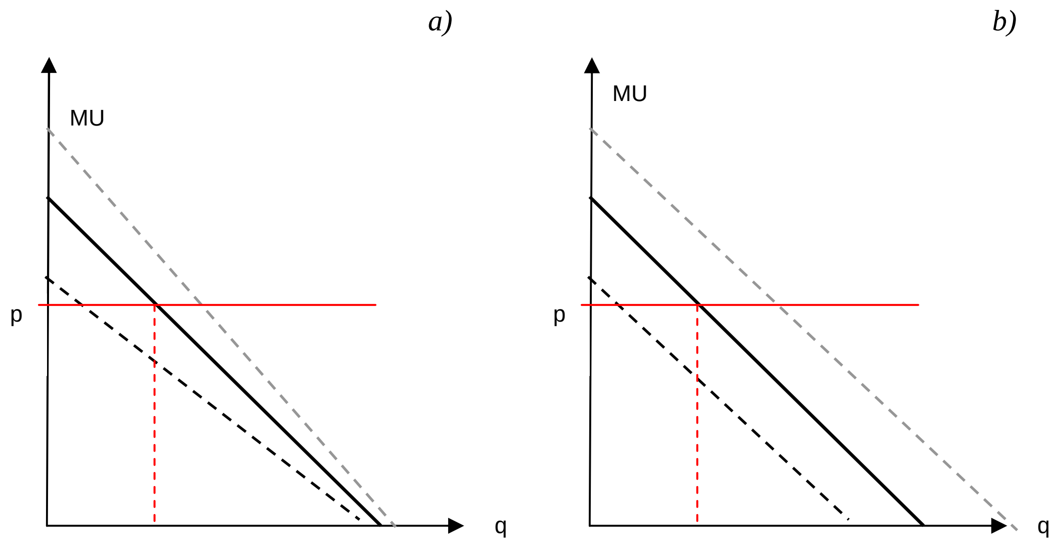
Finally, to get a manageable setup we approximate both the consumer and AIDS indices with the Stone index for both the fish taste and meat-fish taste aggregates. After inserting the Stone index formula (e.g. $P_F^S(\tilde{p}_f, \tilde{p}_l) = \alpha(\tilde{p}_f)^{w_f}(\tilde{p}_l)^{1-w_f}$, see Appendix 5.A for derivation) and the definition $x_F = w_F x_{MF}$ we get:

$$w_f = A_f(\tilde{p}_f, \tilde{p}_l, \frac{w_F x_{MF}}{(p_f)^{w_f}(p_l)^{1-w_f}}) \left(\frac{(p_f)(\tilde{p}_l)}{(\tilde{p}_f)(p_l)}\right)^{1-w_f} \tag{5.8}$$

$$w_F = A_f(P_F^S(\tilde{p}_f, \tilde{p}_l), p_M, \frac{x}{(p_f)^{w_F w_f}(p_l)^{w_F(1-w_f)}(p_M)^{1-w_F}}) \left(\frac{(p_f)^{w_f}(p_l)^{1-w_f}}{(\tilde{p}_f)^{w_f}(\tilde{p}_l)^{1-w_f}}\right)^{1-w_F}$$

This simultaneous system describes the demand for the taste characteristic of the two types of fish and meat while having consistently adjusted for the consumers' evaluation of the two health characteristics – in particular the interactions between the nests are modelled consistently. In addition to health being removed from the taste characteristics in the utility structure the most critical implication of the assumed model is probably that of the functional form for the utility of the two health effects. In many of the papers cited in the introduction, information is added to the shift parameter in the AIDS specification (see e.g. Verbeke and Ward, 2001; Nichèle, 2003; Rickertsen et al., 2003) which is more tractable than our approach. This would in our characteristics context imply that the health characteristics influence utility as illustrated in Figure 5.2a below. The solid black line is marginal utility valuation of taste that declines as the quantity consumed during the current time period (e.g. month or quarter) increases. The black dot-and-dash line indicates total utility net of perceived disutility of dioxin consumed and the grey dot-and-dash line the utility of nutrition consumed during the same period. The difference between the dot-and-dash lines and the solid lines illustrate that the marginal disutility of dioxin and the marginal utility of nutrition would decline over the quantity consumed. Furthermore, this would imply perfect substitution between the taste and health characteristics of each type of fish. Neither of these two implications seems reasonable in the case of dioxin, but might be accepted in the case of nutrition. Figure 5.2b illustrates the functional shape implied with our setup, which seems a lot more reasonable in the case of dioxin and can be accepted in the case of nutrients.

Figure 5.2: Functional form for the utility of the long-term health effects



Consumers have to estimate the content and ultimate the utility value of the two long-term credence health characteristics when deciding their demand for fish. We presume that they might use media news and other current information flows to inform their estimates. In fact there has been a steady flow of newspaper stories and television news about both nutritional benefits and dioxin disadvantages of fish consumption with varying intensity over time. Thus we are looking at media information flows that presumably could affect behaviour at the margin – but probably does not imply major shifts in consumer beliefs about food safety like, for example did the BSE case. Typically news stories about nutritional benefits state that these apply to all fish while stories about dioxin typically state that this mainly is a problem for fatty fish types thus reflecting the established truths about these health effects.

The information model

Verbeke (2005a) has suggested a framework for understanding how information could affect consumers, drawing on both economic and psychological literature. The basic idea (Swinnen et al., 2005) suggests that consumers weigh expected costs and benefits of acquiring and processing information when deciding if and how to use information. Thus consumers may decide not to use the information if the expected benefits from getting the information are small or the perceived costs of processing are large (i.e. it may be rational for a consumer not to use all the information flowing to him). An important twist on this basic idea draws on the psychological literature (Chen and Chaiken, 1999; Petty and Cacioppo, 1986) suggesting that consumers may chose between a heuristic and a systematic information processing strategy. Systematic processing implies extensive investigation and detailed exploitation of the information and takes place when the information relates to an issue of significant importance to the consumer. Heuristic processing on the other hand is based on simple decision rules and uses the information only superficially. Though this does not exploit the full potential of the available information it allows the consumer to make fast decisions without extensive processing costs and so to derive some benefit from information in situations where the stakes involved are limited or the amount of information is large. A number of studies suggests that daily routine food purchasing typically is guided by heuristic information processing as opposed to purchasing of consumer durables such as cars where systematic processing is common (see e.g. Verbeke 2005a). Systematic processing in connection with food purchases is probably mainly seen in trial situations, for example first purchases of new foods. Massive information signalling an important change in, for example food safety, such as the BSE scare might also cause some consumers to undertake a systematic information search. In the following we assume that consumers have weighed the expected costs and benefits of acquiring and processing information concerning health attributes in fish and thereby have chosen their "strategy". This implies that the media information flows experienced by consumers are interpreted by heuristic information processing using rules of thumb. Since no major events concerning either the positive or negative health effects have occurred during our data period, we assume that the media information flows do not initiate a systematic information search. This means that we can assume that the rules of thumb used by a given consumer do not change. As the original cost and benefits of processing information depend on household characteristics such as age and education, the chosen strategy or rule of thumb does as well. Making this assumption in our 'business as usual'-situation seems unproblematic. When consumers use a heuristic information processing strategy a number of studies suggest that if information is to affect behaviour it must give immediate meaning to the consumer (see Guthrie et al., 1995; Nayga, 1996, 2001). Thus the information flow must interact meaningfully with the consumers' prior knowledge. If it does not do so chances are it will be ignored because the heuristic processing strategy is geared towards economising on processing costs.

In our case the consumers' prior knowledge about which types of fish are fatty and lean and whether this is operationalized through the strategy chosen when purchasing fish is critical. Some consumers may differentiate between fatty and lean fish and so may be able to avoid the risk of dioxin poisoning while maintaining the consumption of fish by substituting away from fatty fish and toward lean fish. These are called the sophisticated consumers. Other consumers may not be able to make the distinction. Some consumers with 'unsophisticated' rules of thumb may choose to ignore the information because the fatty/lean distinction does not fit into the decision rules they usually apply. However, other 'unsophisticated' consumers may choose to react to the dioxin information by ignoring the fatty/lean distinction, and instead substitute away from both types of fish towards meat. The way we will capture this is by assuming that information at time *t* can affect consumer evaluations of the health characteristics in the following way:

$$\begin{aligned} d_{ft} &= \gamma_f S_{dt} \\ d_{lt} &= \gamma_l S_{dt} \qquad (5.9) \\ n_t &= \gamma_n S_{nt} \end{aligned}$$

where we let S_d denote current stock of information about dioxin and S_n denote current stock of information about nutritional health. Thus we assume that a larger stock of information indicates that there is more of the characteristic in question in fish or at least the consumers perceive there is.[57] The gamma parameters reflect the households' individual marginal utility valuation of the two characteristics. The main point here is that qualitative differences in how

[57] If consumers were estimating say the alcohol content in a beer and each piece of information was signalling an alcohol content of 5 per cent it would be strange to assume that more information signalling 5 per cent caused consumers to increase there estimates of the alcohol percentage. However, the characteristics considered here are fluffy in the sense that consumers probably only have a vague idea of both content and utility implications. Assuming that they interpret more information about e.g. dioxin in fish as an indication that the problem is greater as initially expected does not seem far fetched especially when remembering that the stock calculation function may have upper and lower limits etc. so that there may be bounds within which the consumers' evaluation fluctuates.

consumers use information are revealed by comparing the γ_f and γ_l parameters. If consumers are sophisticated they know that there is no dioxin in lean fish and so set $\gamma_l = 0$ while γ_f reflects their utility valuation of dioxin. Unsophisticated consumers that do not distinguish will set $\gamma_f = \gamma_l > 0$. Finally, unsophisticated consumers who choose to disregard this type of information will set $\gamma_f = \gamma_l = 0$. Note that we have left some important issues to the specification of the information index and the identification discussion in the estimation section; for example, the relative weight ascribed by consumers to different sources (reflecting how credible they think the source is) and how much information has been gathered prior to the data period

5.3 Data and the fish market

Data

We perform the analysis using a household panel dataset from GfK-Denmark covering weekly purchase of food from approximately 2500 household in the period from 1st January 1997 to the 31st December 2002. This dataset includes daily household level registrations of volume and value of the purchase of more than 35 types of fish. These types of fish are divided into two types of fish; lean fish (including seafood) and fatty fish.[58] The data are aggregated to quarterly observations. Meat is divided into poultry, beef and pork. Prices for lean and fatty fish and for each of the meat types are constructed as average prices for all households in each region (Capital area, East and West Denmark) and each type of shop (supermarket, discount and speciality store)[59]. Then individual prices are weighted according to each household's share of purchase in each type of store[60]. The price of meat is then weighted together according to each household's budget share for each type of meat. We select on households being in the panel for at least 12 quarters buying both meat and fish in all periods. This leaves us with 1050 households. The dataset provides background variables for the households in the panel making it possible to analyse consumers choice of strategy depending on household characteristics such as age, education etc. The various data sources, as well as the linkages between them, are illustrated in Figure 5.3 below.

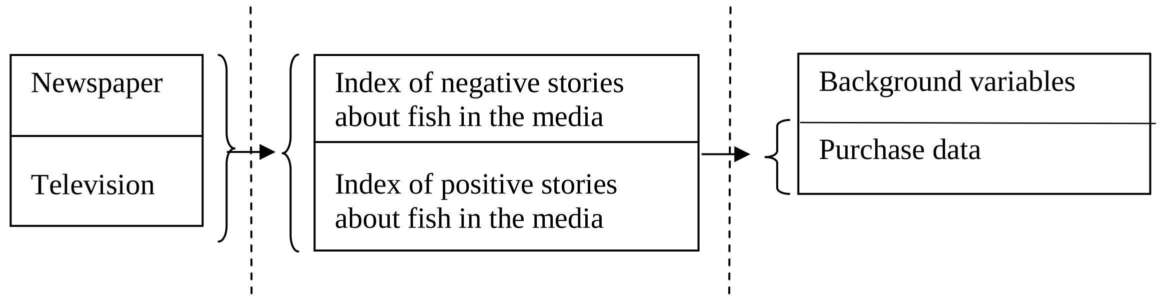
---

[58] Appendix 5.B shows the types of fish, price, volume share and aggregation category.

[59] Speciality stores are fishmonger or butcher.

[60] Prices vary considerably between type of stores. For an example see prices for lean and fatty fish in Appendix 5.C

Figure 5.3: Illustration of the different datasets and the linkages between them



By combining the purchase data with information indices on media coverage of the health consequences of fish consumption, household exposure to new information can be dated and the impact on demand followed closely. The indices are based on an extensive search in a database Infomedia, covering all types of articles in Danish newspapers. The search was limited to include the following widely read newspapers: Aktuelt, Berlingske Tidende, B.T., Ekstra Bladet, Fyens Stifttidende, Information, JyllandsPosten, Politiken, and Weekendavisen (covering most of the market). The search is based on the word “dioxin”, “health”, “fish-oil” and “omega” as search word in combination with “fish”. A search on “dioxin” and “fish” returns 107 hits while the positive search returned 148 hits. Besides articles brought in the written press a request for a search on "fish" is directed to the major Danish television stations: DR and TV2. Each article/feature has been read and the content is described (what it is about and in which week the news have been submitted) and each article/feature is then given a number to show that a news event has been submitted in a specific week for each media (newspaper or TV channel). This gives a number of time series specific for each media.

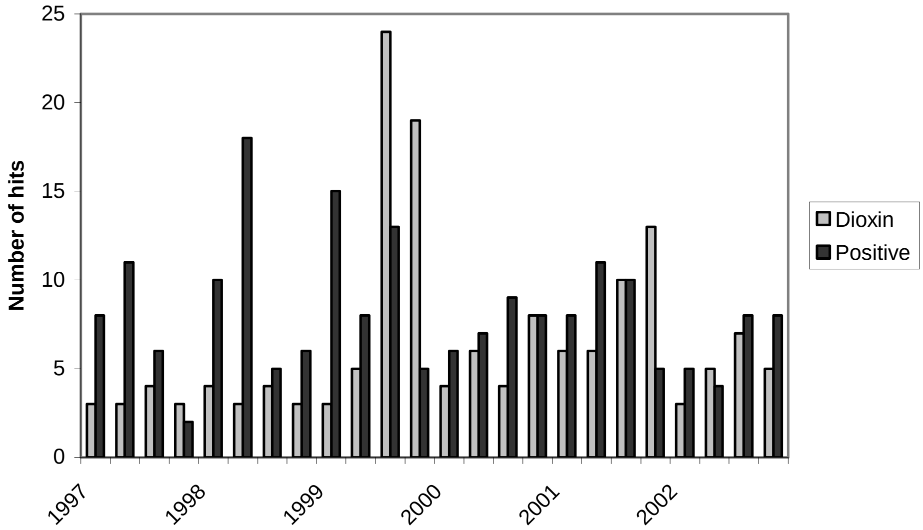
These times series have to be weighted into news indices. Several types of indices have been used in the literature to represent the effect of information, ranging from dummy or trend variables (Tansel, 1993), actual message numbers (Smith et al., 1988) cumulative message numbers (Brown and Schrader, 1990; Chang and Kinnucan, 1991) and cumulative message numbers with a decay (Chern and Zuo, 1995; Kim and Chern, 1999). Some of the indices discriminate between negative and positive news, some include lags and some of them make more complicated structures. Brown and Schrader (1990) introduce an index constructed as the cumulated number of published medical articles supporting a link between heart diseases and cholesterol intake representing the negative news, and the accumulated number of published medical articles questioning a link as positive news. An updated version of the

Brown and Schrader index is used in several studies including Kinnucan et al. (1997) and Chang and Kinnucan (1991). Chern and Zuo (1995) and Kim and Chern (1997, 1999) have made alternative indices based on the approach in Brown and Schrader where the effects of an article are assumed to diminish over time. The basic assumption behind these indices is that the information in these articles is transmitted down to the consumer through newspapers and TV. A more direct approach is used in Piggott and Marsh (2004), where the number of relevant newspaper articles is aggregated without any weights, and in McGuirk et al. (1995) and Schmit and Kaiser (2004), where a cumulative index based on articles (weighted by readership) in popular press periodicals is constructed similar to the Brown and Schrader index. Verbeke and Ward (2001) base a publicity index on TV coverage of the BSE problems together with advertising expenditure, while Smith et al. (1988) let their index be based on articles in major newspapers, weighting the newspaper articles by using the Budd's attention score[61]. This index is furthermore weighted by a probability that the articles are read (newspaper market share). Several of the indices introduced in the literature use a lag structure, as the studies find that press coverage has a cumulative effect (Verbeke and Ward, 2001; Kinnucan et al., 1997; Rickertsen et al., 1995). This includes simple declining shares to lagged index values like in Rickertsen et al. (1995) or more sophisticated structures as in Verbeke and Ward (2001).

In this study the index is made by aggregating the negative news (dioxin) and the positive news (fish oil, omega and health) in separate indices for each newspaper and for each of the two major Danish television stations. First of all, the news is considered to be equally important and is thus all weighted as one. Secondly, the indices are made by weighting the news according to the reading share of the specific newspaper. Television transmissions are not weighted since all households are assumed to have a TV. Furthermore, as our data are aggregated to quarterly observations a floating index is constructed assuming that each article lasts one quarter. That is, an article in the last part of a quarter will have a larger influence in the next quarter than in the current quarter. The number of hits for dioxin and positive news are shown in Figure 5.4.

[61] This system ranges newspaper articles according to their location in the newspaper.

Figure 5.4: Frequency of positive and negative news items over time



The fish market

Fish constitutes approximately 15 per cent of the total budget devoted to fish and meat. The mean budget share increased from 1997 to 2002, mainly due to increasing relative prices for fish as illustrated in Figure 5.5. Lean fish (including seafood) accounts for the largest share of the budget for fish. As indicated in Figure 5.6 the budget share of fatty fish has increased in part due to an increase in the consumption of salmon and trout.

Figure 5.5: Budget share and price indices for fish and meat

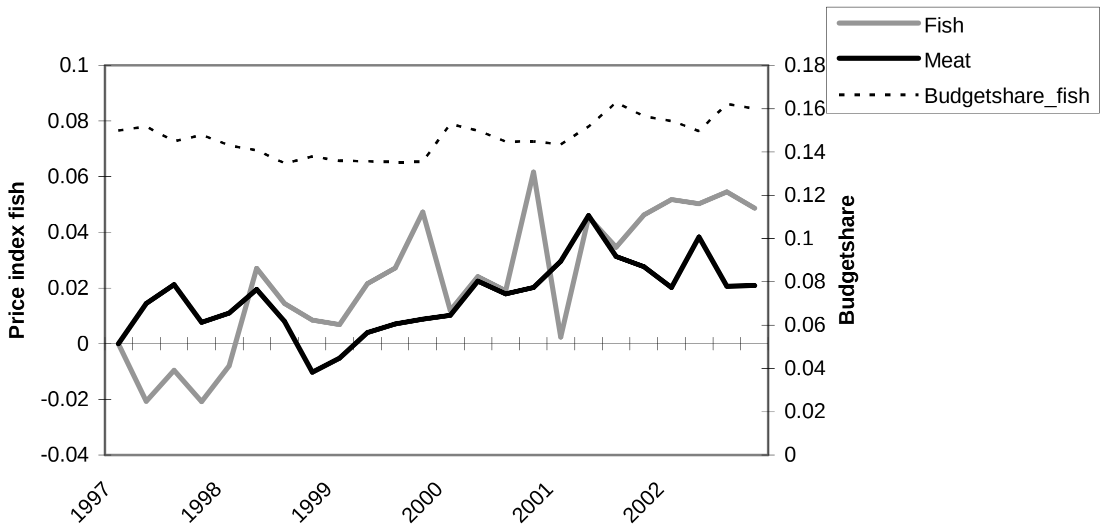
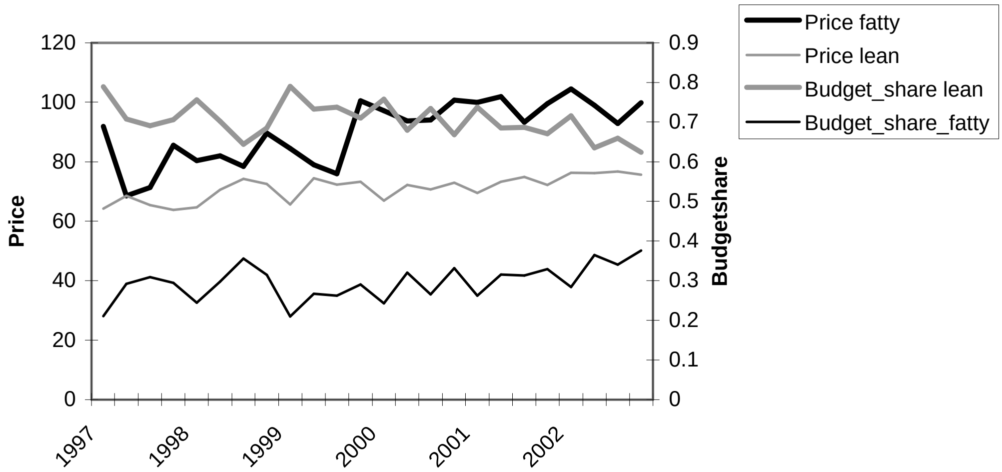
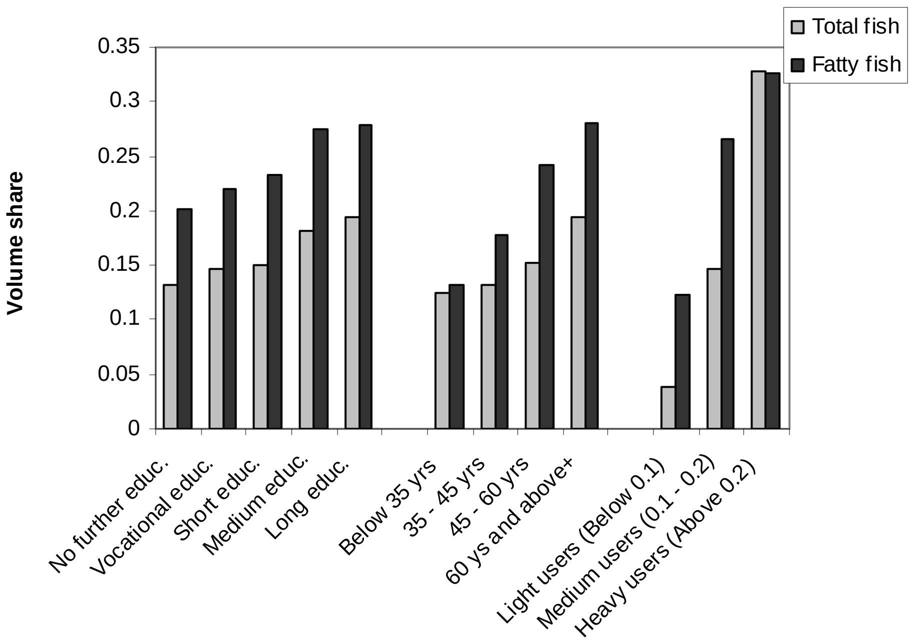


Figure 5.6: Budget shares and prices for lean and fatty fish



The total consumption of fish and the volume share of different types of fish between social and demographic groups vary considerable. Older and more educated households have a larger budget share for fish than the younger and less educated as illustrated in Figure 5.7, which shows the share of fish in the total fish-meat budget and the share of fatty fish in total fish. The budget share for fatty fish seems to follow the share of total fish closely, so whenever more fish is purchased a larger share of it is fatty fish.

Figure 5.7: Difference in fish consumption between social and demographic groups



5.4 Estimation section

Identification

We assume that only information in the current period affects the consumers' evaluations of characteristics and that current information S_{dt}^m is weighted by a credibility parameter c_m indicating how much confidence the consumer has in the media supplying the information.

$$S_{dt} = c_m S_{dt}^m \qquad (5.10)$$

We only observe current media information flows so inserting we get:

$$\begin{aligned} d_{ft} &= \gamma_f (c_m S_{dt}^m) \\ d_{lt} &= \gamma_l (c_m S_{dt}^m) \\ n_t &= \gamma_n \left(c_m S_{nt}^m \right) \end{aligned} \qquad (5.11)$$

We cannot use the exogenous variation to identify the γ_f and γ_l parameters in our model directly since the parameters we estimate are: $\gamma_f c_m$, $\gamma_l c_m$, $\gamma_n c_m$. Among other things this means that when there is a lack of reaction to information this could be because the credibility of the media is low. However, numerous surveys note that television and newspapers were the major information sources for the public, followed by radio, magazines, and other people (Bruhn et al., 1992; Chipman et al., 1995; Hoban & Kendall, 1993), i.e. indicating that the credibility of newspaper and television is rather large. For consumers who do in some way react to one of the information flows the parameters are identified (under the assumption made that the credibility parameter c_m applies to all information flows from the specific source). For these consumers relative values of the lambda parameters can be derived and interpreted.

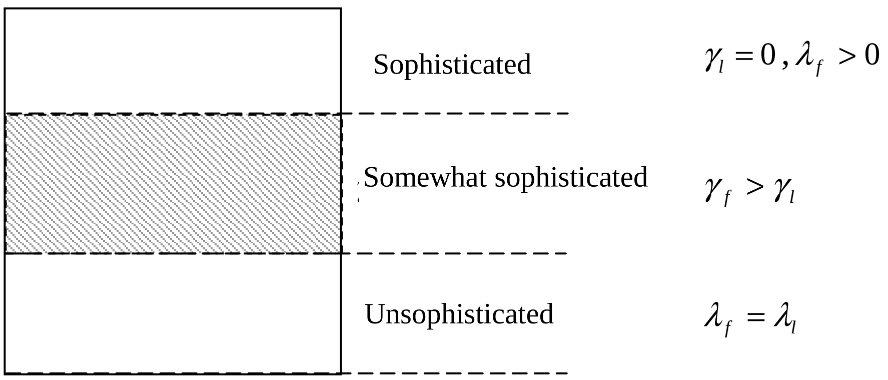
Estimation

To allow for maximal heterogeneity we estimate for each household separately. The budget for fish and meat is instrumented by the total expenditure on foods using the control function approach (Blundell and Powell, 2003). Households do not buy both lean and fatty fish in all periods, which means that the purchase of fish is censored. For the current analysis, households that buy both types of fish in less than 70 per cent of the periods observed are deleted from the dataset. This leaves us with 467 households who buy both types of fish at least 70 per cent of the time. Periods where the household does not buy both types of fish are deleted. This is not the correct way to handle the censoring problem, and as such this is a route for further research to include the correct censoring in the estimations. However, the final results have been compared to results based on estimation on the 219 households that buy both types of fish in all periods without change of the main conclusions. The problem with this reduced dataset is that the number of households is too small for the final Probit estimations. The prices for fish might be endogenous and should as such be instrumented. Usable instruments would be prices for fish in a nearby country. These have not been available for us.

Tests for information processing strategy

Conditional on reacting to negative information (i.e. that one or both of the γ_f and γ_l parameters are 'significantly' positive at the 5% level) the households are separated according to their information processing strategy. Households where it cannot be rejected that $\lambda_f = \lambda_l$ are categorised as unsophisticated, households where it cannot be rejected that $\gamma_l = 0$ and λ_f is significantly larger than zero are the sophisticated ones. Households where $\lambda_l < \lambda_f$ are the somewhat sophisticated.[62]

Figure 5.8: Illustration of information strategy categories



Finally, unsophisticated consumers who choose to disregard this type of information will set $\gamma_f = \gamma_l = 0$. Irrational behaviour is described by households that put $\gamma_l > \gamma_f$. When the households are categorised according to reaction to information a Probit is estimated. Since we estimate the model without any cross household parameter constraints we have not imposed restrictions that by definition generate correlation between choices of strategies

[62] This definition is somewhat more vague than the other definitions.

across households. We can therefore freely investigate different background variables that may be important in explaining heterogeneity in the consumer's choice of information processing strategy including the mean budget share of fish, region, price elasticities, indicator of the households being a heavy reader of newspapers etc.

5.5 Results

The basic model estimation results cannot be presented here since the model is estimated household by household for 467 households. An example of the detailed estimation results is given in Appendix 5.D. The remaining aggregated results are divided into two subsections: the first that concerns the choice of whether to react to information or not and the second that concerns the choice of strategy conditional on reacting to negative information.

Reaction to information

Some households have irrational behaviour (eat more fish when provided with information about dioxin in fish). They are deleted from the dataset leaving us with 447 households. Table 5.1 shows the number of households reacting to positive and negative information, respectively. Out of the 447 households 58 per cent reacts to either of the two. This number has to be compared with a survey from the European Commission where 53 per cent of the respondents state they have changed their consumption permanently or temporarily according to information from the media (European Commission, 2006). As many as 205 households react to negative information, while 143 react to positive information. 88 households react to both. Appendix 5.D shows examples of households reacting to both types of information (household 1 and 3) and only to negative information (household 2).

Table 5.1: Numbers of households reacting to negative and positive information

	Negative information		
Positive information	No reaction	Reaction	*TOTAL*
No reaction	187	117	*304*
Reaction	55	88	*143*
TOTAL	*242*	*205*	

As the model is estimated household by household the Probit can be modelled freely to isolate variables that have an effect on the choice of reacting to information. Two Probit models are estimated. First, a model with the reaction to negative information as dependent variable is estimated, and secondly, a model with the reaction to positive information as dependent variable.[63] Table 5.2 shows the parameter estimates for both models. In both models age and mean volume share for fish are treated as continuous variables. We include a variable for elastic demand (own price elasticity for fish below -1 versus own price elasticity for fish above -1), being a heavy reader of newspapers (households that read at least four out of six weekly editions), being located in the capital versus in other regions. The education variable is defined as either short educated (no or practical education) or theoretical education (households with a medium length or long education). Short education, capital, inelastic and "light" readers are base. Own price elasticity and the heavy reader variable becomes insignificant in both models. Age is negative in both models while the volume share for fish is positive. The parameters for theoretical education are negative for the reaction to negative information and positive for the reaction to positive information. The log likelihood chi squared statistics shows that the composite value of the independent variable differs from zero. The pseudo- R^2 values (McFadden, 1973; Estrella, 1998) are rather low, but reasonable for panel data.

Table 5.2: Parameter estimates for Probit

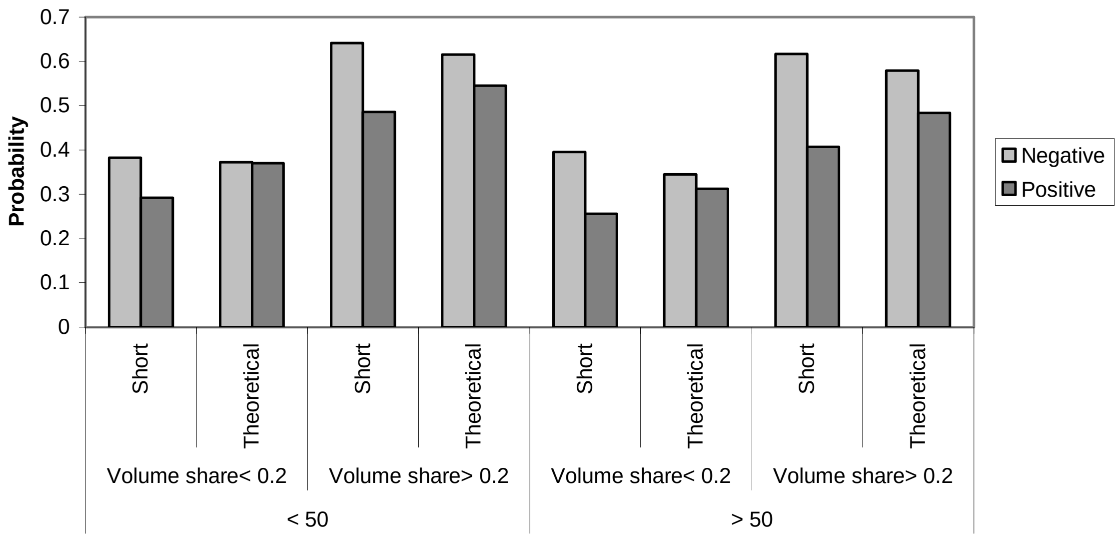
MODEL 1: Dependent variable: Reaction to negative information				
Parameter	**Estimate**	**Std. Error**	**t-Statistics**	**P-Value**
Age	-0.0188	0.00602	-3.12	0.0018
Other region	-0.3618	0.1838	-1.97	0.0491
Theoretical education	-0.229	0.2557	-0.9	0.3704
Volume share fish	4.3839	0.6784	6.46	<.0001
Heavy readers of newspapers	-0.0594	0.1769	-0.34	0.737
Elastic demand for fish (1. Step)	0.2524	0.3546	0.71	0.4766
	Log Likelihood value	-282.00511		
Pseudo R2 values	Log likelihood $\chi^2_{(df=6)}$ = 90.406 $(p < 0.001)$			
	Estrella	0.1085		
	McFadden's LRI	0.0795		

MODEL 2: Dependent variable: Reaction to positive information				
Parameter	**Estimate**	**Std. Error**	**t-Statistics**	**P-Value**
Age	-0.0211	0.006375	-3.32	0.0009
Other region	-0.0891	0.1852	-0.48	0.6303
Theoretical education	0.2031	0.2495	0.81	0.4156
Volume share fish	3.3509	0.7842	4.27	<.0001
Heavy readers of newspapers	-0.0502	0.1845	-0.27	0.7857
Elastic demand for fish (1. Step)	-0.1713	0.3804	-0.45	0.6525
	Log Likelihood value	-270.63748		
Pseudo R2 values	Log likelihood $\chi^2_{(df=6)}$ = 146.124 $(p < 0.001)$			
	Estrella	0.1579		
	McFadden's LRI	0.1166		

[63] It will be reasonable to look at the models as a bivariate Probit model and test for correlation between the error terms, but for the time being the model has been estimated as individual models assuming individual error terms.

Figure 5.9 shows the estimated Probit probabilities for reacting to negative and positive information calculated for each of the models 1 and 2. The estimated probabilities show that there is a general larger probability of reacting to negative information than to positive information. Being a heavy user (households where fish constitutes a share above 0.2 of the total volume of the fish-meat aggregate) also increases the probability of reacting to both negative and positive information. Furthermore, there is a larger probability of reacting if located in the capital than if located in other regions. Age also significantly decreases the probability and more so for the positive information than the negative information.

Figure 5.9: Social and demographic differences in probabilities of reacting to information



Choice of information strategy

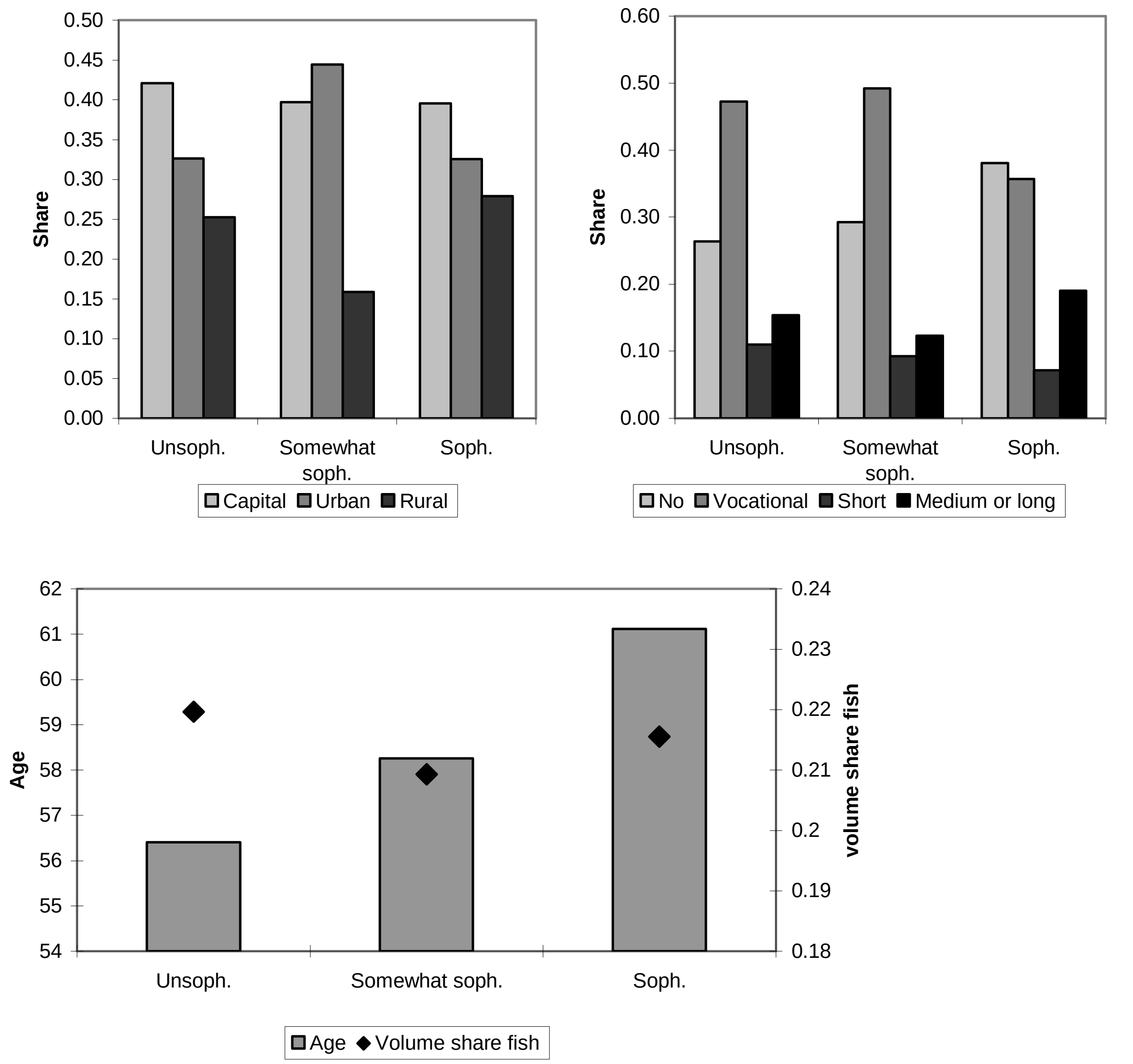
Since there is not any major events concerning either positive or negative health effects during our data period we assume that the media information flows do not initiate a systematic information search. This implies that the media information flows experienced by consumers are interpreted by the heuristic information processing using rules of thumb about how to react to information that they have chosen initially. If consumers choose to react to information, the possible strategies are either to be unsophisticated, i.e. to substitute away from fish when exposed to negative information or to be sophisticated i.e. to substitute between lean and fatty fish. Out of the 205 who react to the negative information only 6 react irrational. The sophisticated strategy is chosen by 43 (for an example see Appendix 5D household 3), the unsophisticated behaviour by 92 (for an example see Appendix 5D household 1), while 64 are somewhat sophisticated (for an example see Appendix 5D household 2).

Table 5.3: Number of households versus strategy choice

	Number of households
Sophisticated	43
Somewhat sophisticated	64
Unsophisticated	92
Irrational behaviour	6

In Figure 5.10 the profile of households choosing different strategies is shown. The figure indicates that the advanced strategy is preferred among rural households, households with medium or long education and with no education, older and households with a somewhat smaller volume share of fish. The crude strategy is preferred by vocational educated households, younger and households with a somewhat large volume share of fish.

Figure 5.10: Profile of households with different information strategies



* Unsoph=Unsophisticated, Soph=Sophisticated, Somewhat soph=Somewhat sophisticated

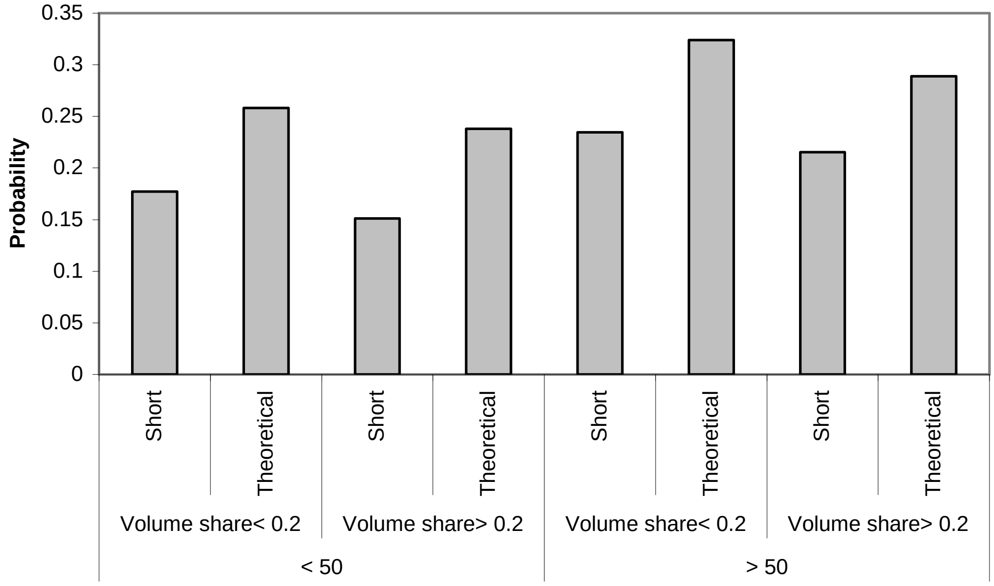
As above a binary Probit model is estimated in order to separate the effects from each other. In the model sophisticated is one choice of strategy and unsophisticated or somewhat sophisticated the other option. Table 5.4 shows the parameter estimates. Region and the newspaper variable becomes highly insignificant for the choice of strategy, the parameter for education is positive, but somewhat insignificant. That the elastic households have a smaller probability is evident since high price elasticity to a certain extent indicates a larger willingness to substitute between meat and fish. The volume share for fish is also highly insignificant indicating that the amount of fish consumed is a determinant for whether to react to information or not, but not a determinant for which strategy to choose. The log likelihood chi squared statistics shows that the composite value of the independent variable differs from zero. The pseudo-R^2 values (McFadden, 1973; Estrella, 1998) are higher than in model 1 and 2 and reasonable for panel data.

Table 5.4: Parameter estimates for Probit with strategy choice, conditional on reaction

MODEL 3: Dependent variable: Strategy choice				
Parameter	**Estimate**	**Std. Error**	**t-Statistics**	**P-Value**
Age	0.0230	0.1216	1.89	0.0583
Other region	-0.0786	0.2926	-0.27	0.7881
Theoretical education	0.4467	0.2495	1.79	0.1006
Volume share fish	-0.8037	1.3305	-0.60	0.5458
Heavy readers of newspapers	0.1574	0.3063	0.51	0.6072
Elastic demand for fish (1. Step)	-1.5844	0.4986	-3.18	0.0015
	Log Likelihood value	-100.97994		
Pseudo R2 values	Log likelihood $\chi^2_{(df=6)}$ =136.736 $(p < 0.001)$			
	Estrella	0.3325		
	McFadden's	0.2529		

Figure 5.11 shows the estimated Probit probabilities for being a sophisticated consumer. The own price elasticity for fish is here left out for the clarity of the figure. Older households have a larger probability of being sophisticated as education also tends to increase the probability of choosing a sophisticated strategy.

Figure 5.11: Probit probability of being sophisticated



5.6 Discussion and conclusion

This paper is, to our knowledge, the first micro-econometric study estimating differences in consumers' information processing strategies. Our approach is based on the assumption that consumers, prior to our data period, choose a strategy about how to react to information concerning unobservable health characteristics in fish. This initial choice is based on consumers weighing the costs of processing information against the expected gains from optimising their food consumption in accordance with this information. This initial decision will result in a "rule-of-thumb" strategy about how to react to information, which will persist until radical information is received. It turns out that some consumers choose to ignore information about dioxin and choose not to react at all. Other consumers choose a 'heuristic' or unsophisticated reaction by substituting away from fish consumption all together ignoring the distinction between lean and fatty fish. Yet another group of consumers realise this distinction and substitute away from fatty fish towards lean fish to avoid the dioxin (the sophisticated reaction). We find that approximately half of the consumers ignore the negative information, while two third choose not to react to the positive information. This is in accordance with the literature that generally states larger influence of negative information compared to positive information. Conditional on reacting to the negative information we find that approximately half of the consumers choose an unsophisticated strategy. We find that the probability of reacting both to negative and positive information increases with the volume share for fish. This is in accordance with our expectations since the more fish you consume the more relevant the information about the health attributes in fish will be. This is also in accordance with Pieniak et al. (2007) who define the enthusiastic fish consumers as those with the highest fish consumption and those who engage most in information search about fish. Heavy users are also assumed to have a larger knowledge of which types of fish are lean and which are fatty, so the budget share for fish is also assumed to be positive correlated with being sophisticated (the cost of processing the information becomes smaller due to the detailed knowledge). We find no significant correlation. We have no immediate explanation for this. But since we estimate the model without any cross household parameter constraints we can freely investigate different background variables that may be important in explaining heterogeneity in the consumer's choice of information reacting strategy. As the reason for this "unexplainable" result may lay in some unexplored variables the full exploitation of these degrees of freedom might be a route of further research. Education presumably decreases the costs of processing information and so we would expect education to be positively correlated with the probability of reacting as well as with the probability of having a sophisticated reaction strategy. Education is generally found to be insignificant for the choice of reacting to information. This is off hand surprising since one would expect that educated consumers more easily can process and decipher information. However, more educated consumers also have a substantial initial stock of information and this may tend to decrease the probability of reacting since additional information will not significantly changed the perceived healthiness of fish. This will be a route of further research. However, we find a positive correlation between education and the probability of choosing a sophisticated strategy. Age significantly decreases the probability of reacting and more so for the positive information than the negative information. This is in accordance with the hypotheses of older households to be more conservative. Pieniak et al. (2007) find the sceptic fish consumers, i.e. consumers with the lowest use of and trust in information about fish, older than the average fish consumer, which also might explain the declining probability of reacting with age. Older households may have more general knowledge about cooking, so they might have a larger probability of choosing an advanced strategy conditional on reacting, which is in accordance with our findings. The probability of being sophisticated is negatively correlated with being price elastic, which is reasonable since price elasticities to a certain extent might indicate the willingness to substitute from one type of food to another.

The differences in the sophistication level of consumers' information reaction strategies that we have found are substantial and seem to provide solid support for the heuristic information processing model of food consumers. Therefore, the results may be relevant in a broader context than just fish consumption. The way of categorising consumers into information strategies should be of general interest to policy makers and marketing strategists when designing and implementing health campaigns or other attempts to regulate or change food consumption behaviour. The need to accentuate the focus on the comprehensibility frame of the target groups is important since consumers only react to information which is relevant to them and choose a reacting strategy which is in accordance with this. The motivation of the "no-reactors" is important, but it is equally important to consider the strategies of the so-called "unsophisticated" reactors. In attempts of regulating food consumption behaviour it is necessary to take possible detrimental effects of their adverse reactions into account. In our case a campaign aimed at shifting consumption towards lean fish, which may be beneficial, may for a large group of consumers have the effect of reducing fish consumption as such – which all in all may be detrimental. This illustrates the importance of considering possible "unsophisticated" reaction before initiating public campaigns.

References

Astrup A., N.L. Andersen, S. Stender and E. Trolle (2005): *Kostrådene 2005*. Publication no. 36, Ernæringsrådet & Danmarks Fødevareforskning, Copenhagen

Becker G.S. (1965): A theory of the allocation of time. *The Economic Journal*, 75(299): 463-517

Bender D.A. (2002): The vitamins. In M.J. Gibney, H.H. Vorster and H.J. Kok (eds.): ´ *Introduction to human nutrition*. Oxford, Blackwell Science Ltd., 125-176

Blanciforti L. and R. Green (1983): An Almost Ideal Demand System Incorporating Habits: An Analysis of Expenditures on Food and Aggregate Commodity Groups. *The Review of Economics and Statistics*, 65(3): 511-515

Blundell R.W. and J.L. Powell. (2003): *Endogeneity in Nonparametric and Semiparametric Regression Models.* In M. Dewatripont, L.P. Hansen and S.J. Turnovsky (eds.): *Advances in Economics and Econometrics*, Chapter 8: 312-357 Cambridge University Press, Cambridge

Brown D.J. and L.F. Schrader (1990): Cholesterol information and shell egg consumption. *American Journal of Agricultural Economics*, 72(3): 548-555

Bruhn C., S. Peterson, P. Phillips and N. Sakovidh (1992): Consumer response to information on integrated pest management. *Journal of Food Safety*, 12(4): 315-326

Chang H-S. and H.W Kinnucan (1991): Advertising, information, and product quality: the case of Butter. *American Journal of Agricultural Economics*, 73(4): 1195-1203

Chen S. and S. Chaiken (1999): The heuristics-systematic model in its broader context. In S. Chaiken and Y. Trope (eds.): *Dual Process Theories in social psychology*. New York: Guilford Press, pp. 73-96

Chern W.S. and J. Zuo (1995): *Alternative measures of changing consumer information on fat and cholesterol*. Paper presented in the annual meeting of the American Agricultural Economics Association, Indianapolis, Indiana, 6-9 August

Chipman H., P. Kendall; G. Auld, M. Slater, T. Keefe (1995): Consumer reaction to a risk/benefit/option message about agricultural chemicals in the food supply. *Journal of Consumer Affairs*, 29(1): 144-163

Domingo J.L. (2007): Omega-3 fatty acids and the benefits of fish consumption: Is all that glitters gold? *Environment International*, 33(7): 993-998

Edgerton, D.L., B. Assarson, A. Hummelmose, I.P. Laurila, K. Rickertsen and P.H. Vale (1996): *The econometrics of demand systems with applications to food demand in the Nordic countries*. Kluwer Academic Publishers, London

Estrella, A. (1998): A new measure of fit for equations with dichotomous dependent variables. *Journal of Business and Economic Statistics*, 16(2): 198-205

European Commission (2006): *Special Eurobarometer Office for official publications of the European Communities*, Luxemburg

Fox, J.A., D.J. Hayes and J.F. Shogren (2002): Consumer preferences for food irradiation: how favourable and unfavourable descriptions affect preferences for irradiated pork in experimental auctions. *Journal of Risk and Uncertainty*, 24: 75-95

Gould, B.W. and H.C. Lin (1994): Nutrition information and household dietary fat intake. *Journal of Agricultural and Resource Economics*, 19(2): 349-365

Greene, W.H. (2003): *Econometric Analysis*, 5th ed., pp. 913f. Prentice-Hall, New Jersey

Grossman M. (1972): On the concepts of health capital and the demand for health. *Journal of Political Economy*, 80(2): 223-255

Guthrie J.F., J.J Fox, L.E. Cleveland and S. Welsh (1995): Who uses nutrition labeling and what effect does label use have on diet quality? *Journal of Nutrition Education*, 27(4): 163-172

Hoban T. and P. Kendall (1994): *Consumer attitudes about food biotechnology*. Final Project Report. North Carolina State University: Raleigh, NC

Ippolito P.M. and A.D. Mathios (1995): Information and advertising: the case of fat consumption in the United States. *The American Economic Review*, 85(2): 91-95

Kim S. and W.S. Chern (1997): Indicative measures of health risk information on fat and cholesterol for US and Japanese consumers. *Consumer Interest Annual*, 43(2):84-89

Kim S. and W.S. Chern (1999): Alternative measures of health information and demands for fats and oils in Japan. *Journal of Consumer Affairs*, 33(1): 92-109

Kinnucan H.W., W.H. Xiao, C.J. Hsia and J.D. Jackson (1997): Effects of health information and generic advertising on U.S. meat demand. *American Journal of Agricultural Economics*, 79: 13-23

Kinnucan, H.W., Ø. Myrland and L. Paudel (2003): Relative impacts of health information and advertising on commodity markets: US meats. In W.S. Chern and K. Rickertsen (eds.): *Health, nutrition and food demand*. CAB international Publishing, Cambridge, USA

McFadden D. (1973): Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.): *Frontiers in economics*. New York: Academic Press, pp. 105-142

McGuirk A., P. Driscoll, J. Alwang and L. Huang (1995): System misspecification testing and structural change in the demand for meat. *Journal of Agricultural and Resource Economics*, 20: 1-21

Mizerski, R.W. (1982): An attribution explanation of the disproportionate influence of unfavourable information. *Journal of Consumer Research*, 9: 301-310

Nayga, R.M. (1996): Determinants of consumers' use of nutritional information on food packages. *Journal of Agricultural and Applied Economics*, 28(2): 303-312

Nayga, R.M. (2001): *Consumer use of nutritional labels: Endogenous health knowledge and sample selection effects*. Paper presented at the 71st EAAE seminar "The food consumer in the Early 21st Century", April 19-20, Zaragoza, Spain

Nichèle, V. (2003): Health information and food demand. In W.S. Chern and K. Rickertsen (eds.): *Health, nutrition and food demand*. CAB international publishing, Cambridge, USA

Petty, R.E. and J.T. Cacioppo (1986): *Communication and persuasion: Central and Peripheral routes to attitude change*. New York: Springer

Pieniak Z., W. Verbeke, J. Scholderer, K. Brunsø and S.O. Olsen (2007): European consumers' use of a trust in information sources about fish. *Food quality and preferences*, 18(8):1050-1063

Piggott, N.E., J.A. Chalfant, J.M. Alston and G.R. Griffith (1996): Demand response to advertising in the Australian meat industry. *American Journal of Agricultural Economics*, 78(2): 268-279

Piggott, N.E. and T. Marsh (2004): Does food safety information impact U.S. food demand. *American Journal of Agricultural Economics*, 86(1): 154–174

Pompa G., F. Caloni and M.L. Fracchiolla (2003): Dioxin and PCB Contamination of Fish and Shellfish: Assessment of Human Exposure. Review of the International Situation. *Veterinary Research Communications*, 27(Supplement 1): 1573-7446

Rickertsen, K. (1998): The effects of advertising in an inverse demand system: Norwegian vegetables revisited. *European Review of Agricultural Economics*, 25(1): 129-140

Rickertsen K., J.A. Chalfant and M. Steen (1995): The Effects of Advertising on the Demand for Vegetables. *European Review of Agricultural Economics*, 22(4): 481-494

Rickertsen K., D. Kristofersson and S. Lothe (2003): Effects of health information on Nordic meat and fish demand. *Empirical Economics*, 28(2): 249-273

Schmit, T.M. and H.M. Kaiser (2004): Decomposing the variation in generic advertising response over time. *American Journal of Agricultural Economic*, 86(1): 139-153

Sidhu, K.S. (2003): Health benefits and potential risks related to consumption of fish or fish oil.*Regulatory Toxicology and Pharmacology*, 38(3): 336-344

Smed S. and J.D. Jensen (2005): Food safety information and food demand. *British Food Journal*, 107(3): 173-186

Smith, M.E., E.O. van Ravenswaay and S.R. Thompson (1988): Sales loss determination in food contamination incidents: An application to milk bans in Hawaii. *American Journal of Agricultural Economics*, 70(3): 513-520

Swinnen, J.F.M., McClusky J.J. and N. Francken (2005): Food safety, the media and the information market. *Agricultural Economics*, 32(s1): 175-188

Tansel, A. (1993): Cigarette demand, health scares and education in Turkey. *Applied Economics*, 25(4): 521-529

Tellis, G.J. (1988): Advertising exposure, loyalty and brand purchase: A two stage model of choice. *Journal of Marketing Research*, 25(2): 134 –144

Variyam, J.N., J. Blaylock and D. Smallwood (1996): A probity latent variable model of nutrition information and dietary fibre intake. *Journal of Agricultural Economics*, 78: 628-639

Verbeke, W. (2005a): Agriculture and the food industry in the information age. *European Review of Agricultural Economics*, 32(3): 347-368

Verbeke, W. and R.W. Ward (2001): A fresh meat almost ideal demand system incorporating negative TV press and advertising impact. *Agricultural Economics*, 25(2-3): 359-374

Verbeke, W., R.W. Ward and J. Viaene (2000): Probit analysis of fresh meat consumption in Belgium: Exploring BSE and television communication impact. *Agribusiness*, 16(2): 215-234

Verbeke, W., I. Sioen, Z. Pieniak, J.V. Camp and S.D. Henauw (2005b): Consumers' perception versus scientific evidence about health benefits and safety risks from fish consumption. *Public Health Nutrition*, 8(4): 422-429

Appendix 5.A: Derivation of LAIDS system with Stone index

Models that use the Stone index are called the linear approximated AIDS following Blanciforti and Green (1983). The Stone index is of the form $\alpha(\tilde{p}_f)^{w_f}(\tilde{p}_l)^{1-w_f}$ where we assume that $\alpha_0 = 0$.

The second step:

$$w_f = A_f\left(\tilde{p}_f, \tilde{p}_l, \frac{x_F}{P_F^S(\cdot)}\right)\frac{p_f}{\tilde{p}_f}\frac{\tilde{P}_F^S(\cdot)}{P_F^S(\cdot)} \Leftrightarrow w_f = A_f\left(\tilde{p}_f, \tilde{p}_l, \frac{x_F}{(p_f)^{w_f}(p_l)^{1-w_f}}\right)\frac{p_f}{\tilde{p}_f}\frac{(\tilde{p}_f)^{w_f}(\tilde{p}_l)^{1-w_f}}{(p_f)^{w_f}(p_l)^{1-w_f}}$$

$$\Leftrightarrow w_f = A_f\left(\tilde{p}_f, \tilde{p}_l, \frac{x_F}{(p_f)^{w_f}(p_l)^{1-w_f}}\right)\frac{(\tilde{p}_f)^{w_f-1}(\tilde{p}_l)^{1-w_f}}{(p_f)^{w_f-1}(p_l)^{1-w_f}}$$

$$\Leftrightarrow w_f = A_f\left(\tilde{p}_f, \tilde{p}_l, \frac{x_F}{(p_f)^{w_f}(p_l)^{1-w_f}}\right)\left(\frac{(p_f)(\tilde{p}_l)}{(\tilde{p}_f)(p_l)}\right)^{1-w_f}$$

$$\Leftrightarrow w_f = \left(\alpha_f + \beta_f \ln(\tilde{p}_f) - \beta_f \ln(\tilde{p}_l) + \theta_f\left(\ln(x_F) - w_f \ln(\tilde{p}_f) - (1-w_f)\ln(\tilde{p}_l)\right)\right)\left(\frac{(p_f)(\tilde{p}_l)}{(\tilde{p}_f)(p_l)}\right)^{1-w_f}$$

The first step:

$$w_F = A_f\left(P_F^S(\cdot), p_M, \frac{x}{P_{MF}^S(\cdot)}\right)\frac{P_F^S(\cdot)}{\tilde{P}_F^S(\cdot)}\frac{P_{MF}^S\left(\tilde{P}_F^S(\cdot), p_M\right)}{P_{MF}^S\left(P_F^S(\cdot), p_M\right)}$$

$$\Leftrightarrow w_F = A_f\left(\tilde{P}_F^S(\cdot), p_{OF}, \frac{x_{MF}}{\left(P_F^S(\cdot)\right)^{w_F}(p_M)^{1-w_F}}\right)\frac{P_F^S(\cdot)}{\tilde{P}_F^S(\cdot)}\frac{\tilde{P}_F^S(\cdot)^{w_F}(p_M)^{1-w_F}}{P_F^S(\cdot)^{w_F}(p_M)^{1-w_F}}$$

$$\Leftrightarrow w_F = A_f\left(\tilde{P}_F^S(\cdot), p_M, \frac{x_{MF}}{\left(P_F^S(\cdot)\right)^{w_F}(p_M)^{1-w_F}}\right)\frac{\left(P_F^S(\cdot)\right)^{1-w_F}}{\left(\tilde{P}_F^S(\cdot)\right)^{1-w_F}}$$

$$\Leftrightarrow w_F = A_f\left(\tilde{P}_F^S(\cdot), p_M, \frac{x_{MF}}{(p_f)^{w_F w_f}(p_l)^{w_F(1-w_f)}(p_M)^{1-w_F}}\right)\left(\frac{(p_f)^{w_f}(p_l)^{1-w_f}}{(\tilde{p}_f)^{w_f}(\tilde{p}_l)^{1-w_f}}\right)^{1-w_F}$$

$$\Leftrightarrow w_F = \begin{pmatrix} \alpha_F + \beta_F\left(w_f \ln(\tilde{p}_f) + (1-w_f)\ln(\tilde{p}_l)\right) - \beta_F \ln(p_M) \\ +\theta_f\left(\ln(x_{MF}) - w_F\left(w_f \ln(\tilde{p}_f) + (1-w_f)\ln(\tilde{p}_l)\right) - (1-w_F)\ln(p_M)\right) \end{pmatrix} \cdot \left(\frac{(p_f)^{w_f}(p_l)^{1-w_f}}{(\tilde{p}_f)^{w_f}(\tilde{p}_l)^{1-w_f}}\right)^{1-w_F}$$

Appendix 5.B: Fish types, aggregation category, volume share and price

Figure 5B.1: Volume shares, price and fatty/lean distinction

	Volume share	DKK/kg.	Seasonality	Aggregation Fatty/lean
Shrimps	18.44%	85.85	No	lean
Trout/salmon	13.80%	93.48	No	fatty
Plaice	12.32%	68.06	No	lean
Cod	7.33%	61.44	Yes	lean
Cream of fish	5.67%	47.60	No	lean
Herring	5.43%	38.39	Yes	fatty
Flounder	5.36%	43.64	No	lean
Mackerel	4.75%	49.82	Yes	fatty
Coalfish	4.04%	47.88	No	lean
Fillet of fish	2.54%	40.81	No	lean
Eel	2.32%	134.52	Yes	fatty
Rainbow trout	1.25%	69.98	Yes	fatty
Cod roe	1.02%	88.26	Yes	fatty
Garfish	0.76%	66.79	Yes	lean
Greenland halibut	0.72%	127.75	No	fatty
Lobster	0.66%	99.26	No	lean
Pollack	0.64%	33.97	No	lean
Grav lax	0.54%	136.22	No	fatty
Dab	0.53%	56.17	No	lean
Cuttlefish	0.49%	41.47	No	lean
Tuna	0.43%	110.58	No	lean
Mussel	0.41%	61.80	No	lean
Smear dab	0.39%	90.08	No	lean
Hake	0.30%	48.28	No	lean
Rockfish	0.30%	75.53	No	lean
Crayfish	0.25%	79.29	No	lean
Crabs	0.18%	79.72	No	lean
Lump sucker	0.16%	55.36	No	fatty
Shellfish mix	0.11%	72.34	No	lean
Spawn	0.09%	113.12	No	lean
Catfish	0.07%	92.67	No	lean
Haddock	0.06%	87.59	No	lean
Hoki	0.03%	69.58	No	lean
Oyster	0.03%	121.84	No	lean
Herring spawn	0.01%	139.69	No	Lean
Caviar	0.00%	410.92	No	lean

Appendix 5.C: Prices and share of fish bought in various stores

Figure 5C.1: Share of fish bought in various stores, selected fish types

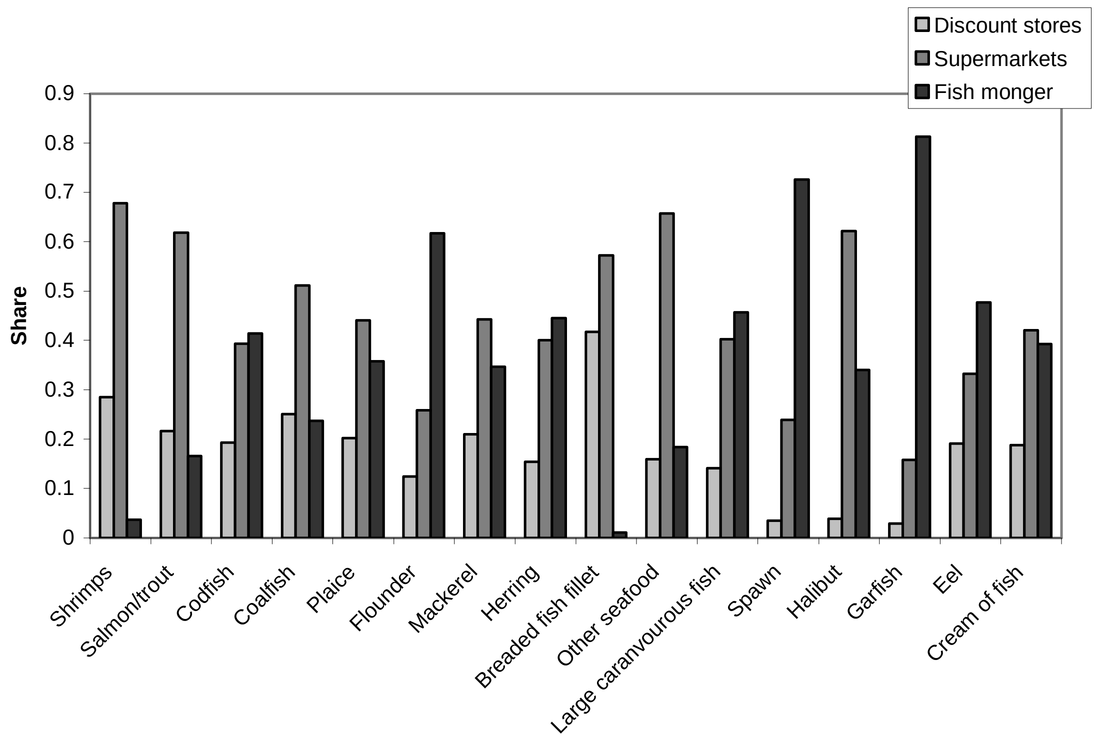
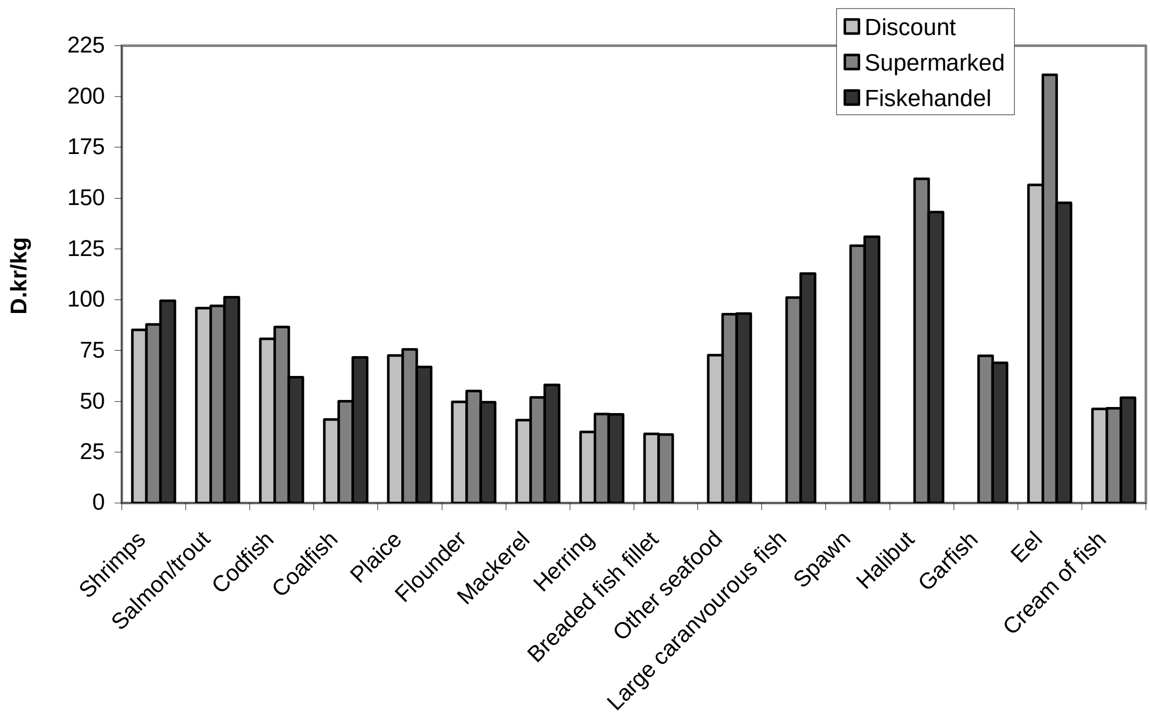


Figure 5C.2: Price of fish in various stores, selected fish types



Appendix 5.D: Example of estimation results

1) Example of an unsophisticated (FIML parameter estimated)

Budget share_fish, Adjusted R^2 = 0.7912, Budget share_fatty, Adjusted R^2 = 0.4649				
Parameter	Estimate	Std Err	t Value	Pr > |t|
α_{fish}	-0.1689	0.3057	-0.5500	0.5884
β_{fish}	0.6196	0.0822	7.5400	<.0001
γ_{fatty}	58.0905	25.8580	2.2500	0.0391
$\gamma_{positiv}$	4.9333	1.1746	4.2000	0.0007
γ_{lean}	55.4453	24.3331	2.2800	0.0368
θ_{fish}	0.0000	0.0000	-0.8000	0.4379
α_{fatty}	0.0148	0.1611	0.0900	0.9278
β_{fatty}	6.6610	2.8988	2.3000	0.0354
$\theta_{fish\theta}$	0.0000	0.0000	1.2800	0.2179

2) Example of a somewhat sophisticated(FIML parameter estimated)

Budget share_fish, Adjusted R^2 = 0.1226 Budget share_fatty, Adjusted R^2 = 0.7780				
Parameter	Estimate	Std Err	t Value	Pr > |t|
α_{fish}	0.7070	0.1937	3.6500	0.0020
β_{fish}	0.1805	0.0644	2.8000	0.0123
γ_{fatty}	67.4503	17.8744	3.7700	0.0015
$\gamma_{positiv}$	-3.8601	2.4445	-1.5800	0.1315
γ_{lean}	8.7302	3.9783	2.1900	0.0424
θ_{fish}	0.0000	0.0000	-2.1500	0.0465
α_{fatty}	0.9697	0.0398	24.3500	<.0001
β_{fatty}	0.0586	0.0313	1.8700	0.0787
$\theta_{fish\theta}$	0.0000	0.0000	2.7400	0.0140

3) Example of a sophisticated(FIML parameter estimated)

Budget share_fish, Adjusted R^2 = 0.5811 Budget share_fatty, Adjusted R^2 = 0.6280				
Parameter	Estimate	Std Err	t Value	Pr > |t|
α_{fish}	0.1661	0.0949	1.7500	0.1183
β_{fish}	0.0794	0.0317	2.5100	0.0366
γ_{fatty}	110.9593	36.6983	3.0200	0.0165
$\gamma_{positiv}$	7.2356	2.1103	3.4300	0.0090
γ_{lean}	2.8022	5.0063	0.5600	0.5910

θ_{fish}	0.0000	0.0000	0.1600	0.8756
α_{fatty}	-1.8195	1.1989	-1.5200	0.1676
β_{fatty}	1.2480	0.4893	2.5500	0.0342
$\theta_{fish\theta}$	0.0000	0.0000	-0.2100	0.8408

Appendix 5.E. Calculation of price elasticities

The elasticities are taken from Edgerton *et al.* (1996).

The income elasticity is calculated as:

$$E_i = \frac{\beta_i}{\overline{w}_i}$$

The own and cross price elasticity is calculated as:

$$\varepsilon_{ij} = \frac{\beta_{ij} - \theta_i \left[\alpha_j + ½ \sum_k \left(\beta_{kj} + \beta_{jk} \right) \ln p_k \right]}{w_i} - \delta_{ij} \Rightarrow \varepsilon_{ij} = \frac{\beta_{ij} - \theta_i \left(\alpha_j \right)}{w_i} - \delta_{ij}$$

First Step

Subscript *M* means meat and subscript *F* means fish. $\beta_{F,M} = -\beta_{F,F}$ and $\beta_{M,M} = \beta_{F,F}$ due to homogeneity, $\alpha_M = (1 - \alpha_F)$ and $\theta_M = -\theta_F$ due to adding up, $\beta_{F,M} = \beta_{M,F}$ due to symmetry

The own price elasticity for fish are calculated as:

$$\varepsilon_{F,F} = \frac{\beta_{F,F} - \theta_F \left(\alpha_F \right)}{\overline{w}_F} - 1, \varepsilon_{M,M} = \frac{\beta_{M,M} - \theta_M \left(\alpha_M \right)}{\overline{w}_M} - 1 = \frac{\beta_{M,M} - (-\theta_F)(1 - \alpha_F)}{\overline{w}_M}$$

The cross price elasticity between fish and meat are calculated as:

$$\varepsilon_{M,F} = \frac{\beta_{M,F} - \theta_M \left(\alpha_F \right)}{\overline{w}_M} = \frac{-\beta_{F,F} - (-\theta_F)(\alpha_F)}{\overline{w}_M}$$

Second step

Subscript *f* means fatty fish while subscript *l* means lean fish.

The own price elasticity for fatty fish are calculated as:

$$\varepsilon_{f,f} = \frac{\beta_{f,f} - \theta_f \left(\alpha_f \right)}{\overline{w}_f} - 1, \varepsilon_{l,l} = \frac{\beta_{l,l} - \theta_l \left(\alpha_l \right)}{\overline{w}_l} - 1 = \frac{\beta_{f,f} - \left(-\theta_f \right)\left(1 - \alpha_f \right)}{\overline{w}_l}$$

The cross price elasticity between fatty fish and lean fish are calculated as:

$$\varepsilon_{f,l} = \frac{\beta_{f,l} - \theta_f \left(\alpha_l \right)}{\overline{w}_f} = \frac{-\beta_{f,f} - \theta_f \left(1 - \alpha_f \right)}{\overline{w}_f}, \varepsilon_{l,f} = \frac{\beta_{l,f} - \theta_l \left(\alpha_f \right)}{\overline{w}_l} = \frac{-\beta_{f,f} - \left(-\theta_f \right)\left(\alpha_f \right)}{\overline{w}_l}$$

The variance is calculated according to the delta method (Greene, 2003).