PhD thesis

Information Provision to Consumers as an Instrument of Environmental Regulation

by

Laura Mørch Andersen

AKF, Danish Institute of Governmental Research
and
University of Copenhagen, Institute of Economics

June 2008

Acknowledgements

I wish to thank my supervisor Martin Browning for useful suggestions and ideas during the process. I am also very grateful to my supervisor at AKF, Lars Gårn Hansen, for his never failing support, and for being very actively involved in my work. I co-authored one of my papers with Sinne Smed, a fellow PhD student at AKF. Without the many laughs, walks and discussions we shared, this thesis might never have seen the light of day. Furthermore, I would like to thank a bunch of current and former colleagues at AKF for help, comments and support. I gratefully acknowledge the financial support of the Danish Social Science Research Council (FSE).

Doing a PhD is always a mixed blessing, in this case especially because I chose to approach the engine room of mixed models. When I started my PhD my oldest daughter had just turned five years old, and thought that the word ‘work’ only denoted physical labour such as maintaining the roads and building new houses. Now she realises that doing research is also work and not ‘just making money’. My youngest daughter had just turned one year old, and for some reason one of the first full sentences from her mouth was ‘don’t you mmmm me!’ (in Danish: ‘du ikke sige mmmm til mig!’), showing how precisely she could tell when my mind was preoccupied with something far from her world.

Somehow we all came through, and I wish to thank Freja, Sarah and my husband Søren for their patience.

Contents

Paper 4 is joint with Sinne Smed

Introduction and summary

If information provision to consumers is to be used as an instrument of environmental regulation, it is important to know whether consumers react to this kind of information, and if so, how the reaction varies within the population. The consumer is generally unable to observe the environmental impact of goods neither in the purchase situation, nor during consumption. Information about environmental effects of specific goods may therefore influence or trigger consumers' preferences for this attribute. The purpose of this thesis has therefore been to investigate preferences for different characteristics of goods.

The thesis consists of four papers. Three papers are empirical (paper 1, 2 and 4) and one (paper 3) is methodological. Paper 1 and 2 concern preferences for the organic label and attempt to reveal which motives for purchasing organic goods make a difference at the counter, and which motives do not. The estimations in these two papers are conducted by Mixed Multinomial Logit (McFadden and Train, 2000). During the work with these two papers I encountered substantial problems estimating this model, which turned out to be caused by the standard simulation techniques I used. Paper 3 is methodological and describes the solution to these problems, and ought therefore to be of interest for almost anyone planning on using the Mixed Multinomial Logit. The solution is of course incorporated in the final estimations presented in paper 1 and 2. The last paper (paper 4) is joint work with another PhD student, Sinne Smed from AKF, Danish Institute of Governmental Research. This paper is empirical and uses a Tobit model with two-sided censoring. The focus of this paper is to estimate preferences for nutritional characteristics of milk, primarily fat, and not least to investigate how information influences these preferences. The main methodological difference between this last paper and the others is that paper 1, 2 and 3 deal with the discrete choices made in each single purchasing situation, whereas paper 4 models aggregated monthly demand, which is continuous in nature.

The three empirical papers are all based on the characteristics model, and the paper 1 and 2 investigate altruistic motives for purchasing organic goods. The structure of this introduction is the following: Section 0.1 introduces the characteristics model and section 0.2 introduces altruistic preferences. Section 0.3 presents the data used in the estimations, section 0.4 summarises the results and section 0.5 concludes.

0.1. Characteristics model

The characteristics model was developed by Gorman (published in 1980) and Lancaster (1966). The model assumes that goods are bundles of characteristics, and that consumers derive utility from these characteristics rather than from the goods themselves. The goods are seen as linear combinations of characteristics, and a given characteristic may appear in different goods. The connection between goods and characteristics can therefore be described through the technology matrix which indicates the level of specific characteristics in a number of different goods. The characteristics are sometimes called attributes of the goods.

Variation in utility of goods may thus originate from at least two different sources: Different perceptions of the nature and amount of characteristics of the goods or different valuation of the characteristics of the goods. The organic attribute is a credence good (Giannakas, 2002), which means that consumers cannot observe the organic characteristic directly, and must rely on information about the organic attribute, e.g. the organic label, instead. It is therefore possible to have different perceptions of the organic attribute, and the data used in this paper indicate that this is the case. Some consumers expect to get improvements for their own or their family's health, some expect to get environmental improvements and some expect to increase the level of animal welfare when purchasing organic goods. But others do not, and the technology matrix therefore varies from individual to individual depending on their perception of the organic label. On the other hand, variations in utility between different socio-demographic groups which are not related to (measurable) differences in the perception of the goods might reasonably be perceived as differences in preferences.

0.2. Altruism

A priori, one may be sceptical of the idea that environmental regulation can be done by information provision. Signalling a public good attribute may not be expected to have a significant effect on consumer purchase decisions because of the inherent free-rider problem associated with public goods. Free-riding means enjoying the benefits of goods without paying, e.g by enjoying the environmental effects of organic production without purchasing organic goods. However, early economic contributions (e.g. Sen, 1973) provide a basis for a more optimistic view. Consumers may, in addition to self-interest, be motivated by what Sen called "sympathy" or "commitment" (and others refer to as "altruism", e.g. Andreoni, 1990) or by the fear of acting socially irresponsibly (non-instrumental or symbolic behaviour). A number of recent contributions discuss the implications of altruistic preferences for environmental regulation in different areas (see e.g. Johansson, 1997; McConnell, 1997; Nyborg, 2000).

Only a handful of econometric demand studies have until now investigated the significance of altruistic demand effects: Teisl et al. (2002) investigate the effect of the dolphin-safe label using aggregate time series data, Bennett et al. (2001) and Blamey and Bennett (2001) investigate the effect of claimed (but not certified) environmental attributes on demand for toilet paper using micro level cross-section data, and Bjørner et al. (2004) investigate the effect of the Nordic Swan label on Danish demand for detergent, toilet paper and paper towels using micro level panel data. All find evidence of consumer reactions that may indicate altruistic behaviour.

When it comes to organic goods, several studies have investigated the motives for purchasing these. Most studies are based on relatively few respondents (Makatouni, 2002; McEachern and McClean, 2002; Thompson and Kidwell, 1998; Wolf, 2002) and/or stated consumption of organic goods (Fotopoulos and Krystallis, 2002; Magnusson et al., 2001 and 2003; Makatouni, 2002; McEachern and McClean, 2002; Wandel and Bugge, 1997). Stated consumption has several disadvantages. First of all, it is stated and is therefore to some extent also a measure of intention to buy, and secondly – and perhaps more importantly – it provides no information about the prices facing the respondents in the actual purchase situation. It is therefore impossible to separate the effect of prices from the effect of socio-demographics and attitudes. This thesis distinguishes itself by combining information about actual purchases (including prices of the purchased goods), socio-demographics and perception of organic goods for each of the households in the sample.

The papers presented in this thesis show that people mainly derive utility from improvements in their own or their family's health, but also from improving the environment or increasing the level of animal welfare in the production. Health is a classical private good, only the one who consumes the healthy food benefits (apart from the obvious positive externalities provided by e.g. reducing the cost of public health care). Environment and animal welfare, on the other hand, are public goods – no one can be excluded from enjoying the improved environment, or the knowledge that animals in organic production have a higher level of animal welfare, and they are therefore prone to free-riding. The fact that consumers are observed to purchase goods in order to improve either the environment or the level of animal welfare therefore indicates that altruistic behaviour is not just a theoretical phenomenon.

0.3. Data

The data used in this thesis come from the GfK-Denmark household grocery consumption panel which consists of more than 2,000 households dating back to 1997 (average participation time during the period 1997 to 2001 is 95 weeks) including background characteristics and detailed data from weekly purchasing diaries with price and quantity information on a very disaggregated good level (close to bar code). The "diary keeper" of each household (typically the person responsible for most of the shopping) fills in information about all types of food and groceries purchased by family members and sends a weekly diary report to GfK. For each purchased good on each shopping trip the following information is recorded:

- Type of Good. The commodities are separated into approximately 100 groups, most of these are further disaggregated at a very detailed level. Each year more than 4,500 different goods are purchased.
- Good characteristics (these vary in number and type between goods, e.g. for milk indicators for fat content and type (chocolate milk, buttermilk etc.)).
- Organic/conventional (this is recorded for all goods that can possibly be organic).
- Number of units.
- Price per unit.
- Whether the good was on sale or not.
- Name/type of store (Kvickly, SuperBrugsen, Bilka, Irma …).
- The day of the week and time of day of the shopping trip.
- Participants in the shopping trip.
- The total value of the goods purchased on the shopping trip.

In addition, households annually supply information on socio-demographics such as education, income, club membership and media use etc. Single authored papers in this thesis have only had access to the data from 1997 to 2001, and the co-authored paper 4 has had access to data from 1997 to 2004.

In 2002 AKF issued a large questionnaire on organic food to the GfK panel. The purpose of the questionnaire was to obtain information about knowledge of and attitudes towards organic foods in general at household level. It is therefore possible to combine actual purchases with socio-demographics, attitudes and perception of specific organic goods and thereby entangle the effects of different motives for purchasing organic goods from each other, and to estimate the impact of these different motives on the propensity to buy specific organic goods. For more on the GfK data see Andersen (2006) and Smed (forthcoming).

0.4. Summary of papers

Paper 1 – *'Organic Milk – Who and Why?'* – uses information about purchases of milk during the last six months of 2000 combined with stated perception of general environmental and health effects of organic goods. 51 per cent believes that organic production has a positive effect on the environment and 41 per cent believes in a positive effect on their own health. The number of households used is 1,022, and the number of purchases is 33,993. Data are analysed by Mixed Multinomial Logit, mixing a general organic attribute and varying the mean utility by type of milk, perception of organic goods and the socio-demographics education, urbanisation, income, age and presence of children 0-6 years or 7-14 years old in the household.

The results indicate that consumption of organic milk increases significantly with level of education, urbanisation and income. Age and presence of children in the household have no significant effects. Combining the purchase data with a questionnaire about attitudes towards organic production issued to the same panel shows that 51 per cent believes that organic production has a positive effect on the environment and 41 per cent believes in a positive effect on their own health. The level of trust in organic products generally increases with level of education, urbanisation and income. Including perception of organic goods in the estimation therefore reduces the effects of these socio-demographics, and thereby demonstrates the strength of this type of data combination.

It turns out that both trust in effect on environment and on health increases the probability of choosing organic milk significantly. The effect of trust in health is more than twice as big as the effect of trust in environment.

Paper 2 – *'Animal Welfare and Eggs – Cheap Talk or Money on the Counter?'* – uses purchases of eggs during the one year period from July 1999 to June 2000.[1] The number of households included in the analysis is 844, and the number of purchases is 10,800. The data is

[1] This is the only period in which the panel members reported whether the eggs were barn or free-range. In the rest of the period only the distinction organic/non-organic eggs is reported. This is problematic because the level of animal welfare related to barn and free-range eggs is higher than for battery eggs.

again analysed by a Mixed Multinomial Logit, but this time the utility of the three types of non-battery eggs (barn-, free-range- and organic eggs) is assumed to follow a three-dimensional normal distribution, allowing for correlation between the utility of the three types of eggs.

The purpose of the paper is to identify actual willingness to pay for animal welfare, again by combining actual purchases with stated perception of organic goods. In paper 1 the questions concerning perceptions were related to organic goods in general, but in paper 2 the questions regard organic broilers and egg laying hens. The households are asked whether they expect organic broilers to be less likely to cause a bacterial infection (food safety, used as an indicator of health effects) and whether they expect the animal welfare to be better for hens laying organic eggs. The result shows that consumers perceiving a stronger connection between animal welfare and the organic label have higher willingness to pay for organic eggs, even when controlling for private good attributes such as food safety also connected to the label. The results suggest that altruistic motives may play an important role in the demand for agricultural products.

The use of the Mixed Multinomial Logit model in paper 1 and 2 turned out to be quite a challenge. The estimated log-likelihood values were often highly dependent on the starting values, although the estimated parameters were often very similar. In some cases the problem was small enough to be ignored, but in other cases it ruled out usable likelihood ratio tests. In one case the log-likelihood value of a restricted model was *higher* than the one estimated in an unrestricted model, which means that the likelihood ratio test statistic became negative, something which ought to be impossible.

The solution to the problem is presented in paper 3 – '*Comparable Likelihood Values – Antithetic Halton Draws in Mixed Multinomial Logit*'. The paper explains how Mixed Multinomial Logit models are typically estimated using quasi-Monte Carlo integration, often by Halton draws. However, characteristics known to apply to the likelihood function are only ensured in the limit – i.e. as the number of draws goes to infinity – and in actual estimations the number of draws is limited by lack of computational power and not least limited amounts of time. The paper shows that using asymmetric draws to estimate the integral of a likelihood function – which is symmetric by definition – leads to large differences in the maximum log-likelihood value between different quadrants (depending on the signs of the estimated parameters for the Choleski decomposition of the covariance matrix). The standard estimation procedures used today use asymmetric Halton draws and the paper shows that this may result in substantial estimation and inference errors within the span of draws typically applied. A similar type of problem occurs if the relationship between primes and mixed parameters is not maintained when testing mixing of parameters, again something which is typically not done with standard estimation procedures.

Even when the problems created by asymmetric draws are eliminated, the data may still not be informative enough to allow for the most sophisticated models. The data on eggs used in paper 2 allow for a three dimensional mixing with correlation, whereas the data on milk used in paper 1 only allows for a one dimensional mixing. Trying to increase the level of complexity leads to multiple maxima in the milk case, even when using antithetic Halton draws. It is therefore extremely important *always* to check the stability of the estimated Mixed Multinomial Logit models by varying the starting values.

The last paper (paper 4) – *'A Censored Structural Characteristics Model for Milk'* – is joint work with Sinne Smed, also from AKF – Danish Institute of Governmental Research. 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 here 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 information about daily purchases from the individual households in the GfK panel spanning the period from 1997 to 2004, combined with information about social and demographic characteristics of the households. The purchasing data have been combined with the number of articles mentioning a link between the intake of fat and health in Danish newspapers for each time period. The panel structure of these 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 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, but are rather price sensitive. The latter is of great importance since households who drink a lot of milk prefer milk with a moderate to high fat share.

0.5. Conclusion

The results presented in this thesis indicate that it is possible to change consumer behaviour by providing information, and therefore perhaps also to include information provision as a tool for environmental regulation. The information may either change the perception of goods by adding new characteristics, or change the preferences for characteristics already known to be part of the goods. In paper 1, the perceived characteristics of organic milk vary between consumers based on different evaluations of information about environmental effects of organic products, and in paper 4 consumers are expected to agree on the content of fat in the different types of milk, but the preferences for fat are observed to change over time due to information about negative effects of fat consumption.

Paper 1 and 2 show that consumers have a significant utility and thereby willingness to pay for public goods such as environmental improvements and animal welfare. In other words, the consumers exhibit altruistic behaviour, not only when they are talking to researchers (as in studies using stated behaviour), but also when they purchase actual goods in actual stores. When it comes to the effects of the more conventional socio-demographics, the results in these two papers indicate that urbanisation and income are the most important factors in explaining consumption of organic goods.

Turning to the more technical outcome of the thesis, the characteristics model developed by Gorman and Lancaster proved to be a powerful tool both when estimating on discrete choices and continuous good space. The three empirical papers all confirm that the degree of heterogeneity among consumers is substantial. Paper 1 and 2 show that part of this heterogeneity can be eliminated by including information about perceptions of goods for the individual households and paper 4 shows that if the data are good enough, it is possible to allow for the maximum degree of individual heterogeneity by estimating on data from each household separately.

Paper 3 shows that antithetic Halton draws eliminate part of the variance of the log-likelihood functions of Mixed Multinomial Logit models, namely the part that arises from asymmetric draws.

References

- Andersen, L.M. (2006): Consumer Evaluation of Environmental and Animal Welfare Labelling. AKF Working Paper. Available at www.akf.dk .
- Andreoni, J. (1990): Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, **100** (401): 464-477.
- Bennett, J., R. Blamey, J.J. Louviere and M. Morrison (2001): Green Product Choice. In J. Bennett and R. Blamey (eds.): *The Choice Modelling Approach to Environmental Valuation*. Edward Elgar, Cheltenham, UK.
- Bjørner, T.B., L.G. Hansen and C.S. Russell (2004): Environmental labelling and consumers' choice – An empirical analysis of the effect of the Nordic Swan. *Journal of Environmental Economics and Management*, **47** (3): 411-434.
- Blamey, R. and J. Bennett (2001): Yea-saying and Validation of a Choice Model of Green Product Choice. In J. Bennett and R. Blamey (eds.): *The Choice Modelling Approach to Environmental Valuation*. Edward Elgar, Cheltenham, UK.
- Fotopoulos C. and A. Krystallis (2002): Purchasing motives and profile of the Greek organic consumer: a countrywide survey. *British Food Journal*, **104** (9): 730-765.
- Giannakas, K. (2002): Information Asymmetries and Consumption Decisions in Organic Food Product Markets. *Canadian Journal of Agricultural Economics*, **50** (1): 35-50.
- Gorman, W. M. (1980): A Possible Procedure for Analysing Quality Differentials in the Egg Market. *The Review of Economic Studies*, **47** (5): 843-856.
- Johansson, O. (1997): Optimal Pigovian Taxes under Altruism. *Land Economics*, **73**: 297-308.
- Lancaster, K. J. (1966): A New Approach to Consumer Theory. *The Journal of Political Economy*, **74 (**2): 132-157.
- Magnusson, M. K., A. Arvola, U. K. Hursti, L. Åberg and P. Sjödén (2001): Attitudes towards organic foods among Swedish consumers. *British Food Journal*, **103** (3): 209-226.
- Magnusson, M. K., A. Arvola, U. K. Hursti, L. Åberg and P. Sjödén (2003): Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behaviour. *Appetite*, **40**: 109-117.
- Makatouni, A. (2002): What motivates consumers to buy organic food in the UK? Results from a qualitative study. *British Food Journal*, **104** (3/4/5): 345-352.
- McConnell, K.E. (1997): Does Altruism Undermine Existence Value? *Journal of Environmental Economics and Management*, **32**: 22-37.
- McEachern, M. G. and P. McClean (2002): Organic purchasing motivations and attitudes: are they ethical? *International Journal of Consumer Studies*, **26** (2): 85-92.
- McFadden, D. and K. Train (2000): Mixed MNL models for discrete response. *Journal of Applied Econometrics*, **15**: 447-470.

- Nyborg, K. (2000): Homo Economicus and Homo Politicus: Interpretation and Aggregation of Environmental Values. *Journal of Economic Behavior and Organization*, **42** (3): 305-322.
- Sen, A. (1973): Behaviour and the Concept of Preferences. *Economica*, **40**: 241-59.
- Smed, S. (forthcoming): *Empirical studies on "Health, Information and Consumer Behaviour"*. PhD thesis, University of Copenhagen.
- Teisl, M.F., B. Roe and R.L. Hicks (2002): Can Eco-labels Tune a Market? Evidence from Dolphin-Safe Labeling. *Journal of Environmental Economics and Management*, **43** (3): 339-359.
- Thompson, G.D. and J. Kidwell (1998): Explaining the choice of organic produce: Cosmetic defects, prices and consumer preferences. *American Journal of Agricultural Economics*, **80**: 277-287.
- Wandel, M. and A. Bugge (1997): Environmental concern in consumer evaluation of food quality. *Food Quality and Preference*, **8** (1): 19-26.
- Wolf, M.M. (2002): *An Analysis of the Impact of Price on Consumer Interest in Organic Grapes and a Profile of Organic Purchasers*. Paper presented at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 28-31, 2002.

Organic Milk – Who and Why?*

Laura Mørch Andersen†

June 2008

Abstract

Using a unique data set where an unbalanced panel of more than 1,000 households have reported their purchases of groceries in great detail over a period of six months it is shown that consumption of organic milk increases significantly with level of education, urbanisation and income. Age and presence of children in the household have no significant effects.

Combining the purchase data with a questionnaire about attitudes towards organic production issued to the same panel shows that 51 per cent believes that organic production has a positive effect on the environment and 41 per cent believes in a positive effect on their own health. The level of trust in organic products generally increases with level of education, urbanisation and income. Including perception of organic goods in the estimation therefore reduces the effects of these socio-demographics, and thereby demonstrates the strength of this type of data combination.

It turns out that both trusts in effect on environment and on health increases the probability of choosing organic milk significantly. The effect of trust in health is more than twice as big as the effect of trust in environment.

Key words: Heterogeneity of preferences, panel mixed multinomial logit, MMNL, MXL, market data, labelling, characteristics model, health, environment, organic

---

* I thank GfK Denmark for providing the purchase and background data, and for issuing the questionnaire.
I thank Kenneth Train, David Revelt and Paul Ruud for making their MMNL software available at Train's MMNL homepage: elsa.berkeley.edu/Software/abstracts/train0296.html (verified 11 June 2008), and I especially thank Kenneth Train for fast and clarifying answers to my questions.
The research was funded by The Danish Social Science Research Council ('FSE').
† Contact information: Laura Mørch Andersen, AKF - Danish Institute of Governmental Research, www.akf.dk
e-mail: LMA@akf.dk

1.1. Introduction

The growing interest in organic agriculture has inspired numerous scientists to investigate the motives for purchasing organic goods. Most studies are based on relatively few respondents (Makatouni, 2002; McEachern and McClean, 2002; Thompson and Kidwell, 1998; Wolf, 2002) and/or stated consumption of organic goods (Fotopoulos and Krystallis, 2002; Magnusson et al., 2001 and 2003; Makatouni, 2002; McEachern and McClean, 2002; Wandel and Bugge, 1997). Stated consumption has several disadvantages. First of all it is stated and is therefore to some extent also a measure of intention to buy, and secondly – and perhaps more importantly – it provides no information about the prices facing the respondents in the actual purchase situation. It is therefore impossible to separate the effect of prices and budget restrictions from the effect of socio-demographics and attitudes. This paper distinguishes itself by using information about actual purchases (including prices of the purchased goods), socio-demographics and answers to a questionnaire about perception of organic goods for each of the 1,022 households in the sample.

The data on prices available for the present study means that it is possible to investigate whether the lack of income effect in stated behaviour studies (e.g. Wolf, 2002) might be due to the absence of budget restriction in the hypothetical settings. The data used for this analysis also make it possible to entangle the effects of attitudes from the effects of socio-demographics. It is therefore possible to investigate whether attitudes are correlated with socio-demographics and to what extent the effect of socio-demographics observed in studies without information about perception of organic products could be ascribed to attitudes rather than socio-demographics.

Methodology: Data on actual purchases of milk during a six-month period from 1,022 households are combined with information about the perception of environmental and health effects of organic goods for each individual household along with information about income, urbanisation, education, age and presence of children in the household. A Lancaster characteristics model (Lancaster, 1966) is estimated as a discrete choice model, using mixed multinomial logit (McFadden and Train, 2000). A model including only purchases and socio-demographics is compared to a model which also includes questionnaire responses. The result is that socio-demographics and attitudes are correlated, and that the effects of soci-demographics may be exaggerated in estimations where individual perceptions of the organic label are not available. Using a discrete model means that the data is investigated as close as possible to the actual purchase situation, which involves discrete choices between alternatives.

The purpose of this paper is twofold: To separate the effects of different motives for purchasing organic milk and to give an introduction to the mixed logit model. Readers who are not interested in estimation technique may skip section 1.6 and 1.7.

The structure of the paper is as follows: Section 1.2 presents the data which combine information about actual purchases with information about not only conventional socio-demographics, but also attitudes towards the organic label. Section 1.3 discusses the different types of milk. Section 1.4 motivates the choice of purchase motives and socio-demographics used in the paper. Section 1.5 explains how the model is related to Lancaster's characteristics model and section 1.6 introduces the Mixed Multinomial Logit model (MMNL). Section 1.7 presents the empirical specification of the utility function. Section 1.8 provides the main results of estimations and section 1.9 concludes.[1]

1.2. Data

The data are collected by GfK ConsumerScan Denmark (GfK). Each week households in the panel report the values and volumes of their actual purchases to the GfK in a 'diary'. Among other attributes, the households report whether the goods are organic or conventional, and for milk the *type* of milk is reported which means that the approximate fat contents and taste are known. All data are self-reported by the households. GfK recommends that the diaries are filled in immediately after each shopping trip to avoid problems with forgotten purchases. Once a year the households answer a questionnaire about household attributes such as e.g. level of education for father and mother, and household income.

The data on milk[2] used in this paper cover the period from 1 July 2000 to 31 December 2000 and are combined with results from a questionnaire about attitudes towards organic production issued to the panel in the summer of 2002 by AKF and GfK. 1,771 households reported purchases of milk during the six months in the data period, and 1,022 of these also

---

[1] Appendix A gives a more thorough definition of the socio-demographics, Appendix B provides information about prices and market shares, Appendix C shows the relationship between socio-demographics, Appendix D shows the relationship between trust and socio-demographics and Appendix E provides a list of tables, both in the paper and in the appendixes.

[2] GfK has collected purchases of all types of food, and since 1997 it has been recorded whether the food was organic. For more on the GfK data see Andersen (2006).

answered the questions used in this paper. The number of observed purchases from these households was between 1,033 and 1,596 per week, with a median of 1,321. 10 per cent of the households reported less than 6 purchases, but half of the households reported more than 28 purchases of milk and 25 per cent reported more than 46 purchases (for more on the GfK purchase data see Andersen, 2006).

As mentioned above, the data include answers to a questionnaire on attitudes towards organic production. The questionnaire was issued to the households participating in the panel during the summer of 2002. Many of the respondents also participated during the last six months of 2000, and for these households it is possible to combine stated preferences with observed purchasing behaviour. This means that it is possible to entangle the effects of trust in an organic effect on environment and on health from each other, and to estimate the impact of these different types of trust on the propensity to buy organic milk. The relationship between the questionnaire and the purchase data is based on the assumption that the perception of organic goods has been unchanged from 2000 to 2002, something which might not be entirely true. If the perception of organic good has changed it means that distinction between the group of households perceiving no effect and the ones expecting environmental or health improvements will be less precise, and that the effects of trust may be underestimated.

The degree of trust in the organic label is determined from the question: 'To what extent do you agree with the following statements? … I think that the rules regarding organic production are good enough to create improvements for …'

- Nature, e.g. wild animals and plants
- My and my family's health

The respondents were allowed to answer on a five-point scale ranging from 'Totally disagree' to 'Totally agree'.

In Table 1, the 'totally disagree' category is merged with the 'disagree' category and the 'agree' category with the 'totally agree' category. This leads to nine possible combinations of the attitudes towards environment and health. Trust in positive effects on environment seems to be a precondition for trust in positive effects on health, as only 46 households have trust in health but not the environment. On the other hand, health is not a precondition for environment, as 152 households trust in environmental effects, but not in health effects. This indicates that many people believe that their own health is related to the 'health' of the surrounding environment, and thereby supports the results found by Makatouni (2002). Makatouni reported results from qualitative interviews with 40 British parents, and found that health (personal or for their families) was the most important factor when trying to explain stated organic consumption. Environment and animal welfare were also important, but mainly through their impact on the health factor.

Table 1 Relationship between trust in organic effect on environment and health

Number of households/ Number of purchases/		Health: Disagree	Health: Uncertain	Health: Agree	*Total*
Environment	**Disagree**	106 3,043	18 656	8 146	*132* *3,845*
	Uncertain	15 537	313 10,679	38 1,202	*366* *12,418*
	Agree	20 545	132 4,437	372 12,748	*524* *17,730*
	Total	*141* *4125*	*463* *15,772*	*418* *14,096*	*1,022* *33,993*

Source: AKF/GfK questionnaire data from 2002.

Table 2 shows the relationship between the organic purchase share and the different combinations of the answers to questions on environment and health. Households who believe in none of the effects still purchase organic milk in 8 per cent of the cases, so environment and health are not the only attributes of organic products that matter. They are, however, very important. Trust in just one of the two practically doubles the purchase share, and trust in both health and environment leads to a purchase share of 43 per cent.

Table 2 Organic purchase shares by perception of organic goods

Organic purchase share		Health: Disagree	Health: Uncertain	Health: Agree	*Total*
Environ-ment	**Disagree**	7.8%	17.8%	20.6%	*10.0%*
	Uncertain	12.1%	13.0%	28.7%	*14.5%*
	Agree	27.2%	21.1%	43.4%	*37.3%*
	Total	*10.9%*	*15.5%*	*41.9%*	*25.9%*

Source: GfK purchase data for milk June to December 2000 combined with questionnaire data from 2002. Only whole, semi-skimmed and skimmed milk.

1.3. Milk

During the last six months of 2000 three main types of milk were available with different contents of fat:

- Whole milk (Sødmælk): 3.5 per cent fat

- Semi-skimmed milk (Letmælk): 1.5 per cent fat
- Skimmed milk (Skummetmælk): Between 0.1 per cent and 0.5 per cent fat

The conventional versions of these types were always homogenised and the organic versions were un-homogenised.[3] Note that the effect of homogenisation is perfectly correlated with the organic label in these types of milk and therefore not identified.

The nature of the data means that only the price of the chosen alternative is recorded. The prices and availability of the different types of milk in each choice situation are imputed from purchases made (in the same chain of stores, within the same week) by other panel members. If nobody purchased a given type of milk in a given chain of stores in a given week it is perceived as rationed, and not included as an alternative in the specific purchase situation. Figure 1 and Figure 2 show the absolute imputed prices and price differences. To avoid systematic differences in the measurement errors of the price all prices are imputed, including the one for the type that was actually chosen.

Figure 1

Absolute prices of different types of milk over time

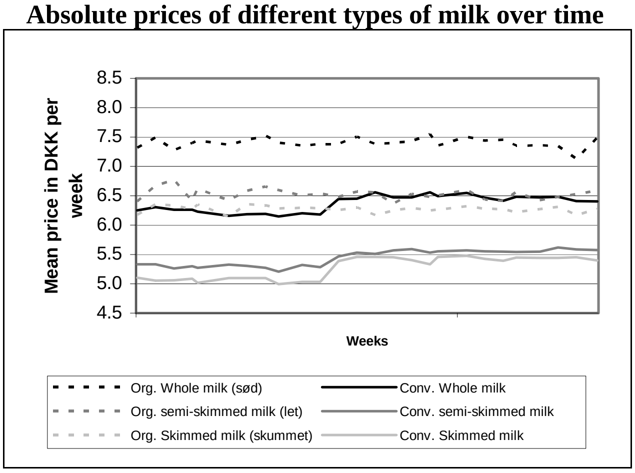
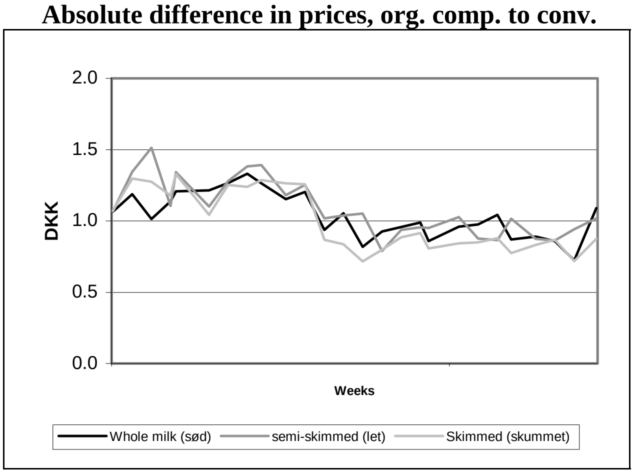


Figure 2

Absolute difference in prices, org. comp. to conv.

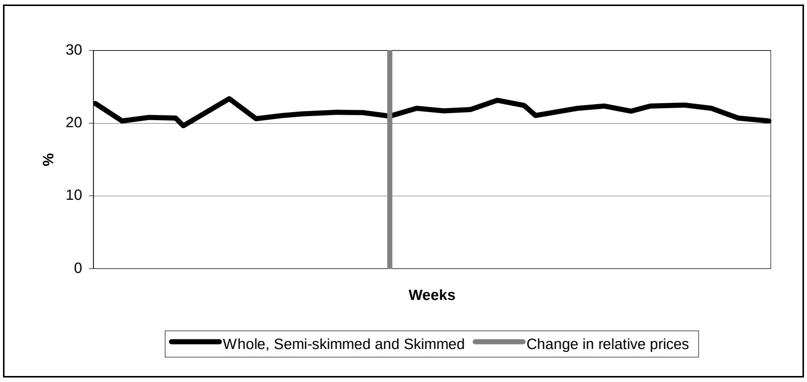


Source: GfK purchase data for milk June to December 2000, only whole, semi-skimmed and skimmed milk.

It is clear that the prices of all types of conventional milk for some reason were increased mid September 2000. As can be seen in Figure 3 the change in relative prices did not affect the organic volume share.

[3] Buttermilk (kærnemælk) and chocolate milk (kakaomælk) are excluded because they taste very different compared to the other types, and are used for different purposes.

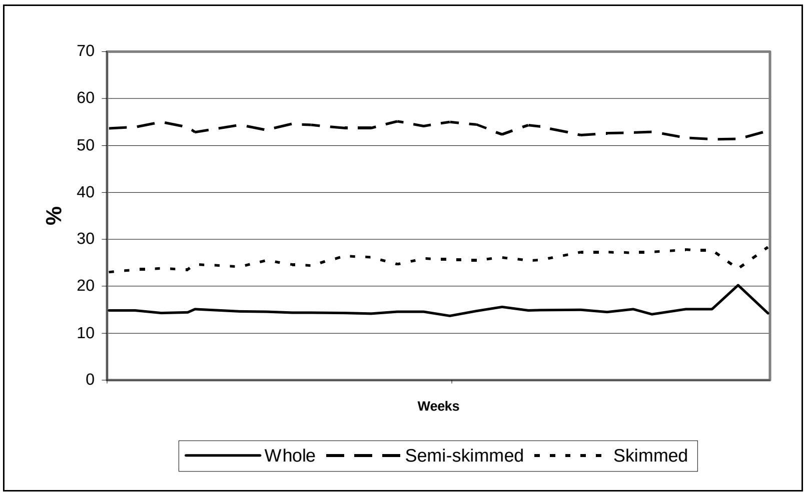
Figure 3 Organic volume share



Source: GfK purchase data for milk June to December 2000, only whole, semi-skimmed and skimmed milk.

Figure 4 shows that the consumption of different types of milk has remained practically unchanged during the data period, and the propensity to buy the three different types of milk is therefore assumed to be constant.

Figure 4 Distribution of the milk market on different types of milk, volume shares[4]



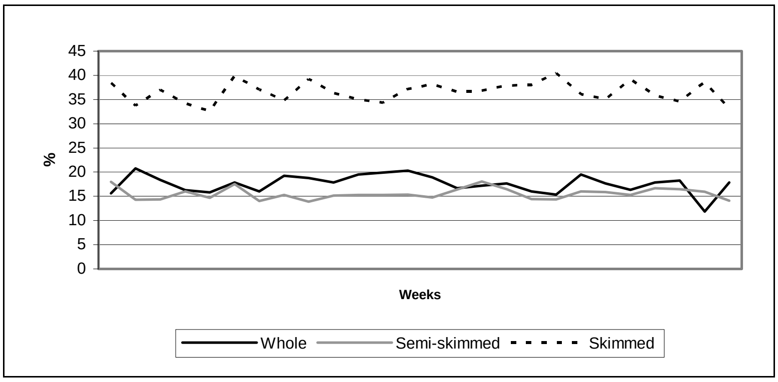
Source: GfK purchase data for milk June to December 2000, only whole, semi-skimmed and skimmed.

The propensity to buy the organic version varies between the different types of milk, as can be seen from Figure 5. The organic share of skimmed milk is much higher than the share of semi-skimmed and whole milk. Skimmed milk has a low fat content and might appeal more to people who are very health conscious. People who are more health conscious may also be more interested in organic products because these are often regarded as healthier. The low fat contents of skimmed milk make the effect of homogenisation smaller compared to semi-skimmed or whole milk. Many people dislike the 'lumpiness' of un-homogenised milk, and

[4] The peak on whole milk corresponds with Christmas where Danes eat various traditional dishes based on pudding rice and whole milk. This peak is clear in each of the five years in the original data.

this negative effect is likely to be smaller for skimmed milk. None of these correlations can be tested using the data at hand, so this is mere hypothesis.

Figure 5 Organic volume share, whole, semi-skimmed and skimmed[5]



Source: GfK purchase data for milk June to December 2000, only whole, semi-skimmed and skimmed milk.

As for the propensity to purchase milk with different levels of fat, the propensity to buy the organic version is invariant during the data period. It is therefore assumed to vary between milk types, but to be constant over time.

1.4. Motives and socio-demographics

Many studies have investigated the motives for purchasing organic goods (e.g. Fotopoulos and Krystallis, 2002, Makatouni, 2002, Magnusson et al., 2003, McEachern and McClean, 2002), and some of the most important motives appear to be environmental and health improvements (Makatouni, 2002, Magnusson et al., 2003). As mentioned before the data used in this paper includes household perception of environmental and health benefits from purchasing organic products. The relationship between the perception of environmental and health effects was presented in Table 1 on page 18, and it is evident that distrust in effect on environment and health is closely related. This means that the effect of distrust in health effects and distrust in environmental effects therefore cannot be identified separably. It is therefore chosen to compare the utility of the organic characteristic for households which agree in a positive effect with the utility for those who either disagree or are uncertain about the effect. The organic purchase shares for these groups are reported in Table 3 below.

When it comes to socio-demographics, this paper focuses on the effect of income, urbanisation, education, age and presence of children in the household.[6] These characteristics

[5] The discrete choice model used in this paper disregards the volumes and focuses on the probability of choosing the different types of milk in a given purchase. The difference between volume shares and purchase shares is not substantial in this case.

do not vary over the period used in the estimations, so in that sense the data are treated as cross-sectional. However, the utility function is assumed to be constant for each household during the estimation period, but to vary between households, thereby utilising the panel dimension of the data.

A quick look at the highest level of education within the household shows that it has a vast effect on the propensity to buy organic milk. The effect ranges from an organic purchase share of 20 per cent for households with no further education to 42 per cent for households with a long further education (Table 3). The question is whether the effect of education is an expression of something else. From Table 10 and Table 11 in Appendix C it is clear that there is a relationship between education and urbanisation and income. A high level of education seems to be associated with having a relatively high income and living in the capital area. It is therefore important to include these explanatory variables in the estimation.

Children (especially young children) are expected to have a positive effect on the propensity to choose organic products because the health of young children may be more important for parents than their own health. Even if the parents are not convinced that organic products are healthier, they may buy them as insurance just in case. Looking at data, it seems that children most likely result in a negative effect (the organic purchase share is 27 per cent for households with no young children and 16 per cent for households with young children where the effect was expected to be the largest, see Table 3). Data indicate that the effect of children may vary with level of education, but the number of families with children is too small to estimate the effect. The presence of children in the household is strongly correlated with age, and it is therefore important to control for this effect too.

The perceptions of environmental and health effects of organic goods mentioned above are related to the socio-demographics. The relationship between trust and socio-demographics is described in Table 25 to Table 36 in Appendix D. Trust in positive effects of organic products is more likely to be present when either income or level of education is high, and when the household is living in the capital area. In many studies information about attitudes is not available for the estimation, and the effect of these socio-demographics may therefore be overestimated. The question is by how much.

---

[6] See Appendix A for a detailed definition of these, and Table 3 for descriptive statistics on the relationship between these variables and the organic purchase share.

Table 3 sums up the information about perception of organic goods and socio-demographic characteristics used in the estimations and provides the organic purchase share for each group. Table 3 also indicates which sub-groups constitute the control group in the estimations. The utility of the organic characteristic in the other groups is measured relative to this group. The estimated utility in the capital area is therefore the difference between the mean utility in households in the capital area and those in the rural municipalities. If the parameter for capital area is significant, it means that the difference between the utility in the capital area and in the rural municipalities is significantly different from zero.

Table 3 Perceptions and socio-demographic data used in estimations

Variable	Sub-groups	Number of households	Share of households	Control group[a]	Organic purchase share
Environment[b]	Disagree or not sure	498	49	X	13
	Positive effect on environment	524	51		37
Health[c]	Disagree or not sure	604	59	X	15
	Pos. effect on own or family's health	418	41		42
Income[d]	Lowest 25 %	283	28	X	22
	Middle 50 %	462	45		23
	Highest 25 %	277	27		36
Degree of urbanisation[e]	Rural municipality	338	33	X	19
	Urban municipality	468	46		25
	Capital area (Copenhagen)	216	21		42
Level of education[f]	No further education stated	277	27	X	20
	Vocationally oriented high-school	347	34		21
	Short further education	172	17		34
	Medium further education	176	17		32
	Long further education	50	5		42
Age[g]	18-29 years	44	4		17
	30-44 years	218	21		24
	45-59 years	363	36		27
	60 years or more	397	39	X	27
Children 0-6 years[h]	No	937	92	X	27
	Yes	85	8		16
Children 7-14 years[h]	No	902	88	X	27
	Yes	120	12		23

Data source: GfK purchase data for milk June to December 2000 combined with background data covering 2000 and questionnaire data from 2002. Only whole, semi-skimmed and skimmed milk. The total number of households is 1,022 and the organic purchase share for all households is 26 per cent.

a: Utility of the organic characteristics in the other groups is measured relative to this group.
b: To what extent do you agree with the following statement: "I think that the rules regarding organic production are good enough to create improvements for nature, e.g. wild animals and plants".
c: To what extent do you agree with the following statement: "I think that the rules regarding organic production are good enough to create improvements for my and my family's health".
d: Income is recorded in brackets of DKK 50,000 (~€6,700). These brackets are divided by the number of persons in the household, weighted by the OECD-modified scale i.e. 1 for the first adult, 0.5 for the next adults and 0.3 for children (OECD). Income is split into three categories indicating relative levels of income.
e: GfK divides the 275 Danish municipalities (2002) into categories depending on how urbanised they are and on their geographical location. The geographical location is ignored here, and the sample is split into rural, urban and capital area municipalities.
f: Highest level of education within the household.
g: Age is defined by the age of the oldest person in the household.
h: Indicates whether children in a specific age group are present in the household.

1.5. Characteristics of milk

As in Lancaster (1966) it is assumed that goods are bundles of characteristics and that consumers derive utility from these characteristics rather than from the goods themselves. [7] The goods are linear combinations of characteristics and the connection between goods q and characteristics z can therefore be described through the technology matrix A:

$$\textit{Goods}\ \begin{cases}1\\ \vdots\\ i\\ \vdots\\ I\end{cases} \overset{\overbrace{1 \quad \cdots \quad j \quad \cdots \quad J}^{\textit{Characteristics}}}{\begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1J} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{iJ} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ a_{I1} & \cdots & a_{Ij} & \cdots & a_{IJ} \end{bmatrix}} \equiv \mathrm{A} \tag{1.1}$$

which means that the relationship between goods purchased and characteristics obtained can be written as $z = \mathrm{A}'q$.

In the case of milk the consumer can choose between an organic and a conventional version of three different types of milk, leading to six goods, each constituting a different combination of characteristics. The goods are presented in Table 4:

Table 4 Definition of choice set

	Organic	***Conventional***
Whole (3.5% fat)	Alternative 1	Alternative 2
Semi-skimmed (1.5% fat)	Alternative 3	Alternative 4
Skimmed (0.1-0.5% fat)	Alternative 5	Alternative 6

The three types of milk (whole, semi-skimmed and skimmed) all share a set of 'milkiness' characteristics which differentiate the product from other goods which also consists of fat, protein, calcium etc. Milk can be used for drinking, coffee and other things where e.g. butter would be inapplicable. The fat percentages of different types of milk leads to differences in taste and other sensory characteristics of the milk, but not necessarily as a linear function of the fat percentage. It is therefore also necessary to include 'whole-milkiness', 'semi-skimmedness' and 'skimmedness' as characteristics of the goods. The organic attribute is also assumed to consist of a general part, and a part which is allowed to depend on the type of milk, mainly because the effect of the non-homogenisation is likely to vary a great deal depending on the fat percentage.

---

[7] Characteristics are the same as attributes.

$$\text{goods}\left\{\begin{array}{c}\overbrace{\left[\begin{array}{c|cccccccc} & org(w): & whole: & org(ss): & semi: & org(s): & skim: & org: & milk: \\ \hline Org\_Whole: & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ Conv\_Whole: & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ Org\_SemiSkim: & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ Conv\_SemiSkim: & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ Org\_Skimmed: & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ Conv\_Skimmed: & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array}\right]}^{\text{characteristics}}\end{array}\right. \equiv \mathrm{A}' \tag{1.2}$$

The general organic attributes is a credence good (Giannakas, 2002), which means that consumers cannot observe the organic characteristic neither in the purchase situation, nor at the point of consumption. Consumers must therefore rely on the organic labelling. It is therefore possible to have different perceptions of the organic attribute, and the data used in this paper show that some consumers expect to get a positive effect on the environment when purchasing organic goods, whereas others do not, just as some expect to get a positive health effect (Table 1). This leads to an individual specific technology matrix A_i indicating that the households receive different sets of characteristics when consuming organic goods. In the case of milk the matrix becomes:

$$\text{goods}\left\{\begin{array}{c}\overbrace{\left[\begin{array}{c|cc} & environment: & health: \\ \hline Org\_Whole: & 1_{\{Env,trust\}} & 1_{\{Health,trust\}} \\ Conv\_Whole: & 0 & 0 \\ Org\_SemiSkim: & 1_{\{Env,trust\}} & 1_{\{Health,trust\}} \\ Conv\_SemiSkim: & 0 & 0 \\ Org\_Skimmed: & 1_{\{Env,trust\}} & 1_{\{Health,trust\}} \\ Conv\_Skimmed: & 0 & 0 \end{array}\right]}^{\text{characteristics}}\end{array}\right. \equiv \mathrm{A}_i' \tag{1.3}$$

Household who neither believes in environmental nor health effects receive only the characteristics of the common technology matrix A, whereas those who trust in environmental or health effects also benefits from these when they purchase organic products. The characteristics obtained from a bundle of goods can therefore be described as:

$$z_i\left(q\right) = \mathrm{A}'q + \mathrm{A}_i'q \tag{1.4}$$

The technology matrix in (1.2) means that if a household with no trust in environmental or health effects purchases a litre of organic whole milk it gets (see equation (1.2)) one unit of whole-organic-ness, one unit of whole-milkiness, one unit of general-organic-ness and one unit of milkiness. A household who believes in positive effects on both the environment and own or family's health gets the same, but also one unit of environmental improvements and one unit of improved health.[8]

Variation in utility of goods may thus originate from at least two different sources: Different perceptions of the characteristics of the goods or different preferences for the characteristics of the goods. In this example, the perception of environmental and health effects of organic goods varies between households, and thus results in different perceptions of the characteristics related to organic goods, whereas the difference in utility of the general organic attribute between socio-demographic groups is interpreted as differences in preferences. The preferences for environment and health are assumed to be the same for all households.[9]

1.6. Mixed logit

When dealing with discrete choices, the parameters of the utility function are often estimated using a conventional multinomial logit model (e.g. Greene 1997) which means that the household likelihood function is

$$L_i^{conv}(\beta) = \prod_{t=1}^{T_i} \left(\frac{\exp(U_{it}(j))}{\sum_{k=1}^{J} \exp(U_{it}(k))} \right) \tag{1.5}$$

where β is a vector containing all of the parameters of the utility function,[10] J is the number of alternatives in the choice set (in this case six) and $U_{it}(k)$ is the utility for household *i* from choosing alternative *k* from the choice set in period *t*.

However, the conventional multinomial logit model suffers from the assumption of Independence of Irrelevant Alternatives (IIA). Imagine that organic skimmed milk leaves the market. Then the IIA in the multinomial logit model would imply that the people who used to buy organic skimmed milk would distribute themselves between the rest of the five combinations of organic/conventional and milk type according to the *market* share of these

[8] The value of these units is likely to vary between different food categories, so this is actually one unit of e.g. organic-milk-healthiness.

[9] The utility of environment and health is assumed to be independent of socio-demographics (a simplifying assumption which could be relaxed in further research).

[10] The details of the empirical specification of the utility function is given in section 1.7.

other combinations. But people who buy organic skimmed milk may very well have a higher propensity to buy either organic semi-skimmed milk or conventional skimmed milk than the population in general and, in particular, have a lower propensity to buy conventional whole milk. IIA is therefore not reasonable in this case.

Investigating data shows that some households buy e.g. organic milk more frequently than others, which contradicts the theory that all households have the same utility of the organic attribute. As in e.g. McFadden and Train (2000), Revelt and Train (1998), Train (1998) or Train (1999) it is therefore assumed that (part of) the household utility is drawn from a distribution (i.e. the household utility is known to the household, but only the distribution is observable to the econometrician). The household likelihood function then becomes the likelihood function in the conventional multinomial logit model integrated over all possible values of β:

$$L_i(\theta) = \int L_i^{conv}(\beta) f(\beta|\theta) d\beta \tag{1.6}$$

where θ are the parameters determining the distribution of the utility β, and $f(\beta|\theta)$ is the density of β given θ. The likelihood function is maximised over θ instead of β. This is known as the Mixed MultiNomial Logit (MMNL or MXL) model (McFadden and Train 2000). As will be seen in the following the MMNL model does not suffer from IIA (as long as at least one parameter is assumed to be drawn from a common distribution (i.e. to be '*mixed*')):

Under the conventional multinomial logit the utility function is assumed to be

$$U_{it}(j) = \beta' x_{ijt} + \varepsilon_{ijt} \tag{1.7}$$

with identical β's for all households and i.i.d. extreme value error terms ε_{ijt}. The fact that the error terms are independent over households *i*, milk types *j* and time *t* creates IIA. As in Train (1998) the utility function in the Random Utility Model underlying the MMNL model can be written as

$$U_{it}(j) = \beta_i' x_{ijt} + \varepsilon_{ijt} = (b' + \eta_i') x_{ijt} + \varepsilon_{ijt} = b' x_{ijt} + \eta_i' x_{ijt} + \varepsilon_{ijt} \tag{1.8}$$

where the household-specific β_i is decomposed into a part, *b*, that is common for all households (the mean of the distribution of household β_i's) and an individual part, η_i, that differs between households and has mean zero in order to separate the effect of *b* from the effect of η_i.

The common part, b, can be estimated by the econometrician, but the individual part η_i remains unobserved by everyone except the household itself. The econometrician will, therefore, observe the error terms

$$\xi_{ijt} = \eta_i x_{ijt} + \varepsilon_{ijt} \tag{1.9}$$

which are correlated over alternatives (j) and time (t) for household i because of the common influence of η_i. This means that the differences in taste make the probability of choosing different types of milk correlated for household i. The households that have tastes different from the mean of the population ($\eta_i \neq 0$) will therefore not distribute their consumption according to the average distribution and thus not substitute according to this average distribution, but according to their own conventional multinomial logit model. The fact the errors are correlated over possible alternatives therefore eliminates IIA, and means that a mixed multinomial logit is more flexible than the conventional multinomial logit model.

1.7. Empirical specification of the model

The utility of the characteristics is assumed to follow a Random Utility Model (RUM) in which the household utility is not perfectly observed by the econometrician. The utility function is assumed to have a simple linear form, depending on the prices of the different alternatives of milk and the characteristics of the alternatives. The linear form of the utility function means that the marginal willingness to pay is simply the utility of the attribute divided by the utility of money, just as in Hanemann (1984).[11] The relationship between goods and characteristics ($z = \mathrm{A}'q$) means that the utility of the characteristics inherent in the goods is

$$U(z) = \beta'\left(\mathrm{A}'q + \mathrm{A}_i'q\right) \tag{1.10}$$

---

[11] Marginal willingness to pay is the amount of money a person is willing to pay in order to receive an extra unit of the good in question. It implies that the person is assumed to be at a given level of utility when he is offered an extra unit of the good. If the consumer is faced with a unit price for the good, he will only accept the purchase if it leaves him with at least the initial level of utility. The point of interest is the unit price that will lead to the same level of utility regardless of whether the person chooses to buy the good or not, since this is the maximum amount the person will be willing to pay. Actually, this is 'marginal *maximum* willingness to pay', but it is often simply referred to as 'marginal willingness to pay' or 'wtp'.

where A and A_i are the technology matrixes are defined in (1.2) and (1.3), and β measures the utility of the characteristics. The utility of the characteristics is defined in Table 5:

Table 5 Utility of characteristics of organic and conventional milk

Characteristic	Utility	Description
Common characteristics of milk:		
Org(w)	β_{ow}	Organic, whole
Whole	β_{cw}	Whole-milkiness
Org(ss)	β_{oss}	Organic, semi-skimmed
Semi	β_{css}	Semi-skimmedness
Org(s)	β_{os}	Organic, skimmed
Skim	β_{cs}	Skimmedness
Org	$\beta_{o,i}$	General organic, mixed with the normal distribution, $E\left(\beta_{o.i}\right)=\beta_0$
Milk	β_m	Milkiness
Household specific characteristics of organic milk:		
Environment	$\beta_{oEnveronmTr}$	Positive environmental effects of organic goods
Health	$\beta_{oHealthTr}$	Positive health effects of organic goods

In a discrete choice model the absolute utility of a given alternative is never observed, only which alternative yields the highest utility. This limits the identification in two dimensions. First of all, only the *difference* between the utility of two alternatives can be estimated, and secondly all parameters are only defined up to a *scale*. If the utility of all alternatives is multiplied with the same number it will have no effect on the choices observed. This is usually solved by normalising the variance of the utility in the RUM model, but it is crucial to remember that the absolute values of estimated parameters cannot be compared with results from other estimations. Only ratios such as the willingness to pay are identified. It is, however, possible to use the estimated parameters to tell whether the utility of one attribute is higher than the utility of another attribute, *within the same estimation.*

It is not possible to identify the utility of all eight characteristics in Table 5 and it is therefore chosen to restrict the utility of milkiness, semi-skimmednes and organic semi-skimmed to zero and measure the utility of the other characteristics relative to this. This means that the utility of the conventional version of the two other milk types is compared to semi-skimmed milk, and that the utility of the organic version is compared to the conventional version for each of the three types of milk, whole, semi-skimmed and skimmed. The utility of the part of the organic characteristic which depends on the type of milk is assumed to be the same for all households, whereas the utility of the part that is common for all types of milk is assumed to vary between different groups of the population, depending on relative income, degree of urbanisation, the highest level of education, age and presence of children in the household. The utility of environmental and health improvements related to organic milk is also assumed to be the same for all households.

The utility of choosing alternative j therefore becomes:

$$U_{it}(j) = \underbrace{\beta_p p_{jt}}_{price} + \underbrace{1_{\{j=organic\}}\beta_{o,i}}_{\substack{\text{individual specific part}\\ \text{of utility of common}\\ \text{organic characteristic}}} + \underbrace{1_{\{j=organic\}}U_{soc}^{org}(x_i)}_{\substack{\text{Variation in preferences for}\\ \text{the organic characteristic}\\ \text{beween socio-demogr. groups}}}$$
$$+\begin{bmatrix} 1_{\{j=1,Org\_Whole\}} \\ 1_{\{j=2,Conv\_Whole\}} \\ 1_{\{j=3,Org\_SemiSkim\}} \\ 1_{\{j=4,Conv\_SemiSkim\}} \\ 1_{\{j=5,Org\_Skimmed\}} \\ 1_{\{j=6,Conv\_Skimmed\}} \end{bmatrix}' \left(\underbrace{A'U(z)}_{\substack{\text{characteristics of}\\ \text{goods common}\\ \text{for all consumers}}} + \underbrace{A_i'U(z_{env}, z_{health})}_{\substack{\text{individual specific}\\ \text{characteristics of}\\ \text{organic goods}}} \right) \tag{1.11}$$

where $(-\beta_p)$ is the utility of money, p_{jt} is the imputed price of alternative j at time t, $1_{\{j=organic\}}$ is a dummy/indicator function indicating that j is an organic good, $\beta_{o,i}$ is the individual specific deviation from the average utility of the organic attribute (mixed with the normal distribution), U_{soc}^{org} is the part of the utility of the organic attribute which varies with socio-demographics and $1_{\{j=1,Org\_Whole\}}$ is a dummy indicating that j equals alternative 1, i.e. organic whole milk. A is the technology matrix for common characteristics defined in (1.2), A_i is the individual specific technology matrix defined in (1.3), indicating whether households trust in positive environmental or health effects.

The utility of the common characteristics of milk is given by the common technology matrix and the parameters defined in Table 5:

$$A'U(z) = \underbrace{\left[\begin{array}{l|cccccccc} & org(w): & whole: & org(ss): & semi: & org(s): & skim: & org: & milk: \\ \hline Org\_Whole: & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ Conv\_Whole: & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ Org\_SemiSkim: & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ Conv\_SemiSkim: & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ Org\_Skimmed: & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ Conv\_Skimmed: & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array}\right]}_{\substack{\text{characteristics of goods}\\ \text{common for all consumers}}} \underbrace{\begin{bmatrix} \beta_{ow} \\ \beta_{cw} \\ \beta_{oss} \equiv 0 \\ \beta_{css} \equiv 0 \\ \beta_{os} \\ \beta_{cs} \\ \beta_o \\ \beta_m \equiv 0 \end{bmatrix}}_{\substack{\text{utility, common}\\ \text{for all consumers}}} \tag{1.12}$$

The utility of the individual specific characteristics (environment and health) is given by the individual technology matrix and the remaining parameters of Table 5:

$$
\begin{aligned}
&A_i' U\left(z_{env}, z_{health}\right) = \\
&\left(\underbrace{\underbrace{\begin{bmatrix} 1_{\{Env,notrust\}} \\ 1_{\{Env,trust\}} \end{bmatrix}'}_{trust} \underbrace{\begin{bmatrix} \beta_{oEnvironmNoTr} \equiv 0 \\ \beta_{oEnvironmTr} \end{bmatrix}}_{utility}}_{environment} + \underbrace{\underbrace{\begin{bmatrix} 1_{\{Health,notrust\}} \\ 1_{\{Health,trust\}} \end{bmatrix}'}_{trust} \underbrace{\begin{bmatrix} \beta_{oHealthNoTr} \equiv 0 \\ \beta_{oHealthTr} \end{bmatrix}}_{utility}}_{health} \right) \underbrace{1_{\{j=organic\}}}_{\substack{only\ for \\ organic \\ goods}} = \\
&\left(\underbrace{1_{\{Env,trust\}} \beta_{oEnveronmTr}}_{environment} + \underbrace{1_{\{Health,trust\}} \beta_{oHealthTr}}_{health} \right) \underbrace{1_{\{j=organic\}}}_{\substack{only\ for \\ organic \\ goods}}
\end{aligned}
\tag{1.13}
$$

where $1_{\{Env,notrust\}}$ indicates no trust in positive environmental effects of organic goods and $\beta_{oEnvironmNoTr}$ is the level of utility obtained without environmental effects. This utility is assumed to be zero, meaning that $\beta_{oEnvironmTr}$ actually measures the difference in utility between believers and non-believers. Trust in positive health effects is treated the same way. $1_{\{j=organic\}}$ indicates that the utility is only obtained by purchasing organic types of milk.

The socio-demographic differences in the utility of the organic characteristic are:

$$
\begin{aligned}
U_{soc}^{org} = &\underbrace{\begin{bmatrix} 1_{\{L\_inc\}} \\ 1_{\{M\_inc\}} \\ 1_{\{H\_inc\}} \end{bmatrix}' \begin{bmatrix} \beta_{o\_L\_inc} \equiv 0 \\ \beta_{o\_M\_inc} \\ \beta_{o\_H\_inc} \end{bmatrix}}_{income} + \underbrace{\begin{bmatrix} 1_{\{Rural\}} \\ 1_{\{City\}} \\ 1_{\{Capital\}} \end{bmatrix}' \begin{bmatrix} \beta_{o\_Rural} \equiv 0 \\ \beta_{o\_City} \\ \beta_{o\_Capital} \end{bmatrix}}_{urbanisation} + \underbrace{\begin{bmatrix} 1_{\{No\}} \\ 1_{\{Vocal\}} \\ 1_{\{Short\}} \\ 1_{\{Medium\}} \\ 1_{\{Long\}} \end{bmatrix}' \begin{bmatrix} \beta_{o\_No} \equiv 0 \\ \beta_{o\_Vocal} \\ \beta_{o\_Short} \\ \beta_{o\_Medium} \\ \beta_{o\_Long} \end{bmatrix}}_{education} \\
&+ \underbrace{\begin{bmatrix} 1_{\{60+\}} \\ 1_{\{45-59\}} \\ 1_{\{30-44\}} \\ 1_{\{18-29\}} \end{bmatrix}' \begin{bmatrix} \beta_{o\_60} \equiv 0 \\ \beta_{o\_4559} \\ \beta_{o\_3044} \\ \beta_{o\_29} \end{bmatrix}}_{age} + \underbrace{\begin{bmatrix} 1_{\{NoCh\}} \\ 1_{\{Ch0-6\}} \\ 1_{\{Ch7-14\}} \end{bmatrix}' \begin{bmatrix} \beta_{o\_NoCh} \equiv 0 \\ \beta_{o\_Ch06} \\ \beta_{o\_Ch714} \end{bmatrix}}_{children}
\end{aligned}
\tag{1.14}
$$

where $1_{\{L\_inc\}}$ indicates relatively low income, $1_{\{Rural\}}$ indicates living in a rural municipality, $1_{\{No\}}$ indicates that no further education is stated for either of the adults in the household, $1_{\{18-29\}}$ indicates that the oldest adult is between 18 and 29 and $1_{\{NoCh\}}$ indicates that there are no children younger than 15 in the household. The control groups in (1.13) and (1.14) are defined in Table 3. Neither the socio-demographics, nor the perception of organic goods varies within the estimation period, and all parameters are assumed to be constant. The price of milk is therefore the only variable which varies from observation to observation.

For a household in the control group (i.e. trusting neither of the positive effects of organic production, with relatively low income, living in a rural municipality, no further education, aged 60 or more with no children younger than 15 years), the utility of e.g. organic whole milk (*j*=1) compared to conventional semi-skimmed milk is:

$$U_{it}\left(j=1\right)=\beta_p p_{1t}+\beta_{o,i}+\beta_{ow}+\beta_{cw}+\beta_o \tag{1.15}$$

The interpretation of this utility function is that the utility is composed of disutility of paying money for the milk ($\beta_p p_{it}$), plus the individual specific utility of the general organic attribute ($\beta_{o,i}$) plus the utility of the fact that the organic attribute comes from whole milk (β_{ow}), plus the utility of the whole attribute compared to the semi-skimmed attribute (β_{cw}), plus the average utility of the general organic attribute (β_o). Note that semi-skimmed milk is used as base for both the conventional and the organic attribute which means that the utility of the general organic attribute is actually the utility of organic semi-skimmed milk. The utility of conventional skimmed milk is therefore β_p, whereas the utility of organic skimmed milk is $\beta_p+\beta_{o,i}+\beta_o$.

The interpretation of U_{soc}^{org} is that households with different levels of education, income, urbanity, age and children have different levels of utility of organic milk in general, independent of milk type. The difference between the utility of organic or whole or skimmed milk and the utility of organic semi-skimmed milk is therefore assumed to be the same for all types of households, just as the difference in utility of different fat levels of conventional milk is assumed to be independent of socio-demographics.

This specification of the utility function means that it is assumed, that the utility of the organic attribute depends on the type of milk, that it varies between households and that it follows a normal distribution. As a further restriction, it is also assumed that the variance is identical for all three types of organic milk, i.e. the level of heterogeneity is the same.[12] The

[12] Estimations allowing the three types of organic milk to have different levels of variance proved to be highly unstable (lots of local maxima, even with Antithetic draws), thus the restriction of one common level of heterogeneity.

utility of the organic attribute is assumed to be a combination of a general utility of the organic attribute and a part which is allowed to vary with the type of milk and with various socio-demographics and perception of the organic good. By mixing only the general utility of the organic attribute it is therefore possible to achieve a multitude of mixed distributions with different means of the utility of the organic attribute.

The structure of the covariance matrix is thus assumed to be:

$$\begin{array}{cc} & \begin{array}{cccccc} \overbrace{\phantom{\sigma_o^2+\sigma_{cw}^2\equiv\sigma_o^2}}^{\text{Organic whole}} & \overbrace{\phantom{\sigma_{cw}^2\equiv 0}}^{\text{Conv. whole}} & \overbrace{\phantom{\sigma_o^2+\sigma_{css}^2\equiv\sigma_o^2}}^{\text{Org. semi-skim.}} & \overbrace{\phantom{\sigma_{css}^2\equiv 0}}^{\text{Conv. semi-skim.}} & \overbrace{\phantom{\sigma_o^2+\sigma_{cs}^2\equiv\sigma_o^2}}^{\text{Org. skimmed}} & \overbrace{\phantom{\sigma_{cs}^2\equiv 0}}^{\text{Conv. skimmed}} \end{array} \\ \Omega = & \begin{bmatrix} \sigma_o^2+\sigma_{cw}^2\equiv\sigma_o^2 & & & & & \\ 0 & \sigma_{cw}^2\equiv 0 & & & & \\ 0 & 0 & \sigma_o^2+\sigma_{css}^2\equiv\sigma_o^2 & & & \\ 0 & 0 & 0 & \sigma_{css}^2\equiv 0 & & \\ 0 & 0 & 0 & 0 & \sigma_o^2+\sigma_{cs}^2\equiv\sigma_o^2 & \\ 0 & 0 & 0 & 0 & 0 & \sigma_{cs}^2\equiv 0 \end{bmatrix} \end{array} \qquad (1.16)$$

This means that the estimated variance of the organic types is actually the *extra* variance induced by the organic attribute, compared to the variance of the conventional version.

The parameters for the utility of environmental and health effects, the parameters which measure the difference in the mean utility of the organic attribute between groups of households and the parameter for the utility of the general organic characteristic are assumed to follow a distribution where the mean varies between groups of households and the standard deviation is common for all groups. It is important to keep in mind that this approach implies that the *variation* in utility of the organic attribute for a given type of milk is assumed to be independent of socio-demographics and milk type, only the *mean* is allowed to vary. It is therefore not possible to make statements about whether the degree of heterogeneity differs between groups of households or types of milk, or whether the utility of the organic characteristic is correlated between different types of milk.

1.8. Estimation results

Table 6 presents the results of estimations using only socio-demographics (model 1 in Table 6) and including both socio-demographics and perception of organic goods (model 2 in Table 6). As mentioned in section 1.6, the scaling of a discrete choice model depends on the magnitude of the variance of the utility, and the results of two different estimations are therefore not directly comparable. In this specific case, however, the utility of money (the parameter for price) is the same in both estimations which means that the sign of the difference in willingness to pay (which is a ratio and therefore can be compared between estimations) can be elicited directly from the differences in the non-price parameters.

The definition of the utility function in (1.11) means that if perception of organic goods matters, the difference between the two estimated models should affect only the mean of the mixed organic attribute (because the control group becomes even more restrictive in the model including perception of organic goods) and the parameters of the socio-demographics (if socio-demographics are correlated with perceptions as indicated in Table 25 to Table 36 in Appendix D). It turns out that this is exactly the case. The parameters for price, type of milk, standard deviation of the utility of the general organic characteristic and the differences between the different types of organic milk are identical in the two models.

Distribution of utility:

In the model without perceptions of the organic attribute (model 1) the mean utility of the general organic characteristic is -5.29 and the parameter has a standard *error* of 0.486, which means that it is significantly different from zero at the 1 per cent level. The standard *deviation* of the mixing distribution of the utility of the general organic characteristic is 4.68. This parameter has a standard error of 0.167 which means that it is also significantly different from zero at the 1 per cent level. Together the two parameters show that in the control group (low income, rural municipality, no further education, 60 years old or more and no children) 13 per cent have a positive utility of the general organic characteristic as long as it is provided in organic semi-skimmed milk. The probability can be calculated from the estimated normal distribution: $P\left(x>0 \middle| x \sim N\left(-5.29, 4.68\right)\right) = 12.9\%$. If the organic characteristic is provided in skimmed milk instead the share with positive utility changes to 17 per cent because the mean utility is increased by 0.81: $P\left(x>0 \middle| x \sim N\left(-5.29+0.81, 4.68\right)\right) = 16.9\%$. Note that the negative mean of a mixed parameter is thus not synonymous with negative utility as it would have been in a conventional logit which assumes that everyone has the same utility. In a mixed logit a negative mean merely indicates that less than 50 per cent have a positive utility. This is an important difference between conventional and mixed logit.

Table 6 Estimation results

		Model 1: Without perceptions Estimate	St. err.		Model 2: With perceptions Estimate	St. err.		LR test, only model with perceptions
β_p ¤	Price	-0.24	(0.068)	***	-0.24	(0.068)	***	
	Type of milk							
β_{cw}	Whole	-0.82	(0.101)	***	-0.82	(0.101)	***	
β_{css}	*Semi-skimmed*							
β_{cs}	Skimmed	-0.86	(0.079)	***	-0.86	(0.079)	***	
	Mixed organic attribute							
$E(\beta_o)$	Mean	-5.29	(0.486)	***	-6.39	(0.489)	***	
σ_o	Standard deviation	4.68	(0.178)	***	4.36	(0.172)	***	
	Type of organic milk							
β_{ow}	Organic whole milk	-0.02	(0.167)		-0.02	(0.167)		
β_{oss}	*Org. semi-skimmed milk*							
β_{os}	Organic skimmed milk	0.81	(0.141)	***	0.80	(0.141)	***	
	Positive effect on environment							
$\beta_{oEnvironmNoTr}$	*Disagree or not sure*							
$\beta_{oEnvironmTr}$	Agree				1.00	(0.400)	**	$\chi^2_1(5.94) = 0.015$
	Positive effect on own or family's health							
$\beta_{oHealthNoTr}$	*Negative or no difference*							
$\beta_{oHealthTr}$	Agree				2.45	(0.401)	***	$\chi^2_1(35.96) = 0.000$
	Income							
$\beta_{o\_L\_inc}$	*Lowest 25%*							
$\beta_{o\_M\_inc}$	Mid 50%	0.19	(0.441)		0.28	(0.411)		$\chi^2_2(9.40) = 0.009$
$\beta_{o\_H\_inc}$	Highest 25%	1.57	(0.574)	**	1.48	(0.551)	**	
	Urbanisation							
β_{o_Rural}	*Rural municipality*							
β_{o_City}	Urban municipality	1.36	(0.411)	***	0.98	(0.417)	**	$\chi^2_2(33.51) = 0.000$
$\beta_{o\_Capital}$	Capital area	3.09	(0.509)	***	2.57	(0.461)	***	
	Education							
β_{o_No}	*No further educ. stated*							
β_{o_Vocal}	Voc.-oriented high-school	-0.06	(0.500)		-0.25	(0.439)		
β_{o_Short}	Short further education	1.66	(0.601)	**	1.33	(0.541)	**	$\chi^2_4(16.68) = 0.002$
β_{o_Medium}	Medium further education	1.15	(0.574)	**	0.85	(0.529)		
β_{o_Long}	Long further education	2.34	(0.859)	**	1.74	(0.775)	**	
	Age							
β_{o_60}	*60+*							
β_{o_4559}	45-59 years	-1.31	(0.463)	**	-1.13	(0.440)	**	$\chi^2_3(7.48) = 0.058$
β_{o_3044}	30-44 years	-0.53	(0.637)		-0.39	(0.575)		
β_{o_29}	18-29 years	-1.01	(0.886)		-0.76	(0.794)		
	Children							
β_{o_NoCh}	*No children*							
β_{o_Ch06}	Children 0-6 years	-0.91	(0.681)		-0.95	(0.621)		$\chi^2_2(2.42) = 0.298$
β_{o_Ch714}	Children 7-14 years	0.49	(0.655)		0.36	(0.639)		
	Number of observations		*33,993*			*33,993*		
	Number of households		*1,022*			*1,022*		
	Number of parameters		*20*			*22*		
	Log-likelihood value		*-39,930.4*			*-39,882.8*		

¤*:* Parameter labels are defined in equations (1.12), (1.13) and (1.14).

Italics means that the parameter is restricted to zero (control group).

Mixed logit with one normally distributed parameter using 2,500 Antithetic Halton draws based in the prime 2, and a convergence criterion of 10-4. Data source: GfK purchase data for milk June to December 2000 combined with background data covering 2000 and questionnaire data from 2002. Only whole, semi-skimmed and skimmed milk. '***' is significant at the 1% level, '**' at the 5% level and '*' at the 10% level. The LR tests show the results of comparing the complete model with a model excluding variables group by group.

When perception of organic goods is included in the estimation (Model 2), it means that the control group is restricted to households who expect no positive effects on either environment or health and are part of the control group in Model 1). The result is that the share with positive utility drops from 13 per cent to 7 per cent because the new mean and standard deviation lead to $P\left(x>0 \middle| x \sim N\left(-6.39, 4.36\right)\right) = 7.1\%$. The standard deviation of the utility of the organic attribute decreases a bit when perceptions are introduced into the model, again a natural effect since the difference in perceptions explain part of the variation in utility.

Comparing the two models:

The data used for this analysis make it possible to entangle the effects of attitudes from socio-demographics. When comparing the results of the two estimations it becomes clear that the utility of the organic characteristic which could easily be seen as a result of living in the capital area or having a long education, partly arises from the fact that these groups generally are more positive towards organic products than the rest of the population. The remaining extra utility of the organic characteristic for households in the capital area must either come from other attitudes not included in the estimation or from structural differences such as easier access to organic goods. This supports the hypothesis that attitudes are correlated with socio-demographics and indicates that part of the effect of socio-demographics observed in studies without information about perception of organic products ought to be ascribed to attitudes rather than socio-demographics.

Likelihood ratio tests on the most sophisticated model:

The likelihood ratio (LR) tests presented in the last column of Table 6 show the results of comparing the full model 2 with models where sets of parameters are restricted to zero. As an example, looking at the parameters for age shows that the difference in utility of the general organic characteristic is not significant between the groups 18-29 and 30-44 compared to those who are 60 years old or more. However, the difference between the group of 45-59 and the 60+ is significant at the 5 per cent level. The LR test shows that the effect of the dummies for the different age groups can be ignored without significant loss of explanatory power (the probability that the model without dummies for age is just as good as the model including age is 5.8 per cent). The effect of children is even less important as the probability of the LR test is 29.8 per cent, which clearly accepts the restricted model without children. The effect of trust in environment is close to being tested out of the model at the 1 per cent level, but the effects of health, income, urbanisation and education are all significant.

Comparing with other studies:

As mentioned in the introduction, several studies have investigated the motives for purchasing organic goods. Bonti-Ankomah and Yiridoe (2006) provide an excellent review of the literature including more theoretical contributions about the nature of organic goods. Most studies are based on relatively few respondents and/or stated consumption of organic goods. This paper distinguishes itself by using information about actual purchases (including prices of the purchased goods), socio-demographics and perception of organic goods for each of the 1,022 households in the sample. The results therefore yield information about the final result of the attitudes and purchase intentions reported in many other studies – namely the actual money put on the counter at the end of the day.

In the present study, the effect of trust in positive effects on *health* is bigger and more significant than the effect of trust in positive *environmental* effects. This corresponds with findings in Makatouni (2002)[13] and Magnusson et al. (2003).[14] As mentioned above, Makatouni (2002) found that health (personal or for their families) was the most important factor when trying to explain stated organic consumption. Environment and animal welfare were also important, but mainly through their impact on the health factor. Magnusson et al. (2003) found that health was the most important predictor of both attitudes towards organic products and purchase intention of these, and that the health factor also was an important predictor of the stated purchase frequency of the four target foods (organic milk, meat, potatoes, and bread). Magnusson et al. also found that perception of the environmental effects of organic foods contributed to the prediction of attitude towards the specific foods, but not to the prediction of stated purchase. The actual purchases under actual budget constraints and prices in this study therefore confirm the findings in studies using stated motives for purchase of organic goods, health seems to be more important than the environment, but environmental improvements are likely to be perceived as related to better human health, and therefore influence the purchase decision positively in a more indirect way than health.

The results on *income* vary. Some studies find a positive correlation between income and propensity to purchase organic products (e.g. Fotopoulos and Krystallis, 2002)[15] others find no

---

[13] Makatouni (2002): Results of qualitative interviews with 40 British parents, stated motives for purchasing organic foods.

[14] Magnusson et al. 2001 & 2003: Mail survey, 1,154 Norwegian respondents, stated consumption.

[15] Fotopoulos and Krystallis, 2002: Face to face interviews, 1,612 Greek respondents, stated purchasing behaviour.

significant differences (e.g. Wolf, 2002).[16] The present study finds a strong positive and significant effect of income, indicating that the lack of effect in stated behaviour studies might be due to the lack of budget restriction in the hypothetical settings.

The effect of *urbanisation* is rarely investigated, perhaps because many studies focus on specific geographical locations, without much variation in urbanity. However, this study proves that urbanisation is a crucial factor in explaining consumption of organic goods. Part of the effect of urbanisation can be ascribed to a positive correlation between trust in positive environmental and health effects of the organic attribute and degree of urbanisation, but even when controlling for the perception of organic goods, the effect of urbanisation is still very strong. The positive effect of urbanisation may partly be caused by structural differences between rural and urban municipalities, leading to a better supply of organic goods in urbanised municipalities. Another possible explanation is a "neighbouring" effect. The trust in positive effects of organic goods is more common in urbanised municipalities and may lead people to purchase organic goods simply because everybody else do so, independent of their own faith in organic products.

The effect of *education* also varies from study to study, but most studies find either an insignificant or a positive effect. One example is Magnusson et al. (2003) who find a positive and significant effect on stated purchase of organic milk, but not on meat, potatoes and bread. Some studies, however, find a negative effect of education on willingness to pay (e.g. Thompson and Kidwell 1998).[17] In the present study, the effect of education is positive, but not as significant as the effect of urbanisation. The organic purchase share is 20 per cent for households with no further education (control group) and between 34 and 42 per cent for households with short, medium or long further education (see Table 3). The difference between the control group and the non-control groups is therefore just as big as for the degree of urbanisation (20 per cent in the control group, 42 per cent in the capital area, see Table 3), but the likelihood ratio test of urbanisation (0.000) is stronger than the test for education (0.002). This might be because income seems to be more closely associated with education than with urbanisation.[18] Part of the difference in organic purchase share between educational

---

[16] Wolf, 2002: Personal interview of 342 randomly selected respondents at food stores in May 2001 in San Luis Obispo County, California, stated willingness to pay.

[17] Thompson and Kidwell 1998: Actual purchases and actual prices, 340 consumers, one shopping trip each, Tucson, Arizona, April 1994.

[18] Table 10 and Table 11 in Appendix C.

levels which is observed in simple one-way tables like Table 3 may therefore be caused by differences in income.

In the present study, *age* has no significant effect on the utility of the organic attribute, however, there is a significant difference between households aged 45-49 and households aged 60 years or more, in favour of the oldest households. This is surprising, because the organic purchase share is the same for the two groups (27 per cent, see Table 3). Again, the relationship between income and other socio-demographics becomes important. According to Table 12 in Appendix C the probability of belonging to the high income group is 48 per cent for the households aged 45-59, but only 10 per cent for the ones aged 60 or more. This means that the elderly households purchase organic goods to the same extent as the somewhat younger households in spite of the fact that they have considerably less money. Their utility of the organic attribute is therefore higher. The higher utility of elderly households might be explained by the findings in Wandel and Bugge (1997).[19] Based on stated purchasing motives Wandel and Bugge (1997) find that the importance of environmental effects was decreasing with age whereas the importance of health was increasing. In the present study the effect of trust in positive effects on environment and health is assumed to be the same for all households, and differences will therefore turn up as differences between socio-demographic groups e.g. depending on age. This could be worth exploring further in future research.

Most studies find a positive or insignificant effect of *children* in the household. McEachern and McClean (2002)[20] find that committed consumers who claim that they always buy organic products are more likely to have children, and Thompson and Kidwell (1998) find that children below 18 years old in the household increase the probability of choosing the organic version of certain vegetables. Magnusson et al. (2001) find no significant differences between respondents with and without children. In the present study both the observed difference in organic purchase shares (Table 3) and the estimated effect of children indicate that especially young children between 0 and 6 years have a *negative* effect on the propensity to purchase organic milk. The estimated effect on utility is not significantly different from zero, but the probability that the utility of the organic characteristic is higher for households with young children is only 6 per cent. The positive results of children in other studies can therefore not be confirmed here.

[19] Wandel and Bugge (1997): Personal interviews, 1,103 Norwegian respondents, stated willingness to pay and stated purchasing motives.

[20] McEachern and McClean (2002): Questionnaires answered by 200 Scottish consumers, stated consumption.

1.9. Conclusion

It appears that higher income, further education and especially living in an urban area has a significant positive effect on the probability of choosing organic milk over conventional. Age and presence of children do not have a significant effect. Compared to other studies it is interesting that the effect of young children is highly unlikely to be positive (6 per cent).

Believing that organic production has an effect on the environment increases the utility of the organic characteristic of organic milk, but not as much as believing in an effect on health. This corresponds with findings in other studies which indicate that the positive environmental effects of organic goods are perceived as an indicator of possible improvements in human health.

The effect of organic production on the environment and especially on human health is still being debated. This study shows that a considerable share of the population derives utility from environmentally friendly and especially healthy production. Proving these effects scientifically and thus making more people trust in them could be a fertile way of increasing the sale of organic goods.

References:

- Andersen, L.M. (2006): *Consumer Evaluation of Environmental and Animal Welfare Labelling*. AKF Working paper. Available at www.akf.dk .
- Bonti-Ankomah, S. and E. K. Yiridoe (2006): Organic and Conventional Food: A Literature Review of the Economics of Consumer Perceptions and Preferences. Report submitted to *Organic Agriculture Centre of Canada*, Nova Scotia Agricultural College. Available at http://www.organicagcentre.ca/ResearchDatabase/res_food_consumer.asp (verified 11 June 2008).
- Fotopoulos C. and A. Krystallis (2002): Purchasing motives and profile of the Greek organic consumer: a countrywide survey. *British Food Journal*, **104** (9): 730-765.
- Giannakas, K. (2002): Information Asymmetries and Consumption Decisions in Organic Food Product Markets. *Canadian Journal of Agricultural Economics*, **50** (1): 35-50.
- Greene, W. H. (1997): *Econometric analysis, Third edition*. Prentice-Hall International Inc.
- Hanemann, W. M. (1984): Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics*, **66** (3): 332-341.
- Lancaster, K. J. (1966): A New Approach to Consumer Theory. *The Journal of Political Economy*, **74 (**2): 132-157.
- Magnusson, M. K., A. Arvola, U. K. Hursti, L. Åberg and P. Sjödén (2001): Attitudes towards organic foods among Swedish consumers. *British Food Journal*, **103** (3): 209-226.
- Magnusson, M. K., A. Arvola, U. K. Hursti, L. Åberg and P. Sjödén (2003): Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behaviour. *Appetite*, **40**: 109-117.
- Makatouni, A. (2002): What motivates consumers to buy organic food in the UK? Results from a qualitative study. *British food Journal*, **104** (3/4/5): 345-352.
- McEachern, M. G. and P. McClean (2002): Organic purchasing motivations and attitudes: are they ethical? *International Journal of Consumer Studies*, **26** (2): 85-92.
- McFadden, D. and K. Train (2000): Mixed MNL models for discrete response. *Journal of Applied Econometrics*, **15**: 447-470.
- *OECD:* What are equivalence scales? Available at http://www.oecd.org/dataoecd/61/52/35411111.pdf (verified 11 June 2008).
- Revelt, D. and K. Train (1998): Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics*, **80:** 1-11.

- Thompson, G. D. and J. Kidwell (1998): Explaining the choice of organic produce: Cosmetic defects, prices and consumer preferences. *American Journal of Agricultural Economics* **80**: 277-287.
- Train, K. (1998): Recreation demand models with taste differences over people. *Land Economics*, **74** (2): 230-239.
- Train, K. (1999): Mixed logit models for recreation demand. In J.A. Herriges and C.L. Kling (eds.): *Valuing recreation and the environment: Revealed preference methods on theory and practice*, pp. 121-140. Edward Elgar, Cheltenham, UK.
- Wandel, M. and A. Bugge (1997): Environmental concern in consumer evaluation of food quality. *Food Quality and Preference*, **8** (1): 19-26.
- Wolf, M. M. (2002): *An Analysis of the Impact of Price on Consumer Interest in Organic Grapes and a Profile of Organic Purchasers*. Paper presented at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 28-31, 2002.

Appendix A: Data definitions

The variables used in the estimations are:

Trust in effect on environment and health:

The question was: 'To what extent do you agree with the following statements? … I think that the rules regarding organic production are good enough to create improvements for …'

- Nature, e.g. wild animals and plants
- My and my family's health

The respondents were allowed to answer on a five-point scale ranging from 'Totally disagree' to 'Totally agree'. The scales are transformed into two-point scales 'disagree or uncertain' and 'agree'. See Table 1 for more details.

Income:

The household income is stated in categories of 50,000 DKK between 0-99,999 DKK and 450,000+. These categories are equalised by the OECD-modified scale (OECD) (1 for first adult, 0.5 for next adult and 0.3 for all others). The result is split into three categories so that the lowest 25 per cent of the whole sample from 1997 to 2001 is labelled 'Low income', the middle 50 per cent is labelled 'Medium income' and the highest 25 per cent is labelled 'High income'. The measure is thus relative and has no meaning in an absolute sense, e.g. in relationship to poverty. For the sample in the last six months of 2000, 28 per cent of the households have 'Low income', 45 per cent have 'Medium income' and 27 per cent have 'High income'.

Urbanisation:

In 2000, Denmark was divided into 275 municipalities ('Kommuner'). GfK who collected the data has divided these municipalities into categories by degree of urbanisation. The categories are 'Capital area', 'City municipalities' and 'Rural municipalities'. For more on the urbanisation see Appendix F in Andersen (2006).

Education:
Highest level of further education after primary and lower secondary school for 7- to 16-year-olds ('folkeskolen') for the father or the mother. Separated into:

- *None stated* (27 per cent of the households)
- *Vocationally oriented high school* (34 per cent of the households)
    - Examples: Basic vocational courses ('EFG'), trainee ('elev'), apprentice ('lærling'), laboratory technician ('laborant'), nursing aide ('sygehjælper'), 'social- og sundhedsassistent'
- *Short further education* (17 per cent of the households)
    - Examples: Policeman, kindergarten teacher ('pædagogiske uddannelser'), technical school ('tekniske uddannelser')
- *Medium further education* (17 per cent of the households)
    - Examples: Teacher in the 'folkeskole', nurse (both of these are not university educations, but requires upper-secondary school ('gymnasium')), Bachelor
- *Long further education* (5 per cent of the households)
    - Examples: Various Master degrees (at least 5 years at the university after upper-secondary school)

Age:
Age is defined by the age of the oldest person in the household.

Children:
A dummy for children between 0 and 6 years in the household and a dummy for children between 7 and 14 years in the household.

Appendix B: Prices and market shares

Table 7 Imputed milk prices from June to December 2000

	Whole		Semi-skimmed		Skimmed	
	Organic	Convent.	Organic	Convent.	Organic	Convent.
Minimum	4.95	2.75	5.00	2.99	5.00	3.33
Mean	7.40	6.36	6.54	5.45	6.27	5.29
Median	7.32	6.36	6.50	5.43	6.25	5.32
Maximum	10.50	14.00	10.50	8.38	10.00	12.00

Source: GfK purchase data for milk June to December 2000. Only whole, semi-skimmed and skimmed milk.

Table 8 Purchase shares for different types of milk

	Number of observed purchases	Share of purchases
Whole	5,791	18.04
Semi-skimmed	16,948	52.78
Skimmed	9,370	29.98
Total	32,109	100.00

Source: GfK purchase data for milk June to December 2000. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 9 Organic purch. shares for diff. types of milk, and distribution of the org. market

	Share of milk type/ Share of the organic market
Whole	19.98 13.71
Semi-skimmed	21.04 42.26
Skimmed	39.66 44.03
Total	26.28 100.00

Source: GfK purchase data for milk June to December 2000. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Appendix C: Relationship between socio-demographics

Table 10 Relationship between income and urbanisation

Income	Urbanisation			Total
Frequency Per cent Row Pct Col Pct	Capital area	Urban municipality	Rural municipality	
Low	61 5.97 21.55 28.24	138 13.50 48.76 29.49	84 8.22 29.68 24.85	283 27.69
Medium	80 7.83 17.32 37.04	207 20.25 44.81 44.23	175 17.12 37.88 51.78	462 45.21
High	75 7.34 27.08 34.72	123 12.04 44.40 26.28	79 7.73 28.52 23.37	277 27.10
Total	216 21.14	468 45.79	338 33.07	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 11 Relationship between income and education

Income	Highest level of education within the household					Total
Frequency Per cent Row Pct Col Pct	No further education stated	Vocationally-oriented high school	Short further studies	Medium long further studies	Long further studies	
Low	143 13.99 50.53 51.62	95 9.30 33.57 27.38	28 2.74 9.89 16.28	13 1.27 4.59 7.39	4 0.39 1.41 8.00	283 27.69
Medium	110 10.76 23.81 39.71	165 16.14 35.71 47.55	93 9.10 20.13 54.07	81 7.93 17.53 46.02	13 1.27 2.81 26.00	462 45.21
High	24 2.35 8.66 8.66	87 8.51 31.41 25.07	51 4.99 18.41 29.65	82 8.02 29.60 46.59	33 3.23 11.91 66.00	277 27.10
Total	277 27.10	347 33.95	172 16.83	176 17.22	50 4.89	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 12 Relationship between income and age

Income	Highest age within the household				Total
Frequency Per cent Row Pct Col Pct	**18-29 years**	**30-44 years**	**45-59 years**	**60 years or more**	
Low	17 1.66 6.01 38.64	22 2.15 7.77 10.09	43 4.21 15.19 11.85	201 19.67 71.02 50.63	283 27.69
Medium	18 1.76 3.90 40.91	140 13.70 30.30 64.22	146 14.29 31.60 40.22	158 15.46 34.20 39.80	462 45.21
High	9 0.88 3.25 20.45	56 5.48 20.22 25.69	174 17.03 62.82 47.93	38 3.72 13.72 9.57	277 27.10
Total	44 4.31	218 21.33	363 35.52	397 38.85	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 13 Relationship between income and children 0-6 years old

Income	Children between 0 and 6 years		Total
Frequency Per cent Row Pct Col Pct	**No**	**Yes**	
Low	274 26.81 96.82 29.24	9 0.88 3.18 10.59	283 27.69
Medium	397 38.85 85.93 42.37	65 6.36 14.07 76.47	462 45.21
High	266 26.03 96.03 28.39	11 1.08 3.97 12.94	277 27.10
Total	937 91.68	85 8.32	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 14 Relationship between income and children 7-14 years old

Income	Children between 7 and 14 years		Total
Frequency Per cent Row Pct Col Pct	No	Yes	
Low	262 25.64 92.58 29.05	21 2.05 7.42 17.50	283 27.69
Medium	371 36.30 80.30 41.13	91 8.90 19.70 75.83	462 45.21
High	269 26.32 97.11 29.82	8 0.78 2.89 6.67	277 27.10
Total	902 88.26	120 11.74	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 15 Relationship between urbanisation and education

Urbanisation	Highest level of education within the household					Total
Frequency Per cent Row Pct Col Pct	No further education stated	Vocationally-oriented high school	Short further studies	Medium long further studies	Long further studies	
Capital area	47 4.60 21.76 16.97	75 7.34 34.72 21.61	39 3.82 18.06 22.67	43 4.21 19.91 24.43	12 1.17 5.56 24.00	216 21.14
Urban municipality	126 12.33 26.92 45.49	155 15.17 33.12 44.67	81 7.93 17.31 47.09	80 7.83 17.09 45.45	26 2.54 5.56 52.00	468 45.79
Rural municipality	104 10.18 30.77 37.55	117 11.45 34.62 33.72	52 5.09 15.38 30.23	53 5.19 15.68 30.11	12 1.17 3.55 24.00	338 33.07
Total	277 27.10	347 33.95	172 16.83	176 17.22	50 4.89	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 16 Relationship between urbanisation and age

Urbanisation	Highest age within the household				Total
Frequency **Per cent** **Row Pct** **Col Pct**	**18-29 years**	**30-44 years**	**45 59 years**	**60 years or more**	
Capital area	10 0.98 4.63 22.73	42 4.11 19.44 19.27	71 6.95 32.87 19.56	93 9.10 43.06 23.43	216 21.14
Urban municipality	22 2.15 4.70 50.00	100 9.78 21.37 45.87	167 16.34 35.68 46.01	179 17.51 38.25 45.09	468 45.79
Rural municipality	12 1.17 3.55 27.27	76 7.44 22.49 34.86	125 12.23 36.98 34.44	125 12.23 36.98 31.49	338 33.07
Total	44 4.31	218 21.33	363 35.52	397 38.85	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 17 Relationship between urbanisation and children 0-6 years old

Urbanisation	Children between 0 and 6 years		Total
Frequency **Per cent** **Row Pct** **Col Pct**	**No**	**Yes**	
Capital area	206 20.16 95.37 21.99	10 0.98 4.63 11.76	216 21.14
Urban municipality	432 42.27 92.31 46.10	36 3.52 7.69 42.35	468 45.79
Rural municipality	299 29.26 88.46 31.91	39 3.82 11.54 45.88	338 33.07
Total	937 91.68	85 8.32	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 18 Relationship between urbanisation and children 7-14 years old

Urbanisation	Children between 7 and 14 years		Total
Frequency **Per cent** **Row Pct** **Col Pct**	**No**	**Yes**	
Capital area	197 19.28 91.20 21.84	19 1.86 8.80 15.83	216 21.14
Urban municipality	418 40.90 89.32 46.34	50 4.89 10.68 41.67	468 45.79
Rural municipality	287 28.08 84.91 31.82	51 4.99 15.09 42.50	338 33.07
Total	902 88.26	120 11.74	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 19 Relationship between education and age

Highest level of education within the household	Highest age within the household				Total
Frequency **Per cent** **Row Pct** **Col Pct**	**18-29 years**	**30-44 years**	**45-59 years**	**60 years or more**	
No further education stated	13 1.27 4.69 29.55	28 2.74 10.11 12.84	68 6.65 24.55 18.73	168 16.44 60.65 42.32	277 27.10
Vocationally-oriented high school	19 1.86 5.48 43.18	90 8.81 25.94 41.28	125 12.23 36.02 34.44	113 11.06 32.56 28.46	347 33.95
Short further studies	7 0.68 4.07 15.91	49 4.79 28.49 22.48	74 7.24 43.02 20.39	42 4.11 24.42 10.58	172 16.83
Medium long further studies	4 0.39 2.27 9.09	38 3.72 21.59 17.43	73 7.14 41.48 20.11	61 5.97 34.66 15.37	176 17.22
Long further studies	1 0.10 2.00 2.27	13 1.27 26.00 5.96	23 2.25 46.00 6.34	13 1.27 26.00 3.27	50 4.89
Total	44 4.31	218 21.33	363 35.52	397 38.85	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 20 Relationship between education and children 0-6 years old

Highest level of education within the household	Children between 0 and 6 years		Total
Frequency **Per cent** **Row Pct** **Col Pct**	**No**	**Yes**	
No further education stated	272 26.61 98.19 29.03	5 0.49 1.81 5.88	277 27.10
Vocationally-oriented high school	310 30.33 89.34 33.08	37 3.62 10.66 43.53	347 33.95
Short further studies	148 14.48 86.05 15.80	24 2.35 13.95 28.24	172 16.83
Medium long further studies	159 15.56 90.34 16.97	17 1.66 9.66 20.00	176 17.22
Long further studies	48 4.70 96.00 5.12	2 0.20 4.00 2.35	50 4.89
Total	937 91.68	85 8.32	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 21 Relationship between education and children 7-14 years old

Highest level of education within the household	Children between 7 and 14 years		Total
Frequency **Per cent** **Row Pct** **Col Pct**	**No**	**Yes**	
No further education stated	260 25.44 93.86 28.82	17 1.66 6.14 14.17	277 27.10
Vocationally-oriented high school	295 28.86 85.01 32.71	52 5.09 14.99 43.33	347 33.95
Short further studies	146 14.29 84.88 16.19	26 2.54 15.12 21.67	172 16.83
Medium long further studies	155 15.17 88.07 17.18	21 2.05 11.93 17.50	176 17.22
Long further studies	46 4.50 92.00 5.10	4 0.39 8.00 3.33	50 4.89
Total	902 88.26	120 11.74	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 22 Relationship between age and children 0-6 years old

Highest age within the household	Children between 0 and 6 years		Total
Frequency Per cent Row Pct Col Pct	**No**	**Yes**	
18- 29 years	35 3.42 79.55 3.74	9 0.88 20.45 10.59	44 4.31
30-44 years	147 14.38 67.43 15.69	71 6.95 32.57 83.53	218 21.33
45-59 years	358 35.03 98.62 38.21	5 0.49 1.38 5.88	363 35.52
60 years or more	397 38.85 100.00 42.37	0 0.00 0.00 0.00	397 38.85
Total	937 91.68	85 8.32	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 23 Relationship between age and children 7-14 years old

Highest age within the household	Children between 7 and 14 years		Total
Frequency Per cent Row Pct Col Pct	**No**	**Yes**	
18- 29 years	44 4.31 100.00 4.88	0 0.00 0.00 0.00	44 4.31
30-44 years	142 13.89 65.14 15.74	76 7.44 34.86 63.33	218 21.33
45-59 years	320 31.31 88.15 35.48	43 4.21 11.85 35.83	363 35.52
60 years or more	396 38.75 99.75 43.90	1 0.10 0.25 0.83	397 38.85
Total	902 88.26	120 11.74	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 24 Relationship between children 0-6 years and children 7-14 years old

Children between 0 and 6 years	**Children between 7 and 14 years**		**Total**
Frequency **Per cent** **Row Pct** **Col Pct**	**No**	**Yes**	
No	848 82.97 90.50 94.01	89 8.71 9.50 74.17	937 91.68
Yes	54 5.28 63.53 5.99	31 3.03 36.47 25.83	85 8.32
Total	902 88.26	120 11.74	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Appendix D: Relationship between trust and socio-demographics

Question about Environment: "I think that the rules regarding organic production are good enough to create improvements for nature, e.g., wild animals and plants"

Table 25 Relationship between perception of effect on environment and income

Environment	Income			Total
Frequency **Per cent** **Row Pct** **Col Pct**	**Low**	**Medium**	**High**	
Uncertain or disagree	132 12.92 26.51 46.64	246 24.07 49.40 53.25	120 11.74 24.10 43.32	498 48.73
Agree	151 14.77 28.82 53.36	216 21.14 41.22 46.75	157 15.36 29.96 56.68	524 51.27
Total	283 27.69	462 45.21	277 27.10	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 26 Relationship between perception of effect on environment and urbanisation

Environment	Urbanisation			Total
Frequency **Per cent** **Row Pct** **Col Pct**	**Capital area**	**Urban municipality**	**Rural municipality**	
Uncertain or disagree	90 8.81 18.07 41.67	223 21.82 44.78 47.65	185 18.10 37.15 54.73	498 48.73
Agree	126 12.33 24.05 58.33	245 23.97 46.76 52.35	153 14.97 29.20 45.27	524 51.27
Total	216 21.14	468 45.79	338 33.07	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 27 Relationship between perception of effect on environment and education

Environment	Highest level of education within the household					Total
Frequency Per cent Row Pct Col Pct	No further education stated	Vocationally oriented high school	Short further studies	Medium long further studies	Long further studies	
Uncertain or disagree	148 14.48 29.72 53.43	158 15.46 31.73 45.53	89 8.71 17.87 51.74	86 8.41 17.27 48.86	17 1.66 3.41 34.00	498 48.73
Agree	129 12.62 24.62 46.57	189 18.49 36.07 54.47	83 8.12 15.84 48.26	90 8.81 17.18 51.14	33 3.23 6.30 66.00	524 51.27
Total	277 27.10	347 33.95	172 16.83	176 17.22	50 4.89	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 28 Relationship between perception of effect on environment and age

Environment	Highest age within the household				Total
Frequency Per cent Row Pct Col Pct	18-29 years	30-44 years	45-59 years	60 years or more	
Uncertain or disagree	19 1.86 3.82 43.18	106 10.37 21.29 48.62	186 18.20 37.35 51.24	187 18.30 37.55 47.10	498 48.73
Agree	25 2.45 4.77 56.82	112 10.96 21.37 51.38	177 17.32 33.78 48.76	210 20.55 40.08 52.90	524 51.27
Total	44 4.31	218 21.33	363 35.52	397 38.85	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 29 Relationship betw. Percept. of effect on environment and children 0-6 yrs old

Environment	Children between 0 and 6 years		Total
Frequency Per cent Row Pct Col Pct	No	Yes	
Uncertain or disagree	453 44.32 90.96 48.35	45 4.40 9.04 52.94	498 48.73
Agree	484 47.36 92.37 51.65	40 3.91 7.63 47.06	524 51.27
Total	937 91.68	85 8.32	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 30 Relationship betw. Percept. of effect on environment and children 7-14 yrs old

Environment	Children between 7 and 14 years		Total
Frequency Per cent Row Pct Col Pct	No	Yes	
Uncertain or disagree	434 42.47 87.15 48.12	64 6.26 12.85 53.33	498 48.73
Agree	468 45.79 89.31 51.88	56 5.48 10.69 46.67	524 51.27
Total	902 88.26	120 11.74	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Question about health: "I think that the rules regarding organic production are good enough to create improvements for my and my family's health"

Table 31 Relationship between perception of effect on health and income

Health	Income			Total
Frequency Per cent Row Pct Col Pct	**Low**	**Medium**	**High**	
Uncertain or disagree	171 16.73 28.31 60.42	279 27.30 46.19 60.39	154 15.07 25.50 55.60	604 59.10
Agree	112 10.96 26.79 39.58	183 17.91 43.78 39.61	123 12.04 29.43 44.40	418 40.90
Total	283 27.69	462 45.21	277 27.10	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 32 Relationship between perception of effect on health and urbanisation

Health	Urbanisation			Total
Frequency Per cent Row Pct Col Pct	**Capital area**	**Urban municipality**	**Rural municipality**	
Uncertain or disagree	112 10.96 18.54 51.85	264 25.83 43.71 56.41	228 22.31 37.75 67.46	604 59.10
Agree	104 10.18 24.88 48.15	204 19.96 48.80 43.59	110 10.76 26.32 32.54	418 40.90
Total	216 21.14	468 45.79	338 33.07	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 33 Relationship between perception of effect on health and education

Health	Highest level of education within the household					Total
Frequency Per cent Row Pct Col Pct	No further education stated	Vocationally-oriented high school	Short further studies	Medium long further studies	Long further studies	
Uncertain or disagree	181 17.71 29.97 65.34	205 20.06 33.94 59.08	94 9.20 15.56 54.65	99 9.69 16.39 56.25	25 2.45 4.14 50.00	604 59.10
Agree	96 9.39 22.97 34.66	142 13.89 33.97 40.92	78 7.63 18.66 45.35	77 7.53 18.42 43.75	25 2.45 5.98 50.00	418 40.90
Total	277 27.10	347 33.95	172 16.83	176 17.22	50 4.89	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 34 Relationship between perception of effect on health and age

Health	Highest age within the household				Total
Frequency Per cent Row Pct Col Pct	18-29 years	30-44 years	45-59 years	60 years or more	
Uncertain or disagree	27 2.64 4.47 61.36	130 12.72 21.52 59.63	213 20.84 35.26 58.68	234 22.90 38.74 58.94	604 59.10
Agree	17 1.66 4.07 38.64	88 8.61 21.05 40.37	150 14.68 35.89 41.32	163 15.95 39.00 41.06	418 40.90
Total	44 4.31	218 21.33	363 35.52	397 38.85	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 35 Relationship between perception of effect on health and children 0-6 years old

Health	Children between 0 and 6 years		Total
Frequency Per cent Row Pct Col Pct	**No**	**Yes**	
Uncertain or disagree	553 54.11 91.56 59.02	51 4.99 8.44 60.00	604 59.10
Agree	384 37.57 91.87 40.98	34 3.33 8.13 40.00	418 40.90
Total	937 91.68	85 8.32	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Table 36 Relationship between perception of effect on health and children 7-14 years old

Health	Children between 7 and 14 years		Total
Frequency Per cent Row Pct Col Pct	**No**	**Yes**	
Uncertain or disagree	533 52.15 88.25 59.09	71 6.95 11.75 59.17	604 59.10
Agree	369 36.11 88.28 40.91	49 4.79 11.72 40.83	418 40.90
Total	902 88.26	120 11.74	1022 100.00

Source: GfK background data covering 2000 and questionnaire data from 2002. Only households who purchased whole, semi-skimmed and skimmed milk during July to December 2000 and answered the AKF questionnaire in 2002.

Appendix E: List of tables

Animal Welfare and Eggs – Cheap Talk or Money on the Counter?*

Laura Mørch Andersen†

June 2008

Abstract

In this paper we utilize a unique combination of household level real purchase panel data and survey data on perceived public and private good attributes of different types of eggs to identify real willingness to pay for animal welfare using a mixed logit model. We find that consumers perceiving a stronger connection between animal welfare and the organic label have higher willingness to pay for organic eggs, even when we control for private good attributes such as food safety also connected to the label. Our results suggest that altruistic motives may play an important role in the demand for agricultural products.

Key words: Animal welfare, MMNL, market data, labelling, willingness to pay, altruism

2.1. Introduction

Do consumers really care about animal welfare, and are they willing to pay for increased animal welfare? The results in this paper suggest that consumers are willing to put money on the counter, and that the stated willingness to pay observed in opinion polls, hypothetical discrete choice experiments or contingent valuation studies is not just cheap talk.

According to the Eurobarometer Survey conducted in the beginning of 2005, 74% of European citizens believe that they can to some degree have a positive impact on the welfare of farm animals by buying animal-friendly products, and more than 60% state that they are willing to pay an additional price premium in order to do so (Eurobarometer 2005). The aim

* I thank Kenneth Train, David Revelt and Paul Ruud for allowing me to use their MMNL software, and for allowing me to use a later version, which allows for correlation between mixed parameters. I also thank Kenneth Train for his clear and speedy answers to my questions. I also thank Martin Browning and Wim Verbeke for useful comments on an earlier version of this paper.

The research was funded by the Danish Social Science Research Council (FSE).

† Contact information: Laura Mørch Andersen, AKF - Danish Institute of Governmental Research, www.akf.dk e-mail: LMA@akf.dk

of this paper is to estimate actual willingness to pay from observed purchases, in order to separate actual willingness to pay from cheap talk.

Interest in animal welfare has been increasing, both within the population in general and among the legislators who try to frame laws to match these new concerns. One example of this concern is the EU Action Plan to improve animal welfare (IP/05/698), which was adopted by the European Union in 2006. If society wishes to improve the level of animal welfare it may either prohibit production methods that lead to unacceptably low levels of animal welfare or it may improve market conditions for producers who use more animal-friendly production methods. Provision of information using labelling allows the more dedicated producers to signal that their production has a higher level of animal welfare than the standard production. Thereby, labelling offers a way of allowing consumers who actually gain utility from improved animal welfare to achieve this increase in utility without decreasing the utility of less caring consumers. The labelling of eggs described in section 2 is aimed at improving animal welfare.

However, since it is not possible to exclude others from enjoying the improved animal welfare induced by one's own purchase of a certified product this attribute is a public good, and therefore prone to free-riding which might undermine the effectiveness of labelling schemes. This potential problem has been addressed in two strands of literature both suggesting that there is a willingness to pay for animal welfare despite the potential free-rider problem. First of all, a number of contingent valuation studies (such as Rolfe (1999) and Bennet (1997)) find a positive stated willingness to pay for eggs with improved animal welfare. Though encouraging, these studies cannot distinguish real willingness to pay from cheap talk and there is a lingering suspicion that when consumers are put to the real market test free-riding kicks in.

Other studies such as Teisl et al. (2002) and Baltzer (2002) use market data. Both studies find positive (revealed) willingness to pay for animal welfare. Teisl et al. find positive willingness to pay for a label indicating dolphin-safe tuna catching and Baltzer finds positive willingness to pay for eggs carrying labels indicating improved animal welfare (non-battery eggs, see below). However, the suspicion here is that other 'private good' attributes like healthiness/safety of the product that consumers perceive as correlated with animal welfare may be driving behaviour.

Though the previous literature suggests this to be the case it is still not clear that consumers are willing to pay higher market prices for increased animal welfare even though it is a public good. The aim of this paper is to estimate actual willingness to pay for the animal welfare attribute from observed purchases. We utilise a unique dataset combining time series of actual purchase data for 2000 households with survey data on the same households with background information about the individual households along with information on the household perception of the organic label with respect to animal welfare and food safety. This allows us to compare willingness to pay between different socio-demographic groups as well as between groups with different perceptions of animal welfare in relation to organic eggs and food safety in relation to organic broilers. This means that we can establish whether the willingness to pay originates solely from 'private good' attributes, such as lower risk of falling ill, or if there is also willingness to pay for 'public good' attributes like animal welfare implying altruistic motives

The results in this paper suggest that consumers are willing to put money on the counter for animal welfare, and that the stated willingness to pay observed in opinion polls, hypothetical discrete choice experiments or contingent valuation studies is not just cheap talk.

The remainder of this paper is structured as follows. Section 2.2 presents the main differences between the egg labels applied in Denmark. In section 2.3 the data are described, while section 2.4 presents an introduction to the theory behind willingness to pay and the mixed multinomial logit model applied in the estimation of the model. Section 2.5 describes the practical problems of using market data at household level, and explains the solutions chosen. The results of the estimations are presented in section 2.6.

2.2. The Egg Labels

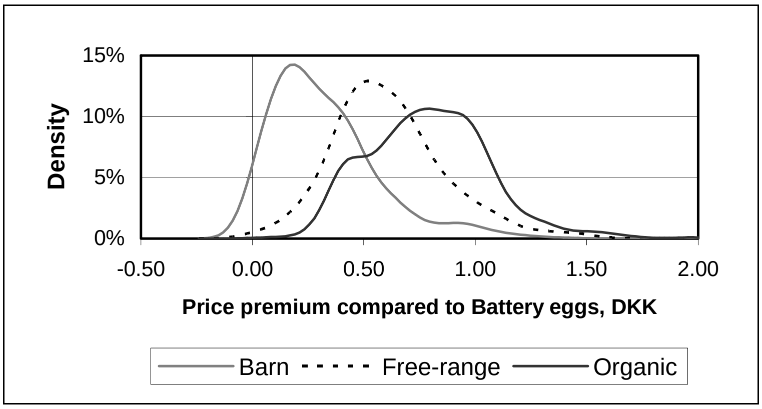
The Danish egg market is dominated by four different labels indicating production methods with different implications for animal welfare. In order to be allowed to bear a given label the production has to meet certain minimum standards, as described in various EU regulations. Table 1 shows the most important differences between the egg labels, and Figure 1 shows the distribution of price premiums compared to the price of battery eggs (the price of a battery egg is close to 1 DKK). For more details of the rules for different production types, see Andersen (2006).

Table 1 Main points of the rules for different production types

Egg label	Conditions for the egg-laying hens
Battery eggs	• Live in cages with 4 hens in each cage • 16 hens per m^2
Barn eggs	• Live in open barns • 7 hens per m^2
Free-range eggs	• Indoors: As for barn hens • Access to outdoor areas • 10 m^2 per hen on outdoor areas
Organic eggs	• 6 hens per m^2 indoors • Access to outdoor areas • 4 m^2 per hen on outdoor areas • Organic feed • No beak trimming

Source: The Danish Poultry Council.

Figure 1 Kernel density of imputed price premiums



Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Only households with answers to questionnaire. Nadaraya-Watson kernel regression estimator using the Gaussian kernel. Purchases made directly from farms excluded. Imputed prices are means of all observed prices within a given chain of stores and a given week. More information on this is provided in Section 5, 'Implementation of the model'.

Battery hens are usually considered to have the lowest level of animal welfare, because they are kept in small cages. Barn hens are allowed to move more freely, but do not have access to outdoor areas, and are therefore usually considered to be better off than battery hens, but worse off than free-range and organic hens. One of the differences between organic hens and free-range hens is that free-range hens may have their beaks trimmed, which is known to cause immediate and subsequent pain. The problem is that in extensive egg production systems, the risk of severe welfare problems such as injurious pecking and cannibalism is much greater in non-trimmed hens (ADAS/IGER/University of Bristol, 2001). Whether organic hens have a better quality of life than free-range hens is therefore sometimes debated, but apart from the differences in rules for production, organic eggs have the advantage of using a familiar label that is used on many different food products (the Danish 'Ø-label', which identifies organically-produced goods). Consumers have a generalised image of goods bearing the Ø-label, and do not have to spend time and energy studying new labels such as 'barn eggs' or 'free-range eggs'. In this paper it is therefore expected that willingness to pay for the different egg labels can be ranked as *battery, barn, free-range* and *organic*, where *battery eggs* are expected to yield the lowest willingness to pay and *organic eggs* are expected to yield the highest willingness to pay. As can be seen in Figure 1, the observed prices of the different types of eggs support this ranking.

2.3. Data

The data are from a Danish panel of approximately 2,000 households reporting all food purchases (GfK ConsumerScan Denmark, GfK). The panel is unbalanced and started in 1997. A substantial number of socio-demographics are collected once a year, and in 2002 a large questionnaire on organic food was issued to the panel. The purpose of the questionnaire was to obtain information about knowledge of and attitudes towards organic foods in general at household level. It is therefore possible to combine actual purchases with socio-demographics, attitudes and perception of specific organic goods. For more on the GfK data see Andersen (2006).

The data on eggs used in this paper cover the period from 26 June 1999 to 30 June 2000. In the analysis, the observed purchases in the GfK data are combined with the results of the 2002 questionnaire. If perceptions about eggs are assumed to be stable over time, the questionnaire makes it possible to use the information about household perceptions of the level of animal welfare and food safety in organic eggs, even though there is a time gap between the purchase data (1999-2000) and the questionnaire (2002). It is therefore possible not only to estimate willingness to pay for labels, but also to allow for different perceptions of the labels, and thereby for different purchasing motives. Among the 1,834 families who reported purchases of eggs during the period from June 1999 to June 2000, 878 families also answered the 2002 questionnaire, and 844 of these answered the questions used in this paper. As can be seen in Table 2, the households who answered the questionnaire represent the sample almost perfectly, at least as far as the overall distribution on types of eggs is concerned.

Table 2 Aggregate consumption of four different types of eggs

Households:	**All**	**With answers to questionnaire in general**	**With answers to both animal welfare and food safety**
Purchase shares:			
Battery eggs	47	47	47
Barn eggs	17	17	17
Free-range eggs	10	10	10
Organic eggs	27	26	26
Total	*100*	*100*	*100*
No. of purchases	20,676	11,178	10,800
No. of households	1,834	878	844

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Purchases made directly from farms excluded, see section 2.5.

Two of the questions in the questionnaire regarded perception of animal welfare related to eggs and food safety related to broilers. As can be seen in Table 3, very few households believe that organic production has a negative impact on animal welfare related to eggs or food safety related to broilers, but a substantial number of households believe it has positive effects. It also appears that trust in better animal welfare and improved food safety are correlated. It is, however, still possible to identify the effects on willingness to pay separately for animal welfare and food safety, as the correlation is not perfect. The answers to the two questions enter separately in the estimation, and the grey cross tabulation in Table 3 is merely included to illustrate the level of correlation. Willingness to pay among households with different perceptions of animal welfare and food safety is measured relative to the groups of households who perceive 'no difference' (control groups).

Table 3 Answers to questionnaire on perception of animal welfare and food safety[1]

No. of households (share of households):			How do you perceive the risk of falling ill with bacteria when you eat organic chicken?		
		Total	Higher (Negative organic food safety)	**No difference**	Lower (Positive organic food safety)
	Total	*844 (100%)*	*27 (3%)*	***571 (68%)***	*246 (29%)*
How do you perceive animal welfare for hens laying organic eggs?	Worse (Negative organic animal welfare)	*23 (3%)*	6 (1%)	9 (1%)	8 (1%)
	No difference	***355 (42%)***	15 (2%)	294 (35%)	46 (5%)
	Better (Positive organic animal welfare)	*466 (55%)*	6 (1%)	268 (32%)	192 (23%)

Source: AKF/GfK questionnaire from 2002.

Bold means Control group: Willingness to pay in the other groups is measured relative to this group. The estimated willingness to pay for households who perceive animal welfare to be better is the difference between the mean willingness to pay among households with perceived *positive* effect and households with *no* perceived effect.

[1] Note that the question about food safety is not related directly to organic eggs, but rather to organic chickens. However, the origin of food safety problems is the same in chickens and eggs (mainly salmonella during the period in question) and the answers are therefore used as a general indication of perception of food safety related to organic poultry, acknowledging that the signal cannot be expected to be as strong as for animal welfare.

One of the attractions of the GfK data is that it is possible to link actual purchases directly to socio-demographic information about individual households. This paper investigates how income, age, degree of urbanisation and level of education influence the willingness to pay for the different types of eggs. Each of the socio-demographic variables is split into sub-groups, and the willingness to pay within each sub-group is estimated relative to the control group indicated in Table 4.

Table 4 Socio-demographic data used in estimations

Variable	Sub-groups	Number of households	Share of households	Control group[a]
Income[b]	Lowest 25%	254	30	X
	Middle 50%	400	47	
	Highest 25%	190	23	
	Total	*844*	*100*	
Age[c]	18 to 44 years	230	27	
	45 to 59 years	304	36	
	60 years or more	310	37	X
	Total	*844*	*100*	
Degree of urbanisation[d]	Rural municipality	247	29	X
	Urban municipality	390	46	
	Capital area (Copenhagen)	207	25	
	Total	*844*	*100*	
Level of education[e]	No further education stated	206	24	X
	Vocationally oriented high-school	304	36	
	Short further education	138	16	
	Medium further education	150	18	
	Long further education	46	5	
	Total	*844*	*100*	

a: Willingness to pay in the other groups is measured relative to this group. The estimated willingness to pay in the Capital area is the difference between the mean willingness to pay in households in the Capital area and those in the rural municipalities. If the parameter for Capital area is significant, it means that the difference between the utility in the Capital area and that in the rural municipalities is significantly different from zero.
b: Income is recorded in brackets of DKK 50,000 (~€6,700). These brackets are divided by the number of persons in the household, weighted as 1 for the first adult, 0.5 for the next adults and 0.3 for children. Income is split into three categories indicating relative levels of income.
c: Age is defined by the age of the oldest person in the household.
d: GfK divides the 270 Danish municipalities into categories depending on how urbanised they are and on their geographical location. The geographical location is ignored here, and the sample is split into rural, urban and Capital area municipalities.
e: Highest level of education within the household.

2.4. Theory

Marginal willingness to pay is the amount of money a person is willing to pay in order to receive an extra unit of the good in question. The utility of household i from purchasing an egg of type j at time t is assumed to depend on the label j (β_i^j, constant over time, varies with household and type of egg). The vector of labels of the eggs purchased by household *i* at purchase 1 to T_i is called e_i. The utility also depends on the money spent purchasing the egg ($\beta^p p_{jt}$). The price vector *p* is allowed to depend on the egg label and the purchase, but the utility of money β^p is assumed to be constant over time, households and type of egg. The utility is not perfectly observed by the econometrician, and the utility therefore also depends on an unobserved error term ε_{ijt}. This is a Random Utility Model (RUM).

As in Hanemann (1984), the utility function is assumed to have the simple linear form

$$U_i\left(e_i, p\right) = \sum_{t=1}^{T_i}\left(\beta_i^j + \beta^p p_{jt} + \varepsilon_{ijt}\right) \tag{2.1}$$

and as in Hanemann (1984), the marginal willingness to pay is therefore the utility of the egg divided by the utility of money:

$$wtp_i^j = \frac{\partial U / \partial\left(egg^j\right)}{\partial U / \partial\left(money\right)} = \frac{\beta_i^j}{-\beta^p} \tag{2.2}$$

The error terms ε_{ijt} in (2.1) are assumed to be extreme value distributed, which means that the parameters can be estimated using a multinomial logit model.

However, the conventional multinomial logit model suffers from the assumption of Independence of Irrelevant Alternatives (IIA). In this application, IIA means that the probability of choosing a free-range egg versus the probability of choosing a battery egg should be independent of the presence of, for example, organic eggs on the market. This is very unlikely. Imagine that organic eggs left the market. Then the IIA in the multinomial logit model would imply that people who used to buy organic eggs would distribute themselves across the remainder of the egg labels according to the market share of those other egg labels. But people who buy organic eggs may very well have a higher propensity to buy free-range eggs, for example, than the population in general, and particularly to have a lower propensity to buy battery eggs. IIA is therefore not reasonable in this case.

Data show that some households buy organic eggs more frequently than others, which suggests variation in the household utility of organic eggs. To capture this variation and to avoid IIA it is therefore assumed that the household utility is drawn from a distribution (i.e. the household utility is known to the household, but only the distribution is observable to the econometrician). The household likelihood function then becomes the likelihood function in the conventional multinomial logit model integrated over all possible values of β:

$$L_i\left(\theta, e_i, p\right) = \int L_i^{conv}\left(\beta, e_i, p\right) f\left(\beta | \theta\right) d\beta \tag{2.3}$$

where $f(\beta|\theta)$ is the density of β given the parameters θ. The parameters θ of the distribution of the utility β are therefore estimated, instead of β itself. This is known as the Mixed MultiNomial Logit (MMNL) model (McFadden and Train, 2000). For applications of this model see for example McFadden and Train (2000), Revelt and Train (1998), Train (1998) or Train (1999). The MMNL model does not suffer from IIA, as long as at least one parameter is assumed to be drawn from a common distribution (mixed); see for example Train (1998).

In this paper it is assumed that the utility of the four types of eggs follows a multivariate normal distribution

$$\beta = \begin{bmatrix} \beta^1 \\ \beta^2 \\ \beta^3 \\ \beta^4 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix} \right) \tag{2.4}$$

where 1 is battery eggs, 2 is barn eggs, 3 is free-range eggs and 4 is organic eggs.

As usual in a discrete model, we can only estimate relative utility, which means that we estimate differences in utility (between types of eggs) and must choose an arbitrary normalisation to identify the scale. In order to estimate willingness to pay for eggs carrying labels indicating higher levels of animal welfare, the differences between the utility of battery eggs (β^1) and the utilities of all other types of eggs $(\beta^2, \beta^3, \beta^4)$ are estimated.

In the simplest version of the model the utility depends only on the type of egg purchased and the price paid:

$$\begin{aligned} U_i(e_i, p) - U_i(battery, p) &= \sum_{t=1}^{T_i} \left(\beta_i^j + \beta^p p_{jt} + \varepsilon_{ijt} \right) - \sum_{t=1}^{T_i} \left(\beta_i^1 + \beta^p p_{1t} + \varepsilon_{i1t} \right) \\ &\equiv \sum_{t=1}^{T_i} \left(\tilde{\beta}_i^j + \beta^p \left(p_{jt} - p_{1t} \right) + \tilde{\varepsilon}_{ijt} \right) \end{aligned} \tag{2.5}$$

where e_i is the vector of labels of the eggs purchased by household i at purchases 1 to T_i, β^p is the utility of money, β_i^j is the utility of an egg of type j for household i, and $\tilde{\beta}_i^j$ is the difference in utility between type j and battery eggs. As in the conventional logit, the problem of the scale is solved by normalising the variance of the extreme value distributed error terms (the ε 's) to $\pi^2/6$.

The estimated parameters of the distribution of differences in utility are

$$\tilde{\beta}=\begin{bmatrix}\beta^2-\beta^1\\ \beta^3-\beta^1\\ \beta^4-\beta^1\end{bmatrix}=\begin{bmatrix}\tilde{\beta}^2\\ \tilde{\beta}^3\\ \tilde{\beta}^4\end{bmatrix}\sim N\left(\begin{bmatrix}b_{21}\\ b_{31}\\ b_{41}\end{bmatrix},\begin{bmatrix}\tilde{\sigma}_{22} & \tilde{\sigma}_{23} & \tilde{\sigma}_{24}\\ \tilde{\sigma}_{23} & \tilde{\sigma}_{33} & \tilde{\sigma}_{34}\\ \tilde{\sigma}_{24} & \tilde{\sigma}_{34} & \tilde{\sigma}_{44}\end{bmatrix}\right) \tag{2.6}$$

Because the utility of each type of egg is assumed to be normally distributed, the differences between the utilities are also normally distributed since $X\sim N(\mu,\Sigma)\Rightarrow g(X)\sim N\Big(g(\mu),(\nabla g)\Sigma(\nabla g)'\Big)$. This means that the relationship between the estimated parameters from (2.6) and the structural parameters of the utility function (2.4) is

$$\begin{bmatrix}\beta^2-\beta^1\\ \beta^3-\beta^1\\ \beta^4-\beta^1\end{bmatrix}\sim N\left(\begin{bmatrix}\mu_2-\mu_1\\ \mu_3-\mu_1\\ \mu_4-\mu_1\end{bmatrix},\begin{bmatrix}\sigma_{22}+\sigma_{11}-2\sigma_{12} & \sigma_{23}+\sigma_{11}-\sigma_{13}-\sigma_{12} & \sigma_{24}+\sigma_{11}-\sigma_{14}-\sigma_{12}\\ \sigma_{23}+\sigma_{11}-\sigma_{12}-\sigma_{13} & \sigma_{33}+\sigma_{11}-2\sigma_{13} & \sigma_{34}+\sigma_{11}-\sigma_{14}-\sigma_{13}\\ \sigma_{24}+\sigma_{11}-\sigma_{12}-\sigma_{14} & \sigma_{34}+\sigma_{11}-\sigma_{13}-\sigma_{14} & \sigma_{44}+\sigma_{11}-2\sigma_{14}\end{bmatrix}\right) \tag{2.7}$$

It is important to note that the structural parameters are not identified, only the relative parameters in (2.6). The estimated variances and covariances do not describe the utility of the different types of eggs, but rather the ‘utility premium’ compared to battery eggs.

In Table 3 and Table 4 the sample was divided into sub-groups on the basis of perception of eggs and socio-demographic factors. Each household belongs to one of the sub-groups for each of the two questionnaire answers and for each of the four socio-demographic variables. The effect of the background variables is estimated non-parametrically by including a dummy for each sub-group. Disregarding the six control groups, the number of dummies is two for perception of animal welfare, two for perception of food safety, two for income, two for age, two for degree of urbanisation and four for education, i.e. a total of 14. See also Table 2 and 4. The simple utility function in (2.5) then becomes

$$\begin{aligned}&U_i(e_i,p)-U_i(battery,p)=\\ &\sum_{t=1}^{T_i}\left(\beta_i^j+\sum_{k=1}^{14}1_{[i\in k]}\beta_k^j+\beta^p p_{jt}+\varepsilon_{ijt}\right)-\sum_{t=1}^{T_i}\left(\beta_i^1+\sum_{k=1}^{14}1_{[i\in k]}\beta_k^1+\beta^p p_{1t}+\varepsilon_{i1t}\right)\equiv\\ &\sum_{t=1}^{T_i}\left(\tilde{\beta}_i^j+\sum_{k=1}^{14}1_{[i\in k]}\tilde{\beta}_k^j+\beta^p\left(p_{jt}-p_{1t}\right)+\tilde{\varepsilon}_{ijt}\right)\end{aligned} \tag{2.8}$$

where $1_{[i\in k]}$ is an indicator function indicating whether household *i* belongs to socio-demographic group *k*. Note that $\tilde{\beta}_i^j$ is the household specific utility of type *j* (for the control group). The individual value is drawn from a normal distribution. Only the mean and the standard deviation of the distribution is estimated, not the individual betas. The estimated parameters for socio-demographics therefore describe the *mean* difference between the utility of households in socio-demographic group *k* and the control group. This is illustrated in Figure 3 on page 82.

The utility of barn and free-range eggs is assumed to depend only on socio-demographics. For a household with a high income (inc=H), aged 45 to 59 (age=45-59), living in the Capital area (urb=Cap) and having a long further education (edu=long) the utility of purchasing a barn egg (*j*=2) at time *t* is:[2]

$$\begin{aligned} &U_i\left(organic, p_t\right) - U_i\left(battery, p_t\right) = \\ &\tilde{\beta}_i^{barn} + \tilde{\beta}_{inc=H}^{barn} + \tilde{\beta}_{age=45-59}^{barn} + \tilde{\beta}_{urb=Cap}^{barn} + \tilde{\beta}_{edu=long}^{barn} + \beta^p\left(p_{2t} - p_{1t}\right) + \tilde{\varepsilon}_{i2t} \end{aligned} \tag{2.9}$$

As described in table 3 we have answers to questions about perception of animal welfare related to organic eggs and food safety related to organic broilers. This means that we can separate private utility (food safety) from altruistic utility (animal welfare) when it comes to organic eggs. It is therefore possible to determine whether altruistic motives actually play a significant role in the willingness to pay for organic eggs.

It is assumed that the effect of trust in animal welfare or food safety is the same for all socio-demographic groups, and the utility of the public good (animal welfare) and the private good (food safety) is therefore added to the utility function without any interaction terms with socio-demographics. No perceived difference is used as control group. If a household with the same characteristics as in (2.9) perceives the animal welfare as better for organic eggs and the food safety as worse the utility is therefore modelled as:

$$\begin{aligned} &U_i\left(organic, p_t\right) - U_i\left(battery, p_t\right) = \\ &\tilde{\beta}_i^{org} + \tilde{\beta}_{animal+}^{org} + \tilde{\beta}_{safety\div}^{org} + \tilde{\beta}_{inc=H}^{org} + \tilde{\beta}_{age=45-59}^{org} + \tilde{\beta}_{urb=Cap}^{org} + \tilde{\beta}_{edu=long}^{org} + \beta^p\left(p_{4t} - p_{1t}\right) + \tilde{\varepsilon}_{i4t} \end{aligned} \tag{2.10}$$

This definition of the utility function means that the variance of utility is assumed to be the same in all subsets of the population; only the mean is allowed to vary between groups of households.

[2] Note that $\tilde{\beta}_i^{barn}$ is the household specific utility of barn eggs (for the control group). The individual value is drawn from a normal distribution. Only the mean and the standard deviation of the distribution is estimated, not the individual betas.

2.5. Implementation of the Model

Only the price of the chosen egg is observed, not the price of the alternatives, nor which alternatives are present in the purchase situation. The prices are therefore imputed as the mean of all observed prices of eggs with a given label within a given week in the chain of stores in which the purchase was actually made.

There are many unknown attributes of the purchased egg. The size of the egg is not recorded, and the store in which the purchase was made is only recorded at chain level. The freshness of the eggs is also unknown. These factors all contribute to unobserved heterogeneity in the prices. Using the observed price as an estimate of the price of the egg that was purchased, and comparing this price to mean prices for the types of eggs that were not purchased (by this household on this occasion) would mean that one was comparing the price of an egg of a given size, purchased in a given store and having a given freshness, with the price of an egg with a mixture of sizes, a mixture of stores and a mixture of different degrees of freshness. This would disturb the estimated effect of the prices, and thereby the estimated effect of the labels and other variables entering the model. It was therefore decided to impute all of the prices, including the price of the egg that was purchased.

The definition of the choice set is also important. It may not be reasonable to expect eggs with all labels to be present in all purchase situations.[3] If eggs with a given label are not present, the label is said to be rationed. If rationing occurs, but is not revealed, it might mean that a person is perceived as choosing not to buy eggs with a specific label even though this label might have been preferred if it had been present. This will lead to a lower estimate of marginal willingness to pay for this label. This is an important fact to keep in mind when interpreting the results of the estimations, especially for barn and free-range eggs that have relatively low purchase shares. In this application, eggs with a specific label are assumed to be rationed if nobody purchased eggs with this label in the relevant group of stores during the week in question.

The mixed multinomial logit models are estimated using a modified version of a programme developed by Kenneth Train, David Revelt, and Paul Ruud. This is an extension of the programme used in for example Revelt and Train (1998) and Train (1998). The extension allows estimation of correlations between normally distributed parameters. One of the virtues

[3] In some purchase situations the labels are not necessarily certified and/or no alternative can be expected to be available. This is e.g. the case for purchases directly from farms. These purchases are therefore excluded from the analysis, along with purchases where the price of battery eggs cannot be imputed.

of this programme is that it takes account of the panel structure of the data. In this paper the simple Halton draws used in the extended programme by Train, Revelt and Ruud are replaced by antithetic Halton draws. This practically eliminates the noise in the log-likelihood values of different models, and thereby improves the reliability of the Likelihood Ratio tests.

The utility of money is assumed to be the same for all households, whereas the utility of eggs with different labels is assumed to follow a multivariate normal distribution. This implies that the estimated marginal willingness to pay is also assumed to be normally distributed. In MMNL language this means that the price parameter is fixed, and the reactions to egg labels are mixed. The utility of money is probably not the same for everyone, but in this case it is a question of semantics. It is not possible to tell whether the difference in willingness to pay origins from differences in utility of money or from utility of non-battery labels. The assumption that everyone has the same utility of money whereas the utility of labels is normally distributed is merely a convenient way of assuming that the willingness to pay is normally distributed.

2.6. Results

First, the model is estimated using only the price and the type of egg as explanatory variables. This version illustrates the results that could be obtained from data with no information on socio-demographics. To illustrate the difference between a conventional and a mixed logit, the results of a conventional model are compared with a mixed version of the same model. The conventional model is rejected, and information about socio-demographic factors and perception of animal welfare and food safety is then included in the mixed model and the results are discussed.

The mixed multinomial logit estimates a distribution of the mixed parameters. The standard deviation of the normal distribution can be used as a measure of the degree of heterogeneity related to the utility of a given type of egg compared to battery eggs, and thereby also to the degree of heterogeneity of willingness to pay. The estimated correlations indicate the extent to which a high willingness to pay for e.g. organic eggs compared to battery eggs is correlated with a high willingness to pay for other types of eggs compared to battery eggs.

The main hypotheses are:

the ranking of willingness to pay for organic, free-range and barn eggs compared to battery eggs follows the animal welfare ranking, which means that willingness to pay is highest for organic eggs and lowest for barn eggs

the correlation between willingness to pay for different types of eggs is highest between organic and free-range eggs, because the production methods are very similar, and lowest between organic and barn eggs (but the correlation is still expected to be positive)

the organic label is familiar from other goods and to some people it also includes a health aspect. This means that there are more potential sources of willingness to pay for organic eggs than for free-range and barn eggs, which only differ from battery eggs in terms of animal welfare. (The variance of a sum is the sum of the variances *plus* twice the covariance). The different sources of willingness to pay are expected to be positively correlated (people who believe that organic products are healthier are more familiar with the organic label). The degree of heterogeneity is therefore expected to be greater for organic eggs than for the other types

households which perceive animal welfare as better for hens laying organic eggs are willing to pay more for them even when perception of food safety is controlled for. (Perception of food safety is observed to be positively correlated with perception of animal welfare, but is a private attribute (non-altruistic)).

Table 5 compares the result of the conventional logit with the results of the simplest mixed logit. In all of the estimated models the utility of price is negative and significantly different from zero, which means that the utility of money is positive, as expected. In the conventional logit the ranking of willingness to pay comes directly from the estimated parameters of the utility function. These are all negative, which means that the willingness to pay for non-battery eggs is lower than the willingness to pay for battery eggs. As an example, the willingness to pay for organic eggs compared to battery eggs is $-0.21/-(-0.45) = -0.47$. The conventional logit thus suggests that all households prefer to buy battery eggs unless the organic eggs are DKK 0.47 cheaper. At a first glance this is somewhat contra intuitive, as the price of non-battery eggs is usually higher than the price of battery eggs. But what it actually means is that the price difference is not enough to explain the low purchase shares of non-battery eggs. The logit model therefore estimates negative utility of the labels. The willingness to pay is higher for organic eggs than for barn eggs, as expected, but the willingness to pay for free-range eggs is lower than for barn eggs.

Table 5 Results of estimations based on all households, including only type of egg and price (Model 1)

		Conventional logit			Mixed logit		
		Estimate	**SD**	**Signific.**	**Estimate**	**SD**	**Signific.**
Price		-0.45	0.032	***	-0.39	0.122	***
Type of egg, utility relative to utility of battery eggs							
Means:	Organic	-0.21	0.026	***	-1.72	(0.200)	***
	Free-range	-1.17	0.024	***	-1.40	(0.134)	***
	Barn	-0.77	0.013	***	-0.80	(0.094)	***
Variance:	Organic				20.73	(1.819)	***
	Free-range				7.56	(0.706)	***
	Barn				4.31	(0.395)	***
Correlation:	(Organic, free-range)				0.84		
	(Organic, barn)				0.66		
	(Free-range, barn)				0.79		
Log-likelihood			-12,950			-8,385	
No. of households			844			844	
No. of observations			10,800			10,800	

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Purchases made directly from farms excluded. Rationing is allowed. Number of antithetic Halton Draws is 7,500. '***' is significant at the 1% level.

This does not correspond with the expectation that willingness to pay for free-range eggs should lie between the willingness to pay for barn eggs and that for organic eggs. On the other hand, it fits well with the fact that free-range eggs have the lowest market share (see Table 2). One explanation is that households may find it difficult to distinguish free-range eggs from barn and organic eggs. If a household believes that there is no difference between barn and free–range eggs, barn eggs will be chosen because they are cheaper. If a household believes that there is almost no difference between free-range and organic eggs, organic eggs are more likely to be chosen, because organic eggs have a familiar label and may even be perceived as healthier, and are often not more expensive than free-range eggs. Baltzer (2002), who used scanner data from individual COOP stores, also found that the willingness to pay for free-range eggs was lower than for organic and barn eggs.

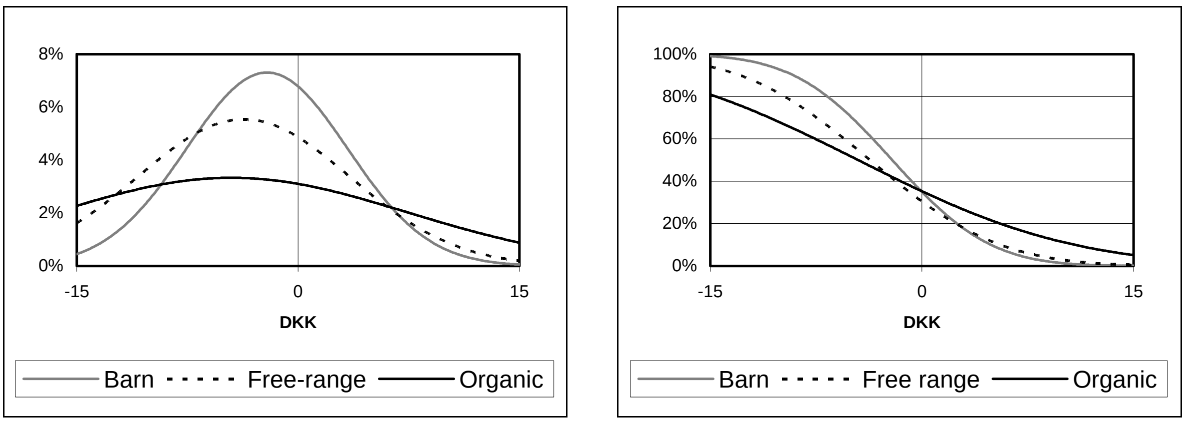
A conventional logit can be seen as the special case of a mixed logit in which all standard deviations are zero. It is therefore possible to test the need for mixing by a likelihood ratio test with degrees of freedom equal to the number of mixed parameters. In the example in Table 5 the likelihood ratio test becomes $-2\cdot\left(-12,950+8,385\right)=9,130$. The degrees of freedom are equal to the number of parameters in the variance covariance matrix in the mixed model i.e. six in this case. The conventional logit is therefore *strongly* rejected. The estimated negative willingness to pay underlines the fact that the conventional logit does a very poor job of explaining the willingness to pay, because it estimates *one* willingness to pay for all households. The rest of the paper therefore focuses on the mixed model.

In a mixed multinomial logit model both the mean and the variance-covariance matrix of the willingness to pay are estimated, so the ranking of willingness to pay now depends on the share of the population who are willing to pay a given percentage extra, compared to the cost of a battery egg. The expectation is that the share of the population with a given willingness to pay for non-battery eggs is largest for organic eggs and smallest for barn eggs.

The mean willingness to pay in the mixed logit in Table 5 becomes negative for all three types of eggs, but now this simply means that the share of households with positive willingness to pay is less than 50%, and this does not seem unreasonable given the market shares presented in Table 2 (between 10% and 26%). The estimated share of households with positive willingness to pay is 35% for organic as well as barn eggs, and 31% for free-range eggs. The densities of willingness to pay are illustrated in Figure 2. This is a very clear illustration of the importance of standard deviations in a mixed logit. As soon as the price premium becomes positive the share of households who are willing to pay the premium is larger for organic eggs than for barn eggs, even though the mean willingness to pay is lowest for organic eggs.

Figure 2 Illustration of the importance of the standard deviation of willingness to pay (WTP) for different types

Density of normally distributed WTP **Share with WTP higher than x DKK**



The willingness to pay for organic eggs has the lowest mean, but the highest standard deviation. In this case the bigger standard deviation implies that the share of the population with willingness to pay higher than a given positive amount is bigger for organic eggs, even though the mean was lower than for the other types. In this specific case the share with willingness to pay higher than zero is exactly the same as for barn eggs, but that is mere coincidence.

The mixing changes not only the magnitude, but also the ranking of the means. However, as illustrated in Figure 2, the ranking of the means is not necessarily the same as the ranking of willingness to pay. This difference between conventional and mixed logit is important to keep in mind whenever one tries to interpret results of a mixed logit.

The standard deviation of the willingness to pay for organic eggs is 4.6 ($20.73^{½}$) and the standard deviations for free-range and barn eggs are 2.7 and 2.1. This supports the hypothesis that the organic label suggests other attributes in addition to animal welfare. As expected, the estimated correlation between organic eggs and free-range eggs is larger than the other correlations, which might indicate that households know that free-range and organic eggs are very close substitutes. The correlation between barn and free-range eggs is also higher than the correlation between barn and organic eggs, confirming that barn eggs and free-range eggs are closer substitutes than barn eggs and organic eggs.

The simple mixed model in Table 5 thus confirms the hypothesis that some share of the population has positive willingness to pay for non-battery eggs, that willingness to pay for barn eggs is lower than for organic eggs, and that the variation in willingness to pay for organic eggs is higher than the willingness to pay for barn and free-range eggs.

The mixed model from Table 5 is repeated in Table 6, together with a model where socio-demographics and perceptions of animal welfare and food safety are included. This model splits the sample into sub-samples with different willingness to pay, as explained in equation(2.8). The new variables are allowed to influence the mean utility of each type of egg, but not the standard deviations. That means that the model estimates differences in mean willingness to pay between households with different perceptions of eggs, and between different socio-demographic groups. The effect of perceptions of animal welfare and food safety is only allowed to influence the willingness to pay for organic eggs, whereas the socio-demographics are allowed to influence the willingness to pay differently for each of the three types of non-battery eggs.

Table 6 Summary of mixing results

		Model 1: Only types		Model 2: With perceptions and socio-demographics		
Explanatory variable:		**Estimate**	**St. dev.**	**Estimate**	**St. dev.**	**LR test**
Price		-0.39	(0.122)***	-0.38	(0.122) ***	
Types of eggs, measured relative to battery eggs						
Means:	Organic	-1.72	(0.200)***	-4.00	(0.512) ***	
	Free-range	-1.40	(0.134)***	-2.43	(0.333) ***	
	Barn	-0.80	(0.094)***	-0.84	(0.241) ***	
Variance:	Organic	20.73	(1.819)***	17.16	(1.552)***	
	Free-range	7.56	(0.706)***	6.77	(0.687) ***	
	Barn	4.31	(0.395)***	4.17	(0.383) ***	
Correlation:	(Organic, free-range)	0.84		0.81		
	(Organic, barn)	0.66		0.67		
	(Free-range, barn)	0.79		0.71		
Perception of animal welfare in organic eggs, no difference is control group						
Organic	Negative organic animal welfare			-0.17	(0.727)	$\chi_2^2(15.24) = 0.000$
	Positive organic animal welfare			0.99	(0.285) ***	
Perception of food safety in organic chickens, no difference is control group						
Organic	Negative organic food safety			-0.55	(0.697)	$\chi_2^2(12.21) = 0.002$
	Positive organic food safety			0.91	(0.274) ***	
Income, lowest 25% is control group						
Organic	Mid 50%			0.24	(0.329)	$\chi_2^2(5.56) = 0.062$
	Highest 25%			0.71	(0.393) *	
Free-range	Mid 50%			0.08	(0.230)	$\chi_2^2(4.95) = 0.084$
	Highest 25%			0.49	(0.278) *	
Barn	Mid 50%			0.10	(0.189)	$\chi_2^2(0.68) = 0.712$
	Highest 25%			0.18	(0.239)	
Age, 60+ is control group						
Organic	Age 18 to 44			-0.88	(0.392) **	$\chi_2^2(5.27) = 0.072$
	Age 45 to 59			-0.53	(0.305) *	
Free-range	Age 18 to 44			-0.49	(0.276)*	$\chi_2^2(4.53) = 0.104$
	Age 45 to 59			-0.47	(0.248) *	
Barn	Age 18 to 44			-0.70	(0.213)***	$\chi_2^2(10.18) = 0.006$
	Age 45 to 59			-0.43	(0.205) **	
Urbanisation, rural municipalities is control group						
Organic	Capital area			2.64	(0.438) ***	$\chi_2^2(39.84) = 0.000$
	Urban municipality			0.72	(0.368) *	
Free-range	Capital area			1.21	(0.292)***	$\chi_2^2(17.42) = 0.000$
	Urban municipality			0.64	(0.270) **	
Barn	Capital area			0.09	(0.240)	$\chi_2^2(0.35) = 0.839$
	Urban municipality			-0.04	(0.209)	
Highest level of education, no further education stated is control group						
Organic	Voc.-oriented high-school			0.51	(0.462)	
	Short further education			0.77	(0.530)	
	Medium further education			1.12	(0.493) **	$\chi_4^2(7.14) = 0.128$
	Long further education			1.69	(0.852) **	
Free-range	Voc.-oriented high-school			0.58	(0.287)**	
	Short further education			0.82	(0.357) **	
	Medium further education			0.80	(0.337) **	$\chi_4^2(7.92) = 0.095$
	Long further education			1.09	(0.692)	
Barn	Voc.-oriented high-school			0.37	(0.229)	
	Short further education			0.26	(0.276)	
	Medium further education			0.59	(0.271) **	
	Long further education			-0.03	(0.499)	$\chi_4^2(5.48) = 0.241$
Log-likelihood		-8,384.65		-8,305.71		
No. of households		844		844		
No. of observations		10,800		10,800		

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000 combined with answers to AKF/GfK questionnaire from 2002. Purchases made directly from farms excluded. Rationing is allowed. Number of antithetic Halton Draws is 7,500. '***' is significant at the 1% level, '**' at the 5% level and '*' at the 10% level. The LR tests show the results of comparing model 2 with a model excluding variables group by group.

One way of interpreting the results in Table 6 is to look at the predicted share of the population with positive willingness to pay for organic eggs. The utility of money is 0.38 per DDK, and is thus practically unchanged compared to the model based only on prices and types. If a household believes in a positive effect on animal welfare (+0.99), has no trust in organic effect on food safety (+0.00, control group), has a high income (+0.71), has a member who is 60 years old or more (control group), lives in an urban municipality (+0.72) and has a long further education (+1.69), then the distribution of the utility of organic eggs has a mean of 0.11 (= -4.00 + 0.99 + 0 + 0.71 + 0 + 0.72 + 1.69) and a spread ($\tilde{\sigma}_{22}$, see equation (2.6)) of $17.16^{\frac{1}{2}} = 4.14$, which means that the willingness to pay has the mean $-\beta^2/\beta^p$ = -0.11/(-0.38) = 0.29 and standard deviation $-\tilde{\sigma}_{22}/\beta^p$ = -4.14/(-0.38) = 10.89 (see equation (2.2)). The model therefore predicts that half of the group are willing to pay at least DKK 0.29 more for an organic egg than for a battery egg, and that $51\% \left(= P\left(x > 0 \middle| x \sim N\left(0.29, 10.89\right)\right)\right)$ of the households in the group have positive willingness to pay for organic eggs compared to battery eggs.

It is important to understand that the parameters for types cannot be compared directly between the two models (and not only because of the change in scale mentioned in the theory section). When socio-demographic factors and perception of organic eggs are included, it means that the estimated means no longer relate to the entire sample, but only to the control group. The mean will therefore change to fit the mean of the control group. The utility of organic eggs is allowed to be influenced not only by socio-demographic factors, but also by perception of organic eggs. The result is that the mean utility drops from -1.72 to -4.00. The estimations show that age is the only socio–demographic factor which influences the utility of barn eggs significantly, and the utility of barn eggs only drops from -0.80 to -0.84.

Introducing socio-demographic factors reduces the estimated variation a little because some of the variation is now captured in the socio-demographics, but the effect is not dramatic. The correlations remain practically the same as in the simple model.

The utility of organic eggs increases significantly when the household trusts that organic production has positive effects on either animal welfare or food safety. The response to the two effects is of approximately the same magnitude. However, the question of food safety is not related directly to eggs, so the effect might be underestimated. This means that purchases of organic eggs are not solely driven by private motives (health), but also by altruistic motives (animal welfare).

Figure 3 illustrates the difference in willingness to pay for households who believe that there is *no difference* between the animal welfare of the hens used in production of organic and battery eggs (the grey line, mean -4.00, see Table 6) and the households who believe in a positive effect (the dark line, mean -4.00 + 0.99 = -3.01, see Table 6). This illustrates the effect of different means, given same standard deviations of willingness to pay. When the standard deviation is the same for the two groups, the group with the highest mean always has the highest willingness to pay, although the difference decreases with the price. Figure 2 and Figure 3 illustrate the importance of knowing the standard deviations when comparing means of a mixed logit.

Figure 3 Households with different perceptions of animal welfare related to organic eggs

Density of normally distributed WTP **Share with WTP higher than x DKK**

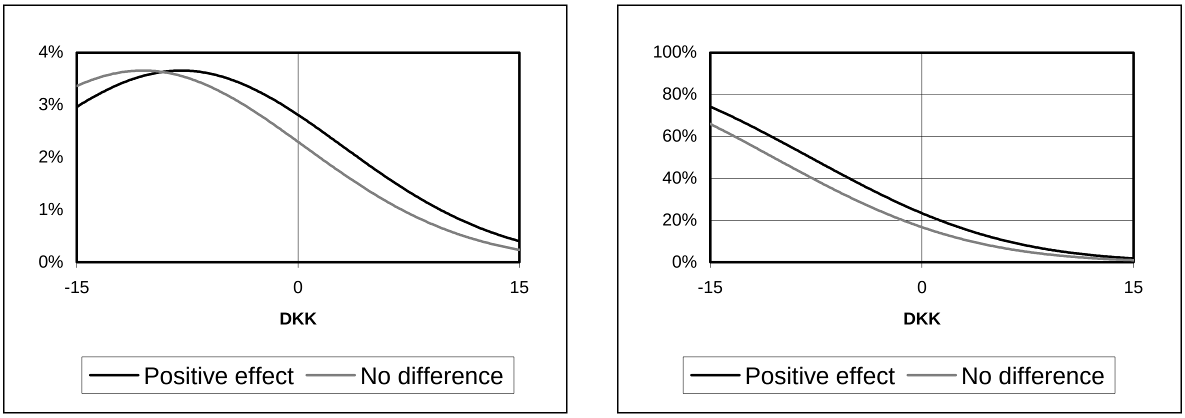


Illustration of the effect of different means, given same standard deviations of willingness to pay (WTP): The willingness to pay for the grey type has the lowest mean, but the same standard deviation as the dark type. This means that the share of the population with willingness to pay higher than a given amount is always bigger for the dark type, although the difference decreases.

The LR tests in Table 6 show the results of comparing the model with perception and socio-demographics with a model excluding variables group by group. The difference between the log-likelihood of model 2 and a model without perception of organic animal welfare is 15.24/2, which means that the LR test rejects that animal welfare can be excluded. At the other end of the scale, the test for the effect of income on barn eggs (probability 0.712) shows that income has no significant effect on barn eggs.

The effect of socio-demographics is very similar for organic and free-range eggs. Income is barely significant, and neither is age. However, urbanisation and to some degree education has a positive effect on the utility of these types of eggs. The picture is somewhat different for barn eggs, where only age seems to make a difference.

2.7. Conclusion

Expressed concern for animal welfare is not just cheap talk. A significant share of the population is willing to put money on the counter in order to increase animal welfare, even when we control for the private attribute food safety.

Willingness to pay for free-range and organic eggs is higher in urbanised municipalities and for households with relatively high incomes. Higher levels of education also influence the willingness to pay positively. The willingness to pay for barn eggs is mainly influenced by age; the older the household, the greater the willingness to pay.

As expected, the willingness to pay for organic eggs displayed more heterogeneity than was the case for barn or free-range eggs (multiple sources of value, e.g. familiar label and health), and the willingness to pay for organic eggs was generally higher than for barn eggs. Contrary to expectation, the willingness to pay was lowest for free-range eggs. However, this result has been seen in at least one other study using completely different methods. A plausible explanation could be that people either confuse barn eggs with free-range and prefer the cheaper barn eggs, or realise that free-range eggs are close to organic both in attributes and price and therefore prefer organic eggs, which yield both a familiar label and perhaps also an expectation of better health.

References

- ADAS/IGER/University of Bristol (2001). Livestock Knowledge Transfer, Fact sheet 'Poultry: 2001:512' (http://www.kt.iger.bbsrc.ac.uk/FACT%20sheet%20PDF%20files/kt32.pdf).
- Andersen, L.M. (2006). Consumer Evaluation of Environmental and Animal Welfare Labelling. *Working paper from AKF, Danish Institute of Governmental Research.* Available at www.akf.dk.
- Baltzer, K. (2002). Efterspørgslen efter fødevarekvalitet og -sikkerhed: Et pilotstudie af danske forbrugeres efterspørgsel efter æg. *Working paper from Danish Research Institute of Food Economics (FØI).* Available at www.foi.dk.
- Bennet, R.M. (1997). Farm animal welfare and food policy. *Food policy*, **22** (4): 281-88.
- Eurobarometer (2005). Attitude of consumers towards the welfare of farmed animals. *Special Eurobarometer 229.*
- Hanemann, W.M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics*, **66** (3): 332-41.
- IP/05/698, press release: Improving animal welfare: EU Action Plan adopted. Brussels, 23 January 2006.
- McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, vol. **15**: 447-70.
- Revelt, D. and K. Train (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics*, **80**: 1-11.
- Rolfe, John (1999). Ethical rules and the demand for free range eggs. *Economic Analysis & Policy*, **29** (2): 187-206.
- Teisl, Mario F., B. Roe and R.L. Hicks (2002). Can Eco-Labels Tune a Market? Evidence from Dolphin-Safe Labeling. *Journal of Environmental Economics and Management*, **43** (3): 339-59.
- The Danish Poultry Council ('Dansk Fjerkræråd'). Available at www.danskfjerkrae.dk (in Danish).
- Train, K. (1998). Recreation demand models with taste differences over people. *Land Economics*, **74** (2): 230-39.
- Train, K. (1999). Mixed logit models for recreation demand, in *Valuing recreation and the environment: Revealed preference methods on theory and practice,* edited by J.A. Herriges and C.L. Kling. Edward Elgar, Cheltenham, UK.

Comparable Likelihood Values – Antithetic Halton Draws in Mixed Multinomial Logit*

Laura Mørch Andersen†

June 2008

Abstract

Mixed logit models are typically estimated using Quasi-Monte Carlo integration and recent developments in sampling algorithms have increased precision substantially. However, characteristics known to apply to the likelihood function are only ensured in the limit – i.e as the number of draws goes to infinity. This paper shows that there are substantial gains to be reaped by a priory imposing the symmetry characteristic that applies to likelihood function dimensions describing standard deviations of mixed parameters. This is not done in the standard estimation procedures used today and this paper shows that this may result in substantial estimation- and inference errors within the span of draws typically applied. A similar type of problem occurs if the relationship between primes and mixed parameters is not maintained when testing mixing of parameters, again something which is typically not done with standard estimation procedures.

Key words: Antithethic Halton draws, panel mixed multinomial logit, MMNL, MXL, Quasi-Monte Carlo integration, Likelihood Ratio tests, simulated likelihood

3.1. Introduction

With improved computing capacity and simulation techniques the Mixed multinomial logit (MMNL, McFadden and Train, 2000) is becoming an attractive way for researchers to introduce heterogeneity into discrete models. Clearly, reliable estimation, validation and inference techniques are a prerequisite for sound models and analysis. At the core of MMNL

---

* I thank Kenneth Train, David Revelt and Paul Ruud for making their MMNL software available at Train's MMNL homepage: elsa.berkeley.edu/software/abstracts/train0296.html, and for allowing me to use a later version, which allows for correlation between mixed parameters.

The research was funded by The Danish Social Science Research Council ('FSE')

† Contact information: Laura Mørch Andersen, AKF - Danish Institute of Governmental Research, www.akf.dk
e-mail: LMA@akf.dk

estimation and inference is of course the likelihood function. MMNL models are estimated by simulated maximum likelihood, and restrictions easily tested with likelihood ratios.

The use of simulated likelihoods are of course bound to induce some approximation error, and it is therefore important to validate the results, e.g. by varying the starting values and checking the stability of the results. After working with mixed logit models for a while, one (almost inevitably) realizes that simulated likelihoods are not always trustworthy. When working with real data the estimated log-likelihood values are often highly dependent on the starting values, even when the estimated parameters are very similar. In some cases the problem is small enough to be ignored, but in other cases it rules out usable likelihood ratio tests, e.g. when likelihood ratio test statistics of restrictions become negative. This has been observed on real data, but is theoretically impossible.

The purpose of this paper is to investigate the problem systematically. Using simulated data we are able to understand how approximation error causes the observed types of simulated likelihood function instability, its implications for estimation and inference and to propose solutions that dramatically reduce the implications of approximation error for estimation and inference. We show that the same mechanisms appear in a real data set with invalidating implications for LR-inference tests.

The result is, that the asymmetric halton draws used in most standard estimation procedures gives rise to two types inference problems: 1) The asymmetry of the draws can generate substantial variation in likelihood values between quadrants in dimensions describing standard deviations of mixed parameters that by definition should be identical. 2) Another type of variation of the likelihood value is generated if the relationship between primes and mixed parameters is not maintained when testing mixing of parameters.

Erroneous variation in the simulated values of the log-likelihood function may first of all lead to falsely accepted or rejected hypotheses. Secondly it may also falsely indicate that data is not informative enough to support the model, and therefore lead to unnecessary reductions in model complexity. The problem is of course reduced when approximation error is reduced by increasing the number of Halton draws. However, the problem remains critical within the span of draws that is feasible today. This paper shows that the problem is completely removed when one uses antithetic instead of conventional Halton draws and maintains the relationship between primes and mixed parameters when testing mixing of parameters.

The structure of the paper is: Section 3.1 presented the problems associated with likelihood ratio tests performed on simulated log-likelihood values. Section 3.2 describes the mixed logit model for panels, section 3.3 explains how the likelihood function of mixed logit can be simulated by Quasi-Monte Carlo integration and section 3.4 briefly presents the concept of likelihood ratio tests. Section 3.5 investigates the lack of symmetry of the simulated log-likelihood function on artificial data, comparing results in the optimum and using conventional Halton draws and section 3.6 investigates the effect of the lack of symmetry outside the optimum. Section 3.7 illustrates the problem using real data and section 3.8 introduces antithetic Halton draws. Section 3.9 presents the results of using this type of draws. Section 3.10 discusses the problems related to restrictions on the number of mixed parameters and section 3.11 concludes.[1]

3.2. Panel Mixed logit

Discrete choice models are based on the assumption that individuals derive different levels of utility from different alternatives. The utility is assumed to depend on the characteristics of the alternative, and perhaps also of the individual making the choice. The utility is assumed to vary randomly from choice to choice. The decision-maker knows the exact utility in each period, but the econometrician can only estimate the non-random part of the utility function. The utility function for individual *i* at time *t* can therefore be written as

$$U_{it}\left(j\middle|\beta_i, x_{ijt}\right) = x_{ijt}\beta_i + \varepsilon_{ijt} \tag{3.1}$$

where β_i is an $m \times 1$ vector of parameters giving the utility of the variables in the $1 \times m$ vector of attributes x_{ijt} for individual *i*. If the error terms are assumed to follow the extreme value distribution, it is possible to estimate the parameters of the utility function using a conventional multinomial logit model (McFadden 1973) which means that the likelihood function is:

[1] Appendix A explains how to draw from a distribution, Appendix B presents the artificial data used in this paper and Appendix C defines the eight quadrants of $\mathbb{R}^3$. Appendix D illustrates the differences in log-likelihood values in optimum in different quadrants. Appendix E compares the log-likelihood functions of different quadrants outside the optimum and Appendix F illustrates the degree of symmetry of conventional Haltons by picturing the skewness coefficients for different primes and different numbers of draws.

$$L_i^{conv}\left(y_i,\beta_i,x_i\right)=p\left(y_i\middle|\beta_i,x_i\right)=\prod_{t=1}^{T_i}p\left(y_{it}\middle|\beta_i,x_{iy_{it}t}\right)=\prod_{t=1}^{T_i}\left(\frac{\exp\left(U_{it}\left(y_{it}\middle|\beta_i,x_{iy_{it}t}\right)\right)}{\sum_{k=1}^{J}\exp\left(U_{it}\left(k\middle|\beta_i,x_{ikt}\right)\right)}\right) \tag{3.2}$$

where J is the number of alternatives available, T_i is the number of choices made by individual i, y_i is a $T_i\times 1$ vector of choices actually made by individual i, β_i is a $m\times 1$ vector containing all of the parameters, x_i is a $T_i\times m\cdot J$ matrix of attributes of the alternatives and $U_{it}\left(k\middle|\beta_i,x_{ikt}\right)$ is the utility for household i from choosing element k from the choice set in period t given the parameters β_i and the attributes x_{ijt}. In a conventional logit it is assumed that all individuals have the same utility function, i.e. that β is the same for all individuals.

In a Mixed MultiNomial Logit (MMNL or MXL) model[2] (McFadden and Train 2000), it is assumed that (part of) the individual utility is drawn from a distribution (i.e. the individual utility is known to the individual, but only the distribution is observable to the econometrician). The individual likelihood functions then become the likelihood function of the conventional multinomial logit model integrated over all possible values of β:

$$L(\theta)=\prod_{i=1}^{I}L_i(\theta)=\prod_{i=1}^{I}\left(\int p\left(y_i\middle|\beta_i,x_i\right)f\left(\beta_i\middle|\theta\right)d\beta_i\right) \tag{3.3}$$

where θ are the parameters determining the distribution of the parameter β, and $f\left(\beta\middle|\theta\right)$ is the density of β given θ. The likelihood function is maximised over θ instead of β.

3.3. Quasi-Monte Carlo integration

Calculating the likelihood function in (3.3) is very cumbersome, especially if θ describes a multivariate distribution. The problem can be reduced significantly by using Monte Carlo integration (see e.g. Morokoff and Caflisch (1995) for asymptotic properties). Monte Carlo integration generally means drawing a set of values of β from the distribution given by θ, calculating the value of the integral for each draw, and taking the mean of these values.[3] Quasi-Monte Carlo integration means that the values of β are drawn quasi randomly from the distribution, instead of randomly. Halton draws are one type of quasi random draws.[4]

---

[2] Also known as Random Parameter Logit (RPL).

[3] Appendix A explains how to draw from any given distribution, and illustrates the difference between random and Halton draws.

[4] Halton draws are drawn from a Halton sequence. Halton sequences were first presented by Hammersley (1960) and Halton (1960). Halton ascribes the idea to Hammersley and talks about Hammersley sequences, but the name Halton sequences seem to have stuck. The efficiency of Halton sequences is discussed in detail in both

In a panel mixed logit the β's are assumed to be drawn from a common distribution, but to be constant for each individual. The β's therefore vary over individuals, not over observations from the same individual. For each of the *I* individuals one must therefore draw *R* sets of β's. For each individual, the value of the likelihood function of the conventional likelihood is then calculated for each of the randomly drawn β's and the mean is an approximation of the integral in equation (3.3). This means that the likelihood function for the entire sample is:

$$\begin{aligned} L(y,\theta,x) &= \prod_{i=1}^{I} L_i(y_i,\theta,x_i) \\ &= \prod_{i=1}^{I}\left(\int\left(p(y_i|\beta_i,x_i)\, f(\beta_i|\theta)\right)d\beta_i\right) \\ &\approx \prod_{i=1}^{I}\left(\frac{1}{R}\sum_{r=1}^{R}\left(p(y_i|\beta_{ir},x_i)\right)\right) \end{aligned} \tag{3.4}$$

where the $R \cdot I$ β_{ir}'s are drawn from the mixing distribution. According to Hensher and Greene (2003), there is no standard for the number of draws needed, but they find that 100 draws appears to be a "good" number. In order to validate the model Hensher and Greene suggest that the models are estimated over a range of draws from 25 to 2,000.

3.4. Likelihood ratio tests on simulated log-likelihood values

In many cases the purpose of estimating a likelihood function is twofold: Maximising the likelihood function leads to the set of parameters which fit the data best, and comparing the best likelihood values of different models makes it possible to determine whether the models are significantly different. The latter is done by Likelihood Ratio (LR) tests. Twice the difference in logs of the likelihood values of the unrestricted (L_U) and the restricted (L_R) model is chi-square distributed with degrees of freedom equal to the number of restrictions imposed (Greene 1997), as described in equation (3.5):

$$LR = -2\left(\ln(L_R) - \ln(L_U)\right) \sim \chi^2_{df} \tag{3.5}$$

If the difference between the restricted and the unrestricted likelihood is very small, the LR test will be unable to reject that the two models have the same explanatory power. The critical

---

Train (1999) and Baht (2001). Both find that Halton sequences greatly improve accuracy with far fewer draws and faster computation.

difference in log-likelihood value depends on the number of restricted parameters (degrees of freedom) and the level of statistical significance. Table 1 shows the critical difference between log-likelihood values, given the degrees of freedom and significance level. Note that the numbers should be multiplied by two to give the value of the likelihood ratio test statistic in (3.5).

Table 1 Critical differences in log-likelihood values (ln(L_U)-ln(L_R))

	log-likelihoods significantly different at:		
Degrees of freedom	**1% level**	**5% level**	**10% level**
1	3.32	1.92	1.35
2	4.61	3.00	2.30
3	5.67	3.91	3.13
4	6.64	4.74	3.89
5	7.54	5.54	4.62
6	8.41	6.30	5.32

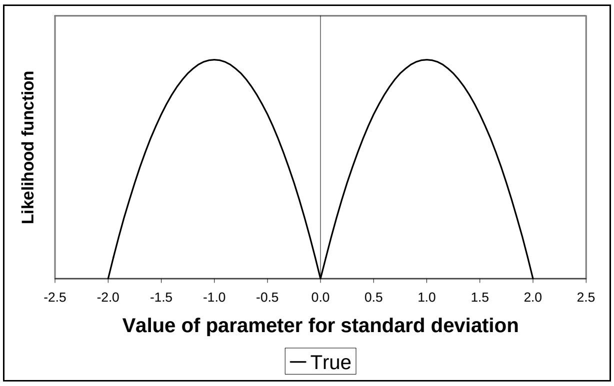
If the difference between the restricted and the unrestricted log-likelihood function is 3.32 – and only one parameter is restricted – the values of the LR test becomes $\chi_1^2(2 \cdot 3.32) = \chi_1^2(6.63) = 1\%$ which means that the probability that the models have the same explanatory power is one per cent, which usually leads to the conclusion that the two models are significantly different. If the difference is only 1.35 the probability of equal explanatory power is ten per cent, since $\chi_1^2(2 \cdot 1.35) = \chi_1^2(2.71) = 10\%$.

From Table 1 it is clear that even rather small variations in the log-likelihood value can have a significant impact on the results. The absolute level of the log-likelihood function is of no interest, but if the standard deviation of the simulated log-likelihood is above e.g. one, testing hypotheses may easily lead to false conclusions. Sometimes the LR value becomes too small, other times too big, and in very unfortunate cases it may have the wrong sign if the log-likelihood value of the restricted model becomes higher than the log-likelihood value of the unrestricted model. The varying values of the log-likelihood function may first of all lead to falsely accepted or rejected hypotheses. Secondly, it may also falsely indicate that data are not informative enough to support the model, and therefore lead to unnecessary reductions in model complexity.

3.5. Symmetry of simulated log-likelihood, artificial data

If a logit model is mixed with e.g. the normal distribution, each mixed parameter leads to two mixing parameters, a mean and a standard deviation. Both of these parameters are maximised over the entire real axis $\mathbb{R}$. The distribution actually depends on the mean and the variance, and since the standard deviation is the square of the variance there is no mathematical problem in a negative standard deviation. As illustrated in Figure 1, the true likelihood function will be symmetric around zero, when focusing on one dimension relating to a parameter for a standard deviation.

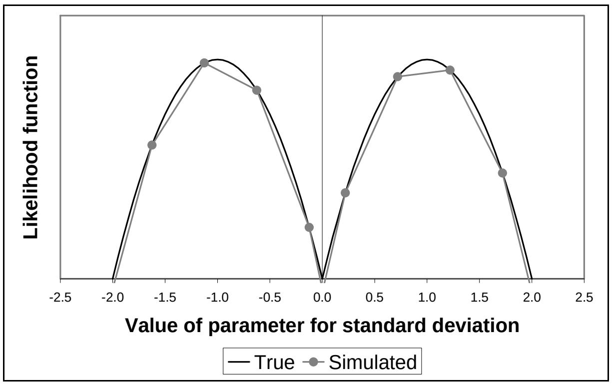
Figure 1 True symmetric likelihood function, one dimension



Hypothetical log-likelihood function on a one dimensional parameter space.

When simulating a likelihood function by Monte Carlo integration the likelihood value is calculated in a number of points (draws). If these draws are not symmetric around zero the result is not likely to be symmetric either. In a mixed multinomial logit model the likelihood function is an integral over the likelihood function of a conventional logit model (see equation (3.4)), which means that the likelihood function depends on the *area* below the simulated likelihood function presented in Figure 2. As illustrated in Figure 2, the lack of symmetry may lead to some variation in the optimal parameter, and especially to different absolute values of the likelihood optimum.

Figure 2 Lack of symmetry of simulated likelihood function, one dimension



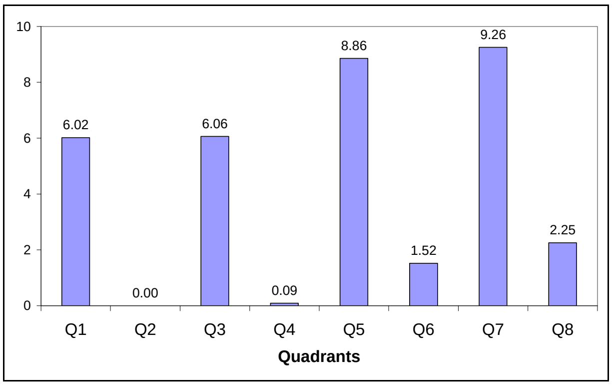
True and simulated hypothetical log-likelihood function on a one dimensional parameter space.

In order to investigate the magnitude of the problem under controlled conditions, an artificial data set has been created. The data are panel data with 1,000 individuals each choosing 20 times between 4 alternatives. The utility of the alternative specific constant is zero for the alternative which is used as base (the base alternative has no alternative specific constant in the estimation). The utility of the three other alternatives follows a three-dimensional normal distribution with no correlation. For more on the definition of the artificial data, se Appendix

B. One of the virtues of artificial data is that the true parameters are known.[5] In the following, there will be no optimisation on the artificial data. Instead the likelihood values calculated from the true parameters will be compared for different quadrants. This illustrates the magnitude of the problems that may potentially arise from actual estimations, but it does not tell much about the probability of encountering these problems. The probability of ending up in a given quadrant may well vary in actual estimations, and the calculations without optimisation treats all quadrants as equal.[6]

In the case of three mixings, the parameters are estimated in $\mathbb{R}^3$, which means that the number of different quadrants is $2^3 = 8$, and the likelihood function must therefore be symmetric in all eight quadrants. Figure 3 shows the difference between the likelihood values of the optimal parameters calculated in different quadrants (see definition of quadrants in Appendix C) using 100 conventional Halton draws. Note that 100 is the number of draws per *individual*, i.e. in a model with 1,000 individuals the total number of draws is 100,000. The difference between the log-likelihood values in Q2 and Q7 is 9.26 which is definitely not zero.

Figure 3 Log likelihood in optimum by quadrant, 100 conventional Halton draws



The lowest estimated log-likelihood value (-17,265.14 from quadrant 2) is normalised to zero. Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws.

Table 2 compares the results of the eight quadrants for increasing numbers of draws. The approximation error decreases as the number of draws increases, simply because the distance

[5] When simulating data, the empirical mean and standard deviation of the mixed parameters differ slightly from the parameters used in the simulation of the data (when the number of observations is finite). The 'true parameters' are the actual parameters which are the result of the data simulation. See Appendix B for the difference.

[6] As will be illustrated later, estimations on actual data lead to results in all eight quadrants, so the problem also exists when parameters are optimised rather than known a priory.

between draws is reduced, but it does not disappear within a feasible span of draws, and certainly not for the low number of draws recommended in Hensher and Greene (2003).

As mentioned above, one of the problems caused by the difference between the values of the log-likelihood function evaluated at different quadrants is that it influences the results of likelihood ratio tests. Table 2 also reports the results of calculations of the likelihood values for a restricted model where the mean of the utility of one of the non-base alternatives is restricted to zero. Again the results vary from quadrant to quadrant.

Table 2 Variation in simulated log-likelihood, by number of draws, artificial data[7]

	Number of draws per individual					
	100	500	1,000	1,500	5,000	7,500
Unrestricted model:						
Highest absolute difference in simulated log-likelih.	9.26	4.44	1.09	0.88	0.47	0.19
Standard deviation of simulated log-likelihood	3.771	1.657	0.345	0.352	0.180	0.073
Restricted model:						
Highest absolute difference in simulated log-likelih.	10.02	5.10	1.41	1.02	0.48	0.22
Standard deviation of simulated log-likelihood	3.949	1.884	0.474	0.384	0.187	0.074

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws. The restricted model assumes that the mean utility of alternative B is zero.

The utility of alternative B has a mean of 0.0834 and a standard deviation of 0.9934 (Appendix B). The restricted model assumes that the mean is zero, and places no bounds on the standard deviation. Testing away the mean is therefore *not* the same as testing whether the utility of alternative B is the same as the utility of the base alternative A. Comparing the choices made in the artificial data set from appendix B with the choices made in an identical data set except that the mean is restricted to zero shows that 99.16 per cent of the 20,000 choices are identical in the two data sets.[8] The restricted model should therefore be accepted.

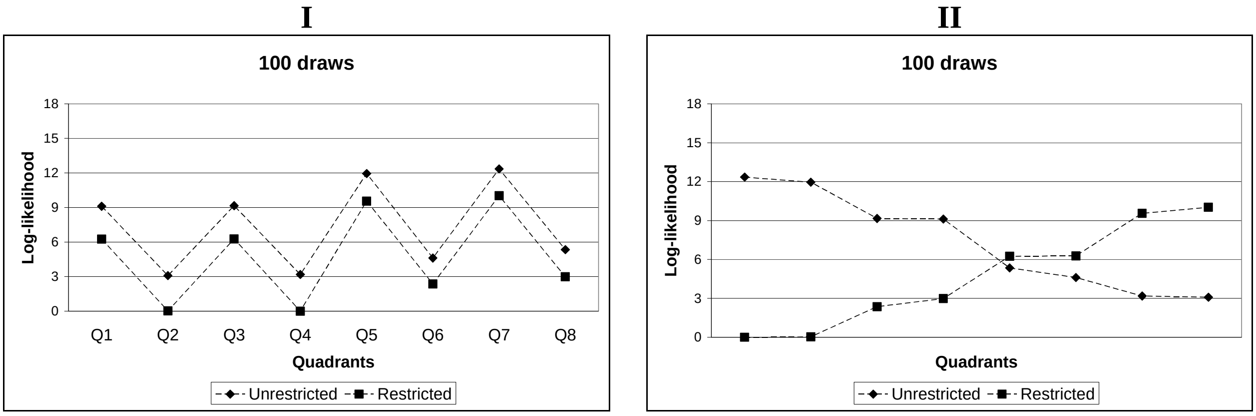
The large variation in the value of the log-likelihood function means that the value of the restricted model in one quadrant may be higher than the value of the unrestricted model in another quadrant, but never within quadrants. Figure 4 shows the log-likelihood values for the unrestricted and the restricted model using 100 Halton draws. **I** shows the relationship between the two models in each quadrant and **II** ignores the quadrants and sorts the values by

[7] Note that the standard deviation of the likelihood values is based on the eight results from the different quadrant, and might change if the estimations were optimised which would allow the probability of the quadrants to vary.

[8] The random extreme value distributed noise added to the utilities is the same in the two datasets, only the mean utility varies.

size. Especially from **II** it is evident that the value of the restricted model will sometimes be higher than the value of the unrestricted model, leading to negative values of the LR test statistic.

Figure 4 Likelihood values of unrestricted and restricted model, 100 conventional Halton draws



Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws. The restricted model assumes that the mean utility of alternative B is zero.

Table 3 shows that for 100 conventional Halton draws, the LR test statistic will become negative in 20 of the 64 different combinations of restricted and unrestricted log-likelihood values, corresponding to 31 per cent of the cases. The problem decreases with the number of draws, but is still present at 1,000 draws. Appendix D illustrates the log-likelihood values for all the different numbers of draws presented in Table 3.

Table 3 Testing away one mean using conventional Halton draws

	Number of draws per individual:					
	100	**500**	**1,000**	**1,500**	**5,000**	**7,500**
Share of negative LR values	.31	.31	.05	.00	.00	.00
Results of positive LR values						
Lowest test probability	.00	.00	.03	.05	.08	.12
Highest test probability	.64	.66	.77	.65	.27	.20
Standard deviation of test prob.	.13	.16	.15	.13	.05	.02

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws. The restricted model assumes that the mean utility of alternative B is zero.

Table 3 also summarises the results of the likelihood ratio tests that can be performed on the positive LR tests statistics. The test probabilities vary from zero to 64 per cent for 100 draws leading to a standard deviation of 13 per cent. This is of course deeply problematic. The problem decreases as the number of draws increases but even for 5,000 draws the restricted model will sometimes be accepted at the 10 per cent level, and other times rejected. Table 4 shows the probability of rejecting the restricted model at different significance levels.

Table 4 Probability of rejecting the restricted model

	Number of draws per individual:					
	100	500	1,000	1,500	5,000	7,500
At the 1 per cent level	.55	.25	.00	.00	.00	.00
At the 5 per cent level	.89	.45	.07	.05	.00	.00
At the 10 per cent level	.91	.59	.28	.30	.17	.00

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws. The restricted model assumes that the mean utility of alternative B is zero.

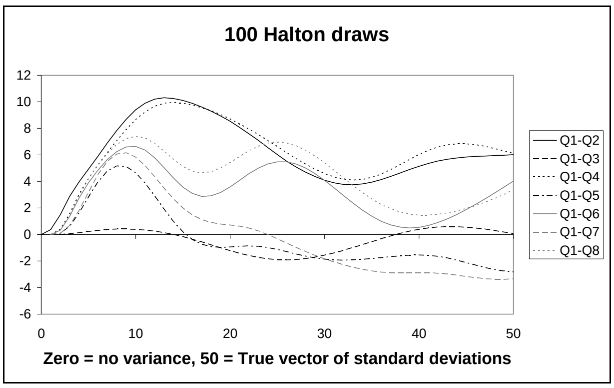
Using 100 draws, 55 per cent of the 64 combinations of unrestricted and restricted log-likelihood values rejects the restricted model at the 1 per cent significance level, indicating that in 55 per cent of the cases the probability that the restricted model has the same explanatory power as the unrestricted model is less than one per cent. In 91 per cent of the cases the restricted model is rejected at the 10 per cent level. Using 7,500 draws, the model is never rejected.

3.6. Comparing likelihood values outside the optimum

In the discussion of symmetry above, focus has been on differences in optimum, i.e. in the true parameters of the artificial data. During the search for optimum the optimisation routine has to perform outside the optimum, and it is therefore important to know what happens there.

Figure 5 illustrates the difference between the values of the log-likelihood function in the eight different quadrants when the parameters for the standard deviations vary from zero to the true value (obtained in the point 50). The difference between the highest and the lowest value in optimum is the 9.26 presented in Table 2. Appendix E presents the differences for higher numbers of draws.

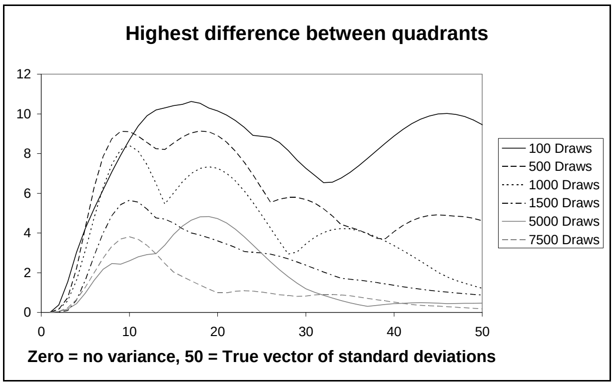
Figure 5 Comparing the likelihood values of different quadrants outside the optimum



Artificial data (see Appendix B). Difference between the value of the log-likelihood function evaluated in different quadrants. The set of standard deviations varies linearly from a vector of zeros to a vector of true standard deviations. The mean is always the true mean and the log-likelihood function is therefore evaluated in the points: True mean, s*(true std)/50, s=0, 1, …,50 on the positive quadrant (Q1) and true mean, s*(-(true std))/50, s=0, 1, …,50 on the negative quadrant (Q8, see Appendix C).

The difference decreases as the vector of standard deviations approaches the true value (the point 50 in Figure 5), but it does not disappear. The difference is likely to arise because the Halton draws are asymmetric. If the support for the Monte Carlo integration is different on each side of the real axis, the simulated integral will also be different. This problem ought to decrease with the number of draws, and as illustrated in Figure 6 this is to some extent true, but even with 7,500 draws the difference still does not disappear, and the problem is sometimes smaller for 5,000 draws than for 7,500 draws.

Figure 6 Comparing the likelihood values, increasing the number of draws



Artificial data (see Appendix B). Difference between the value of the log-likelihood function evaluated in different quadrants. The set of standard deviations varies linearly from a vector of zeros to a vector of true standard deviations. The mean is always the true mean and the log-likelihood function is therefore evaluated in the points: True mean, s*(true std)/50, s=0, 1, …,50 on the positive quadrant (Q1) and true mean, s*(-(true std))/50, s=0, 1, …,50 on the negative quadrant (Q8, see Appendix C).

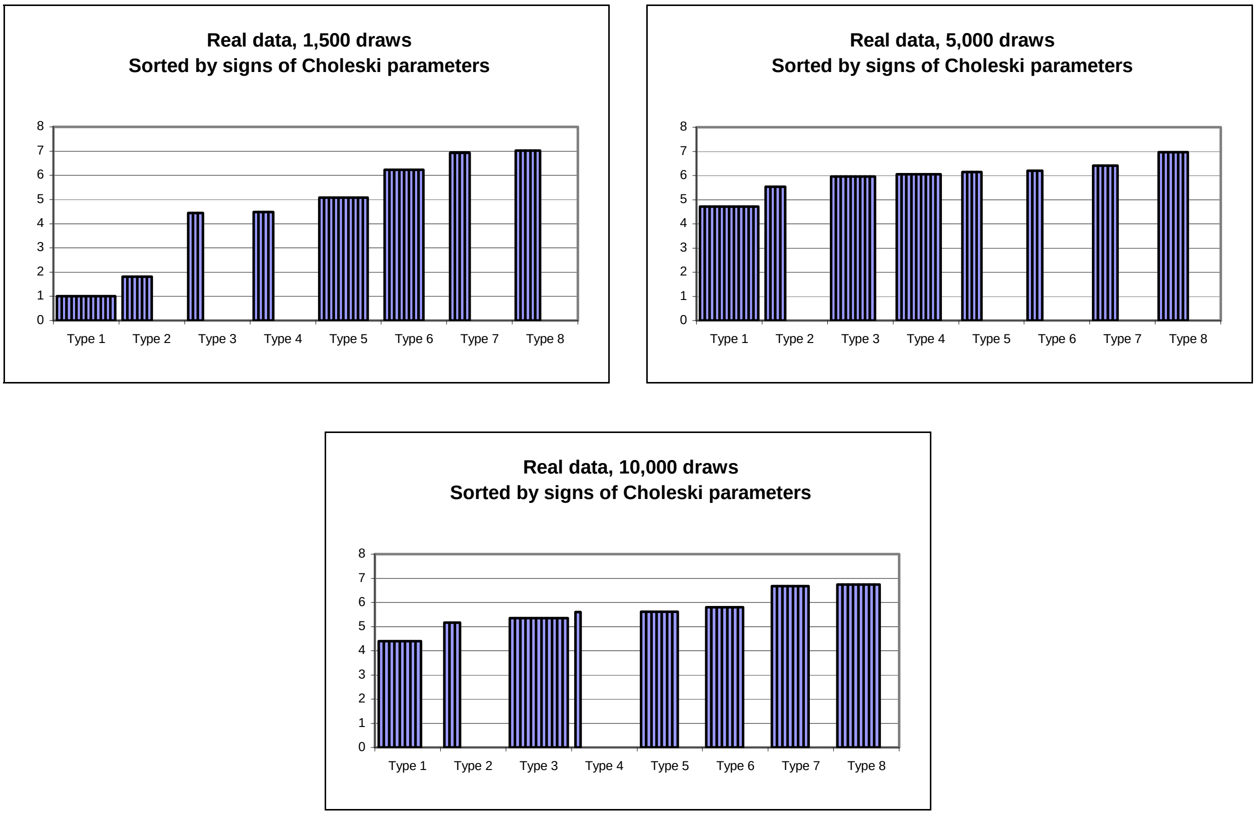
This has to do with the symmetry of the Halton draws. As illustrated in Appendix F, the degree of symmetry increases as the number of draws increases, but not monotonically. Increasing the number of draws by a few thousand may therefore lead to set of draws with a lower degree of symmetry.

3.7. An example using real data

The problems described above have also been experienced on real data. The example below is based on 10,971 observations from 848 individuals, choosing between four different alternatives. The utility of the non-base alternatives is assumed to follow a tree-dimensional normal distribution with correlation. In this example the true parameter values are not known, and the model is therefore optimised using different sets of starting values.

Figure 7 shows the actual distribution of the optimised simulated log-likelihood values for a model with four alternatives and three mixed alternative specific constants (on real data). It is clear that increasing the number of draws reduces the scale of the problem, but it does not solve the problem. The maximal log-likelihood values of 52 different sets of starting values have been sorted by the sign of the Choleski parameters.[9] As can be seen in Figure 7, the estimated log-likelihood values differ significantly between the eight quadrants, but not within quadrants. The number of starting values varies between quadrants, so the figure cannot tell whether the optimised result is more likely to be found in one quadrant than in another. Comparing the quadrants of the starting values and of the optimised results shows that there is apparently no connection between the starting point and the final result.

Figure 7 Absolute levels of maximum simulated log-likelihood values by quadrant



The lowest estimated log-likelihood value (-8390,10, from the estimation with 1,500 draws) is normalised to one. Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants with correlation. 52 sets of starting values. Real data, unbalanced panel, 848 individuals, 10,971 observations. Estimations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws.

It is also clear from Figure 7 that the level of the maximum simulated log-likelihood generally is higher for estimations with 5,000 draws than for estimations with 1,500 draws, and that the difference between the quadrants is smaller for 5,000 draws than for 1,500 draws. The problem thus decreases with the number of draws, but even with 10,000 draws (which is in most cases too time consuming) the problem is still there. The problems described using

[9] The Choleski factorisation (Q) is a triangular matrix with the property $QQ' = \Omega$, where Ω is the covariance matrix (Train 2003). The Choleski matrix is therefore the 'square root' of the covariance matrix, and if the covariance matrix is diagonal (i.e. no correlations) the Choleski matrix is merely a diagonal matrix of standard deviations.

artificial data therefore also appear on real data. In this example, the variation in likelihood values is sufficiently large to disturb LR tests, even for high numbers of draws. It is important to note that whereas the variation is large (and highly problematic) for the values of the log-likelihood function, the estimated parameters do not vary to the same extent (more on this in Table 7 on page 101).

3.8. Antithetic Haltons

If the model includes more than one mixing parameter, symmetry in one dimension is not enough. If the number of mixed parameters is n – and if perfect symmetry is the goal – for each point in a given quadrant a corresponding point must be present in all of the other 2^n-1 quadrants. Even with a two dimensional mixing based on 2 and 3, this is never the case. The problem is solved by creating antithetic Halton draws. As in Train (2003), the draws are created so that each point is “mirrored” into the 2^n-1 other dimensions.

For a case with three mixed parameters a Halton draw $d_1 = \begin{bmatrix} d_1^1 & d_1^2 & d_1^3 \end{bmatrix}$ (between zero and one) is drawn, and then paired with 7 mirrors in the following way:

$$\begin{bmatrix} d_{11} \\ d_{12} \\ d_{13} \\ d_{14} \\ d_{15} \\ d_{16} \\ d_{17} \\ d_{18} \end{bmatrix} = \begin{bmatrix} d_1^1 & d_1^2 & d_1^3 \\ 1-d_1^1 & d_1^2 & d_1^3 \\ d_1^1 & 1-d_1^2 & d_1^3 \\ d_1^1 & d_1^2 & 1-d_1^3 \\ 1-d_1^1 & 1-d_1^2 & d_1^3 \\ 1-d_1^1 & d_1^2 & 1-d_1^3 \\ d_1^1 & 1-d_1^2 & 1-d_1^3 \\ 1-d_1^1 & 1-d_1^2 & 1-d_1^3 \end{bmatrix} \tag{3.6}$$

The Haltons must be symmetric for each individual in the panel, and it is therefore important that each ‘set’ of symmetric draws is assigned to one individual only, and not distributed over different individuals. The number of draws per individual in a model with n-dimensional mixing must therefore be a multiple of 2^n. In the case of 1,500 draws and three mixings this means that the number of draws must be e.g. $63 \cdot 2^3 = 1{,}504$ instead of 1,500 to ensure symmetry. Antithetic draws always have perfect symmetry, and therefore always skewness coefficient equal to zero. See Appendix D for more on skewness.

3.9. Results of antithetic Halton draws

When the simulated likelihood function for the artificial data is calculated using antithetic Halton draws the difference between the log-likelihood over positive and negative standard deviations is always zero as desired and the likelihood ratio test of the restricted model therefore no longer varies. However, the result still changes as the number of draws increases. Table 5 presents the test probabilities already presented in Table 3 above, combined with the results of the antithetic draws.

Table 5 Testing away one mean using conventional or antithetic Halton draws

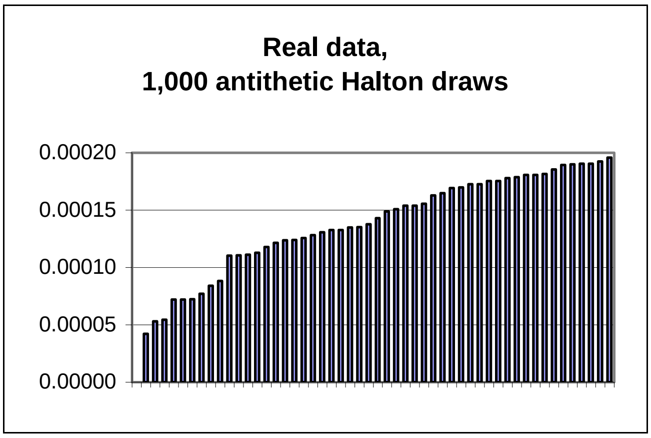
	Number of draws per individual:					
	100	**500**	**1,000**	**1,500**	**5,000**	**7,500**
Conventional Halton draws						
Lowest test probability	.00	.00	.03	.05	.08	.12
Highest test probability	.64	.66	.77	.65	.27	.20
Antithetic Halton draws:						
Test probability	.01	.08	.11	.09	.15	.15

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard or antithetic Halton draws. The restricted model assumes that the mean utility of alternative B is zero.

The results of estimations using antithetic Halton draws for the real data are presented in Figure 8. The results should be compared with the ones in Figure 7, now with all results within the same quadrant. The precision of the optimisation is set to 10^{-4}, and the highest difference between two results is now lower than twice this level, and thereby completely acceptable.[10] Differences of this magnitude will have absolutely no effect on likelihood ratio tests, and the antithetic Halton draws therefore solve the problem of instability in the simulated likelihood of the mixed logit. At least the part caused by lack of symmetry of the likelihood function.

[10] The precision of the optimisation indicates how close to zero the gradient of the log-likelihood function must be to be perceived as a maximum.

Figure 8 Results of antithetic draws



Zero is the lowest estimated value. Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants with correlation. 52 sets of starting values. Real data, unbalanced panel, 848 individuals, 10,971 observations. Estimations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using 1,000 antithetic Halton draws.

Table 6 summarises the results presented in Figure 7 and Figure 8, and clearly demonstrates the effect of the antithetic Halton draws.

Table 6 Standard deviation of simulated log-likelihood, by number of draws, real data

	Number of draws per individual			
	1,000	**1,500**	**5,000**	**10,000**
Conventional Haltons				
Highest absolute difference in simulated log-likelihood	10.88	6.02	2.26	2.35
Standard deviation of simulated log-likelihood	2.90	2.29	0.71	0.78
Antithetic Haltons				
Highest absolute difference in simulated log-likelihood	0.000196			
Standard deviation of simulated log-likelihood	0.000273			

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants with correlation. 52 sets of starting values. Real data, unbalanced panel, 848 individuals, 10,971 observations. Estimations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using either standard Halton draws or antithetic Halton draws.

In this example, the variation in likelihood values is sufficiently large to disturb LR tests, even for high numbers of draws. It is important to note that whereas the variation is large (and highly problematic) for the values of the log-likelihood function, the estimated parameters do not vary to the same extent. Table 7 shows the standard deviation of the 52 estimated parameters for each type and number of draws mentioned in Table 6. The estimated parameters are not truly identical, but the variance is far from the variance of the log-likelihood values. Again, the antithetic draws reduce the variance dramatically.

Table 7 Standard deviation of parameter estimates, by number of draws

	Number of draws per individual			
	1,000	1,500	5,000	10,000
Conventional Haltons				
Highest standard deviation of estimated parameters[a]	0.078	0.022	0.035	0.021
Antithetic Haltons				
Highest standard deviation of estimated parameters	0.000612			

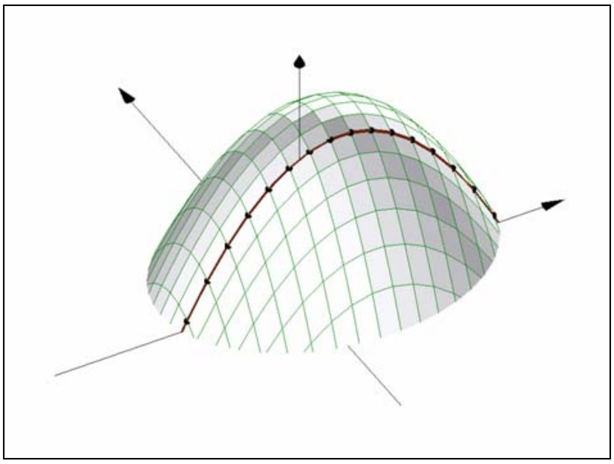
Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants with correlation. 52 sets of starting values. Real data, unbalanced panel, 848 individuals, 10,971 observations. Estimations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using either standard Halton draws or antithetic Halton draws.

a: Parameters of standard deviations are evaluated at their absolute value.

3.10. Testing away mixing dimensions

Even when the problem of symmetry is solved, the problem of comparing log-likelihood values of models with different dimensions still remains. In a model with two mixed parameters (β_1 and β_2) the Halton draws will be based on two primes, e.g. 2 and 3 (2 representing β_1 and 3 representing β_2). If one of the mixed parameters (e.g. β_1) is restricted to be fixed (standard deviation restricted to zero), the dimension of the log-likelihood function is decreased by one, and the Halton draws will be based on only one prime. The standard choice would be the first prime, i.e. 2, independent of which dimension is restricted. Figure 9 illustrates the simulated conventional likelihood function which is to be integrated to form the likelihood function of the mixed logit. The heavy black line shows the likelihood function when one of the parameters is restricted to zero, and the dots on this line show the points in which the one-dimensional log-likelihood function would be evaluated for the given grid.

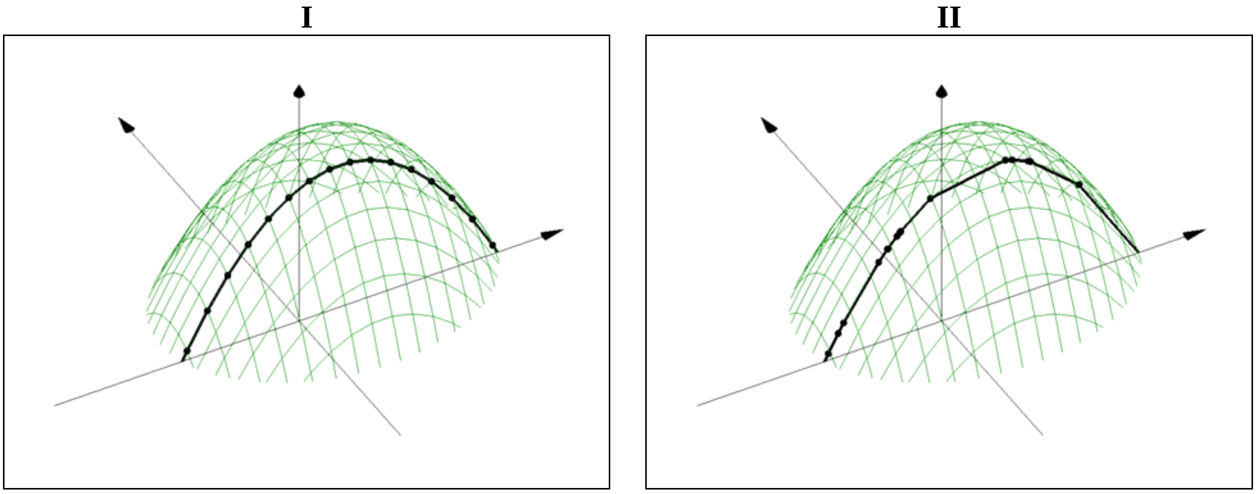
Figure 9 Simulated log-likelihood function in one and two dimensions



The figure describes a hypothetical log-likelihood function on a two dimensional parameter space. The heavy black line shows the likelihood function when one of the parameters is restricted to zero.

The symmetry of antithetic Haltons is needed to ensure that the log-likelihood functions of the different quadrants are identical, but as illustrated in Figure 10, the choice of prime may also matter. Figure 10 describes the same hypothetical log-likelihood function as Figure 9. The dots show the points in which the one-dimensional log-likelihood function would be evaluated for different draws. The dots in **I** illustrate a case where the one-dimensional draws correspond with the two-dimensional grid, and **II** illustrates a case where the one-dimensional draws are not part of the two-dimensional grid. The area under the one-dimensional likelihood function is clearly not the same in **I** and **II**.

Figure 10 Different one-dimensional likelihood functions given by different draws



The figure describes the same hypothetical log-likelihood function as in Figure 9. **I** illustrates a case where the one-dimensional draws correspond with the two-dimensional grid, **II** illustrates a case where the one-dimensional draws are not part of the two-dimensional grid.

To investigate the size of the problem we return to the artificial data used above (1,000 individuals and 20 observations per individual, Defined in Appendix B). The restriction is now placed on the utility of alternative C instead of the utility of alternative B, which was restricted in the mean-restriction case above. The utility of alternative C has a mean of 0.9981 and a standard deviation of 0.0984. The restricted model assumes that the standard deviation is zero, but places no bounds on the mean. This means that the restricted model does not assume that the utility of alternative C is the same as the utility of the base alternative A. Table 8 shows the results of evaluating the log-likelihood function in the true parameters of the restricted model, using different primes for the antithetic Haltons. The difference is substantial.

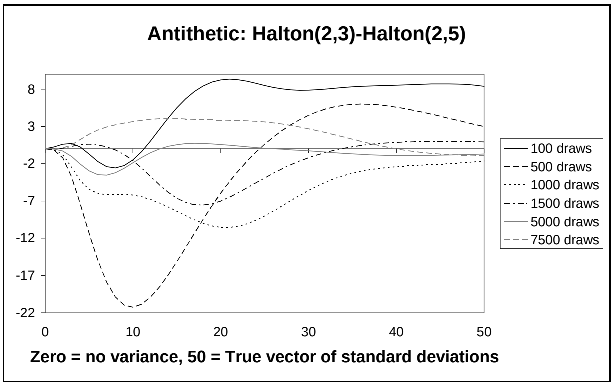
Table 8 Log-likelihood values in the optimum of the restricted model using antithetic Haltons

	Number of draws per individual:					
	100	500	1,000	1,500	5,000	7,500
Antith. Haltons based on 2 and 3	-18,466	-18,413	-18,412	-18,410	-18,409	-18,408
Antith. Haltons based on 2 and 5	-18,475	-18,415	-18,410	-18,411	-18,409	-18,408
Difference	8.37	2.98	-1.64	0.92	-0.70	-0.85

Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws. The restricted model assumes that the standard deviation of the utility of alternative C is zero.

Again, it is also important to know how the likelihood functions behave outside the optimum, and Figure 11 illustrates what happens when the log-likelihood function is simulated with different sets of primes.

Figure 11 Difference in simulated log-likelihood on artificial data outside the optimum, using different primes for antithetic draws



Artificial data (see Appendix B). The standard deviation of alternative C is always set to 0 and the standard deviation of the the alternatives 2 and 4 varies linearly from a vector of zeros to the true standard deviations. The mean is always the true mean and the log-likelihood function is therefore evaluated in the points: True mean, s*(true std)/50, s=0, 1, …,50. This is done using the primes 2 and 3 which would be a standard solution, and 2 and 5 corresponding to the dimensions of the unrestricted log-likelihood, since alternative C corresponds to the prime 3 in the unrestricted model. Antithetic draws, i.e. the quadrant is not important.

The problem obviously has the same magnitude as the symmetry problem alone for models without restrictions on the mixing dimensions (see Figure 6). However, the effect of restricting the standard deviation of a mixed parameter to be zero is very different from the effect of restricting the mean of a mixed parameter to be zero. The true value of the restricted mean in Table 5 is 0.0834 (see Appendix B) and the true value of the restricted standard deviation in Table 8 is 0.0984, i.e. the absolute values of the parameters are almost identical. Yet the restricted mean-model is accepted as the number of draws increases, but the restricted variance-model leads to differences in the log-likelihood value above 1,000 even for 7,500 draws, so it is unequivocally rejected. The difference between the log-likelihoods based on different primes in Table 8 and Figure 11 is therefore of no importance. But for ‘close’ LR tests it will be important to keep track of the relationships between primes and mixing dimensions.

3.11. Conclusion

In mixed logit it is assumed that the signs of parameters for the standard deviation of the mixing distribution have no influence on the value of the likelihood function. When the Monte Carlo integration of the likelihood function is done by standard Halton draws, this assumption breaks down because the Halton draws differ from quadrant to quadrant. This means that sets of optimal parameters with different signs of the standard deviaton can lead to a number of different values of the log-likelihood function, even though the estimated variance is the same. If the solution to an unrestricted and a restricted model is found in different quadrants the likelihood ratio test will make no sense. This paper demonstrates that using antithetic Halton draws eliminates this problem. The paper also illustrates that when testing restrictions on the number of mixed parameters, the relationship between primes and mixed parameters must be maintained in the restricted model.

Note, however, that local maxima may still occur if the model cannot be empirically identified by the data. The stability of the simulated log-likelihood should therefore always be investigated by estimating with different sets of starting values.

References:

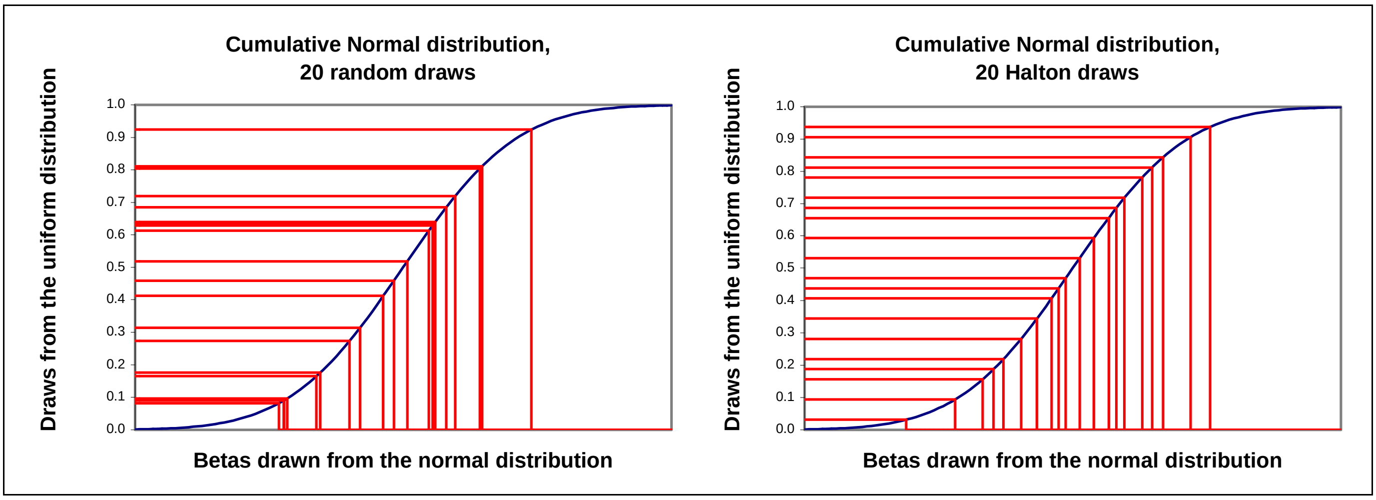
- Baht, C. R. (2001): Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, **35**: 677-693
- Greene, W.H. (1997): *Econometric analysis*. Third edition. Prentice Hall, Upper Saddle River, New Jersey, USA.
- Halton, J.H. (1960): On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, **2**: 84-90.
- Hammersley, J.M. (1960): Monte Carlo methods for solving multivariable problems. *Annals of the New York Academy of Sciences*, **86**: 844-874.
- Hensher, D. A. and W. H. Greene (2003): The Mixed Logit model: The state of practice, *Transportation,* **30**: 133–176.
- McFadden, D. (1973): Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (ed.): *Frontiers in econometrics*. Academic Press, New York. USA, 105-142.
- McFadden, D. and K. Train (2000): Mixed MNL models for discrete response. *Journal of Applied Econometrics*, **15**: 447-470.
- Morokoff, W.J. and R.E. Caflisch (1995): Quasi-Monte Carlo Integration. *Journal of Computational Physics*, **122**: 218-230.
- Train, K. (1999): Halton sequences for mixed logit. *Working paper*, Department of Economics, University of California, Berkeley.
- Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

Appendix A: Drawing from a distribution

Draws from any given distribution can be created by drawing from the uniform distribution (greater than zero and smaller than one) and taking the inverse of the desired cumulative distribution of these draws (Train 2003). The results will follow the desired distribution. The draws from the uniform distribution can be random draws, Halton draws or other types of quasi-random draws. The efficiency of Halton draws is discussed in detail in both Train (1999) and Baht (2001). Both find that Halton draws greatly improve accuracy with far fewer draws and faster computation. Halton draws are used in this paper, but the properties of the antithetic draws can be generalised to other types of draws.

Figure 12 illustrates how normally distributed draws can be created from random or Halton draws from the uniform distribution.

Figure 12 Drawing from a normal distribution, random and Halton draws



This principle is also formulated in equation (3.7):

$$\left.\begin{array}{l}\gamma_r \sim uniform\,]0,1[\\ \beta_r = F^{-1}(\gamma_r)\\ F = cdf \text{ of mixing distribution } f\end{array}\right\} \Rightarrow \quad \beta_r \sim f \qquad (3.7)$$

Appendix B: Artificial data

The number of individuals is 1,000 and each individual makes 20 choices. The number of alternatives is 4.

The artificial utility of the four alternatives is defined as:

$$mean\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.1 \\ 1 \\ 2 \end{bmatrix}, \quad std.dev\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0.1 \\ 3 \end{bmatrix}, \quad corr\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix} \tag{3.8}$$

which means that the artificial covariance is defined as:

$$\mathrm{cov}\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 \\ & & 0.01 & 0 \\ & & & 9 \end{bmatrix} \tag{3.9}$$

The realisation of the utility is

$$\begin{gathered} mean\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 \\ 0.0834 \\ 0.9981 \\ 2.0682 \end{bmatrix}, \quad std\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 \\ 0.9934 \\ 0.0984 \\ 3.0343 \end{bmatrix}, \\ \mathrm{cov}\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ & 0.9868 & 0.0021 & -0.0633 \\ & & 0.0097 & -0.0004 \\ & & & 9.2070 \end{bmatrix} \end{gathered} \tag{3.10}$$

Alternative number 1 is used as the base. The mean and the standard error of the base alternative is zero and since differences of normally distributed parameters are also normally distributed, the result is:

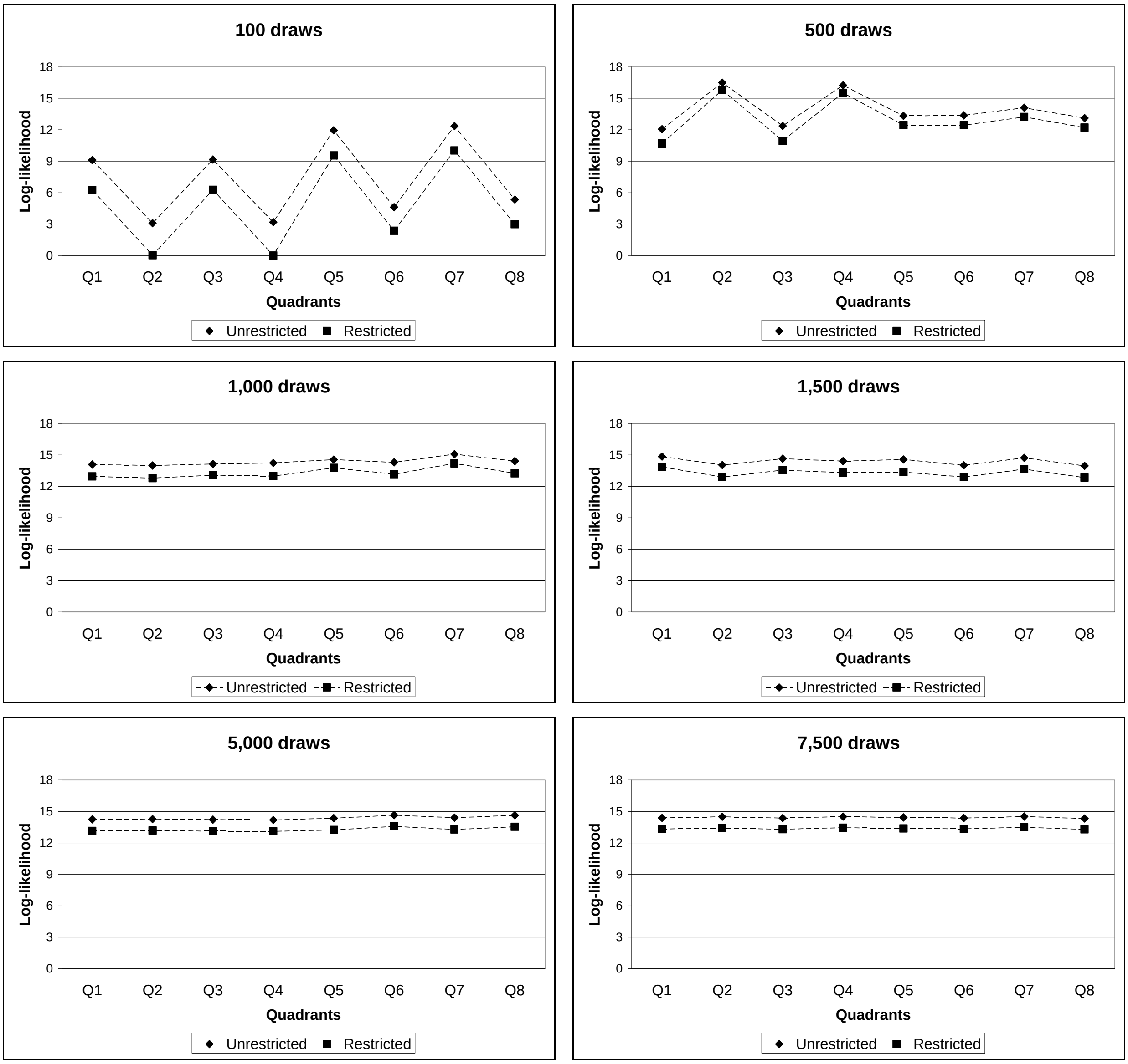
$$mean\begin{bmatrix} u_B - u_A \\ u_C - u_A \\ u_D - u_A \end{bmatrix} = \begin{bmatrix} 0.0834 \\ 0.9981 \\ 2.0682 \end{bmatrix}, \quad \mathrm{cov}\begin{bmatrix} u_B - u_A \\ u_C - u_A \\ u_D - u_A \end{bmatrix} = \begin{bmatrix} 0.9868 & 0.0021 & -0.0633 \\ & 0.0097 & -0.0004 \\ & & 9.2070 \end{bmatrix} \tag{3.11}$$

Appendix C: Definition of quadrants

$$\text{Quadrants}\left\{\begin{array}{c:rrr} & \multicolumn{3}{c}{\overbrace{\hphantom{B\quad C\quad D}}^{\substack{\text{Sign of standard}\\ \text{deviation of utility}\\ \text{of alternative:}}}} \\ & B & C & D \\ \hdashline Q_1 & 1 & 1 & 1 \\ Q_2 & 1 & 1 & -1 \\ Q_3 & 1 & -1 & 1 \\ Q_4 & 1 & -1 & -1 \\ Q_5 & -1 & 1 & 1 \\ Q_6 & -1 & 1 & -1 \\ Q_7 & -1 & -1 & 1 \\ Q_8 & -1 & -1 & -1 \end{array}\right. \tag{3.12}$$

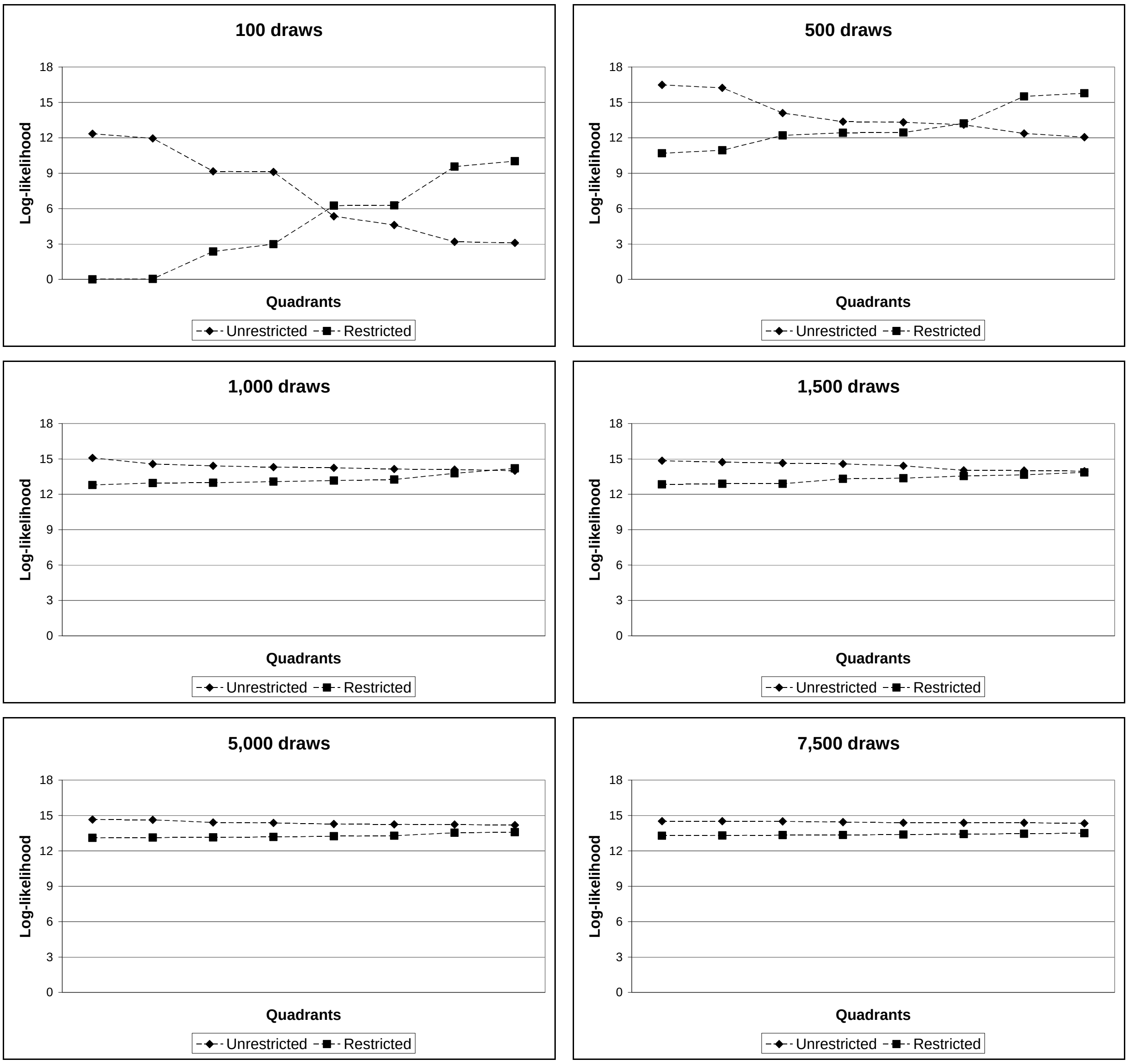
Appendix D: Illustration of differences in log-likelihood values in optimum from different quadrants

Table 9 Log-likelihood function evaluated in the true parameters, by quadrant



Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws. The restricted model assumes that the mean utility of alternative B is zero.

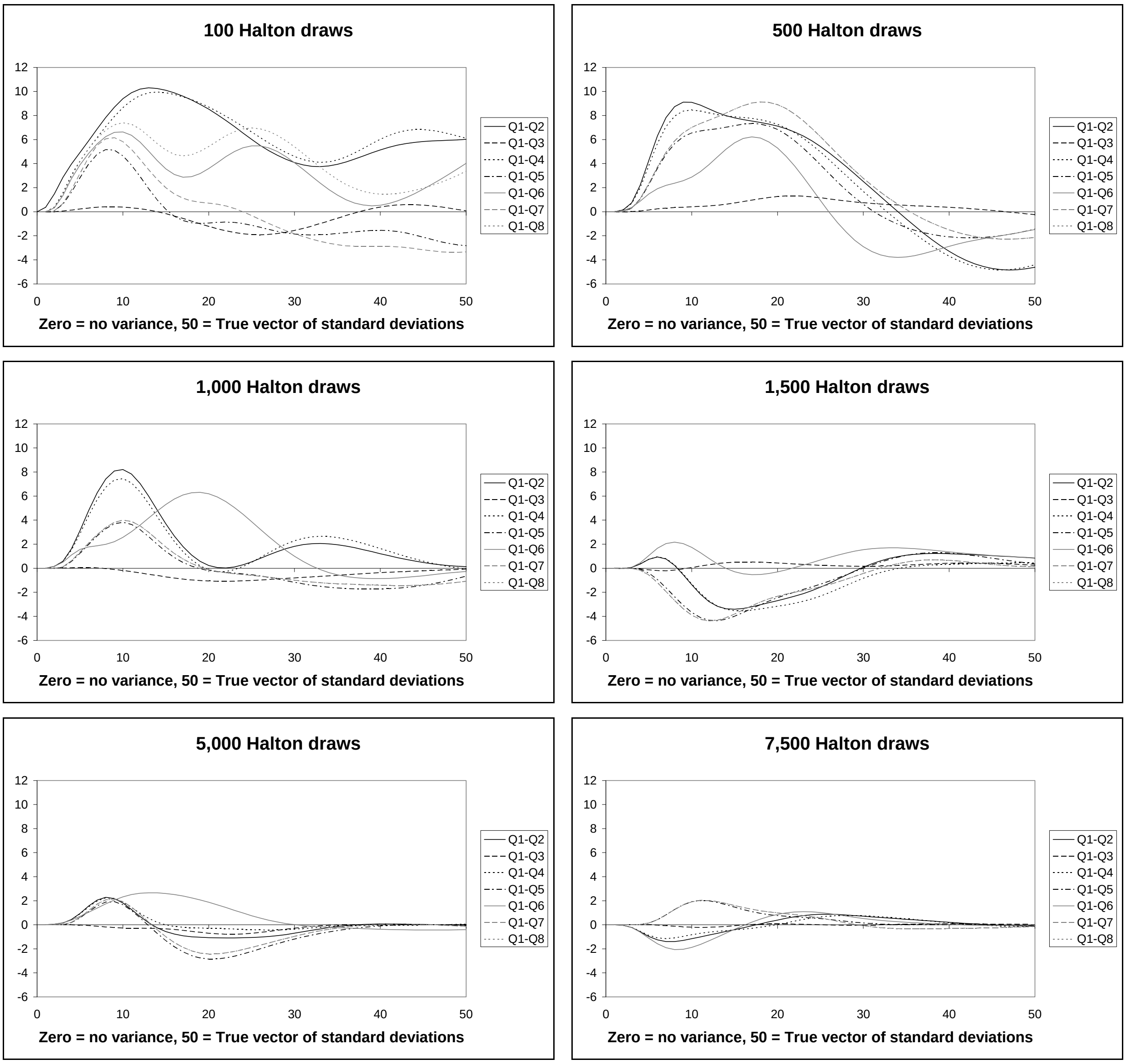
Table 10 Log-likelihood function evaluated in the true parameters, by size



Simulated log-likelihood in a model with 4 alternatives and 3 mixed alternative specific constants. Artificial data, 1,000 individuals, 20 observations per individual. Data is defined in Appendix B. Calculations conducted in the MMNL GAUSS programme developed by Train, Revelt and Ruud, using standard Halton draws. The restricted model assumes that the mean utility of alternative B is zero.

Appendix E Comparing log-likelihood functions for different quadrants outside the optimum

Figure 13 Comparing log-likelihood functions outside the optimum



Artificial data (see Appendix B). Difference between the value of the log-likelihood function evaluated in different quadrants. The set of standard deviations varies linearly from a vector of zeros to a vector of true standard deviations. The mean is always the true mean and the log-likelihood function is therefore evaluated in the points: True mean, s*(true std)/50, s=0, 1, …,50 on the positive quadrant (Q1) and true mean, s*(-(true std))/50, s=0, 1, …,50 on the negative quadrant (Q8, see Appendix C).

Appendix F: Skewness coefficient of random and Halton draws

The skewness coefficient is a measure of symmetry and can be calculated for any distribution. As in Greene (1997), the skewness coefficient is calculated as:

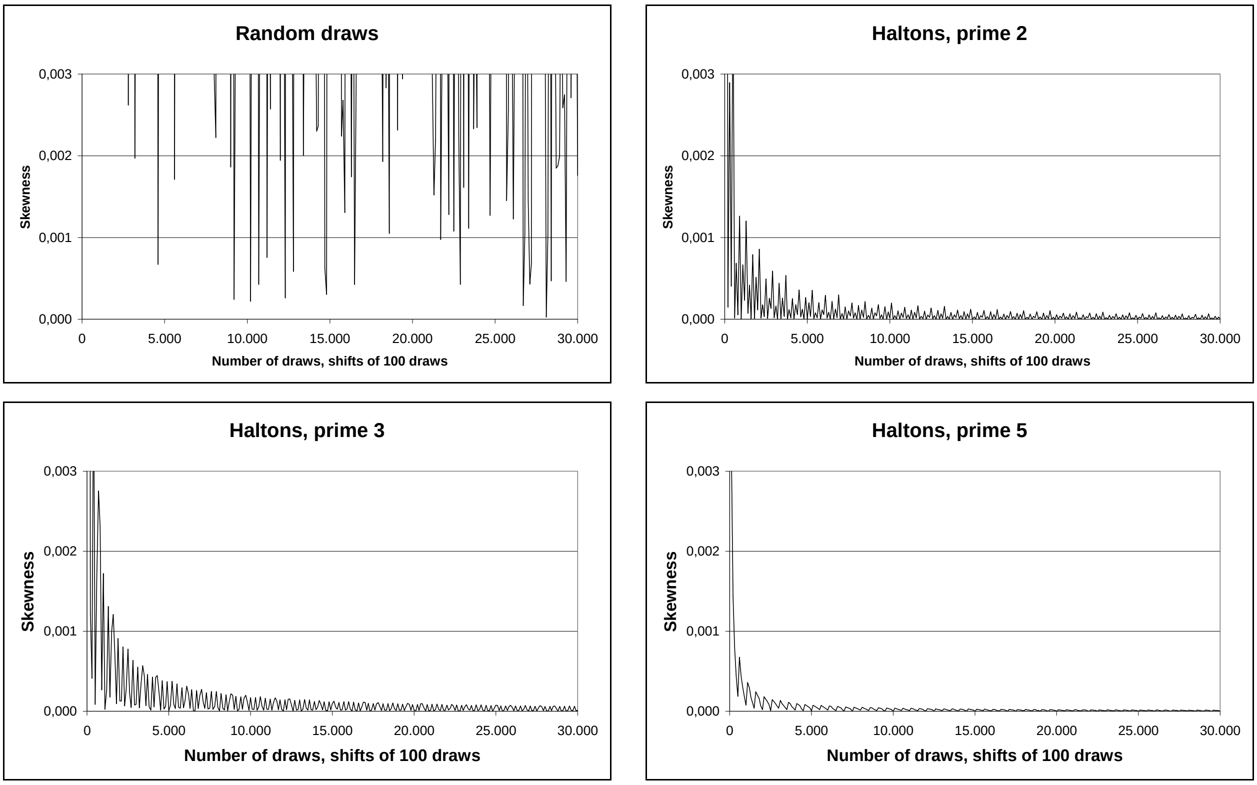
$$skewness\ coefficient = \frac{E\left(\left(x-E(x)\right)^3\right)}{\left(E\left(\left(x-E(x)\right)^2\right)\right)^{3/2}} \qquad (3.13)$$

The skewness coefficient is only measured in one dimension at a time, which means that the symmetry of the multivariate distribution may be far smaller than for each of the single dimensions.

Figure 14 shows the skewness coefficients of random draws and Haltons based on the primes 2, 3 and 5. First of all, it is clear that Halton draws are far more symmetric than random draws. Secondly, it is clear that the symmetry of the Halton draws increases with the number of draws, but that it keeps fluctuating, even for very high numbers of draws. Increasing the number of draws by a few thousand may therefore lead to set of draws with a lower degree of symmetry.

The antithethic Haltons presented in this paper always have skewness coefficient zero.

Figure 14 Skewness coefficients of random draws and Halton draws based on the primes 2, 3 and 5



A Censored Structural Characteristics Model for Milk*

Laura M. Andersen† and Sinne Smed‡

July 2008

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 either 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 households that drink a lot of milk prefer milk with a moderate to high fat share.

* The research was funded by the Danish Social Science Research Council (FSE).

† Contact information: Laura Mørch Andersen, AKF - Danish Institute of Governmental Research, www.akf.dk e-mail: LMA@akf.dk. ‡Sinne Smed, AKF - Danish Institute of Governmental Research, e-mail: SIS@akf.dk.

4.1. 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,[1] 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 to a massive campaign by 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, 1980). 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 if they vary non-systematically through an error term placed directly in the utility function.

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

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 4.2 starts out with the basic theory of the characteristics model and then the data and the milk markets are described in section 4.3. Section 4.4 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 4.5 summarises the results of the introductory model. In section 4.6 the model is reformulated according to the best suited model to allow estimation of a Tobit model with two-sided censoring. Finaly, section 4.7 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 4.8 is devoted to a discussion and conclusion.

4.2. 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.

$$\begin{array}{cc} & \overbrace{\begin{array}{ccccc} 1 & \cdots & j & \cdots & J \end{array}}^{\textit{Characteristics}} \\ \textit{Goods} \left\{ \begin{array}{c} 1 \\ \vdots \\ i \\ \vdots \\ I \end{array} \right. & \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1J} \\ \vdots & \ddots & \vdots & \unicode{x22F0} & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{iJ} \\ \vdots & \unicode{x22F0} & \vdots & \ddots & \vdots \\ a_{I1} & \cdots & a_{Ij} & \cdots & a_{IJ} \end{bmatrix} \equiv \mathrm{A} \end{array} \tag{4.1}$$

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 enter the market). For each household we observe the quantity purchased of each good: $q_t^h = \left(q_{1t}^h, \dots, q_{it}^h, \dots, q_{It}^h\right)'$ and we also observe a unit price for each good in each period: $p_t^h = \left(p_{1t}^h, \dots, p_{it}^h, \dots, p_{It}^h\right)'$. 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:

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

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):

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

where λ is the utility value of increasing the binding constraint (the budget) $\partial u/\partial x$. If we, furthermore, substitute the technology into the budget restriction – which we assume are binding – we get the following estimation problem:

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

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, $\left(\pi = \partial x/\partial z\right)$. 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:

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

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

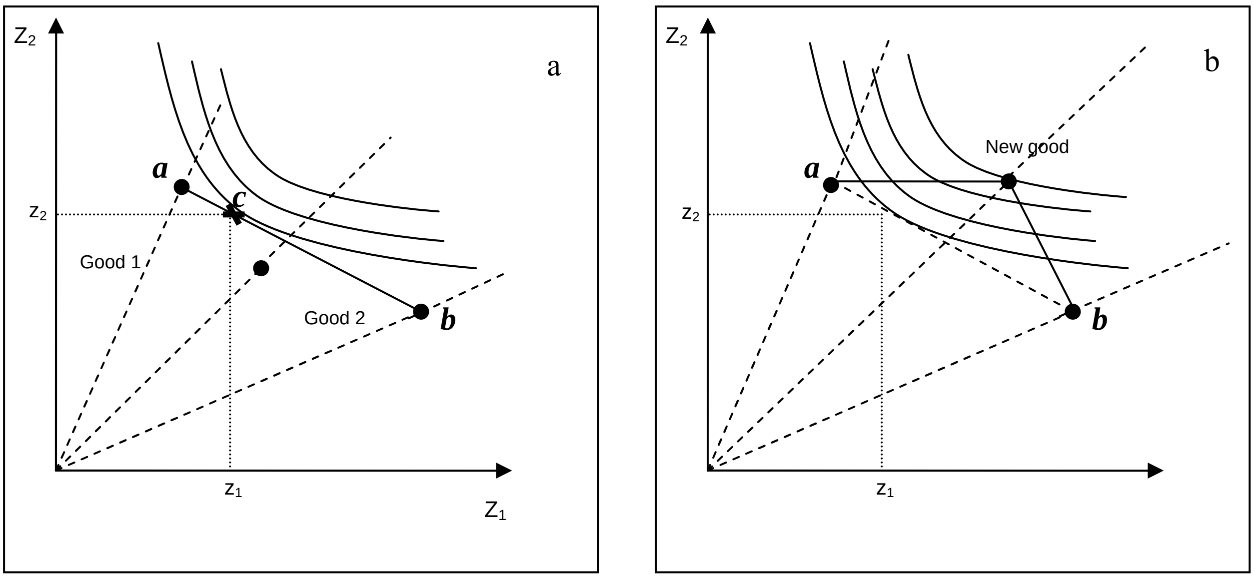
I.e. $\pi_1 = p_1\tilde{a}_{11} + p_2\tilde{a}_{12}$ and $\pi_2 = p_1\tilde{a}_{21} + p_2\tilde{a}_{22}$, where $\tilde{a}_{ij}$ are the elements in the inverse technology matrix. In this simple universe where the unit price is independent of the quantity, the implicit price of a characteristic is simply 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[2] x, we can calculate the amount of each characteristic $\left(z_1, z_2\right)$ that household h would obtain in period t spending all the money on good one (point a in Figure 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

[2] 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.

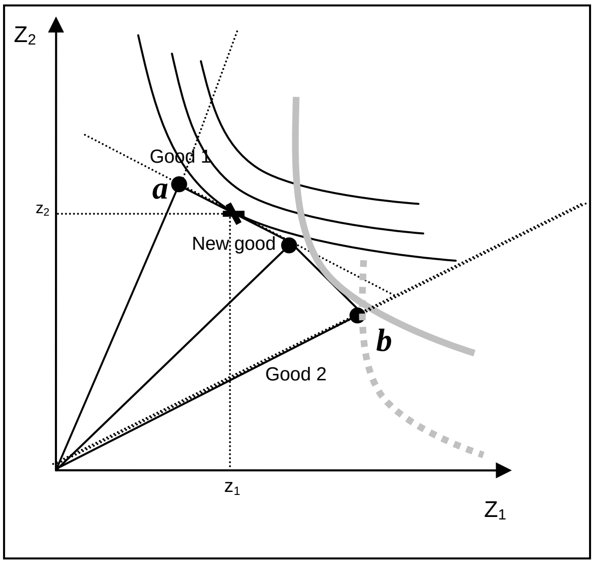
$(a, b, 0)$ due to the technology restriction. On the Danish milk market it is not possible to purchase milk with less than 1 gram or more than 35 grams of fat per litre. We assume that all goods can be purchased in continuous quantities and the line between the highest obtainable level of characteristics (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 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 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 1 Consumers' optimisation problem in a two characteristics world



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 2. It is not possible to buy goods outside the triangle consisting of zero and the lines running through *a* and *b* in Figure 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 dashed household in Figure 2. We will return to that later.

Figure 2 More consumers in a two goods, two characteristics world



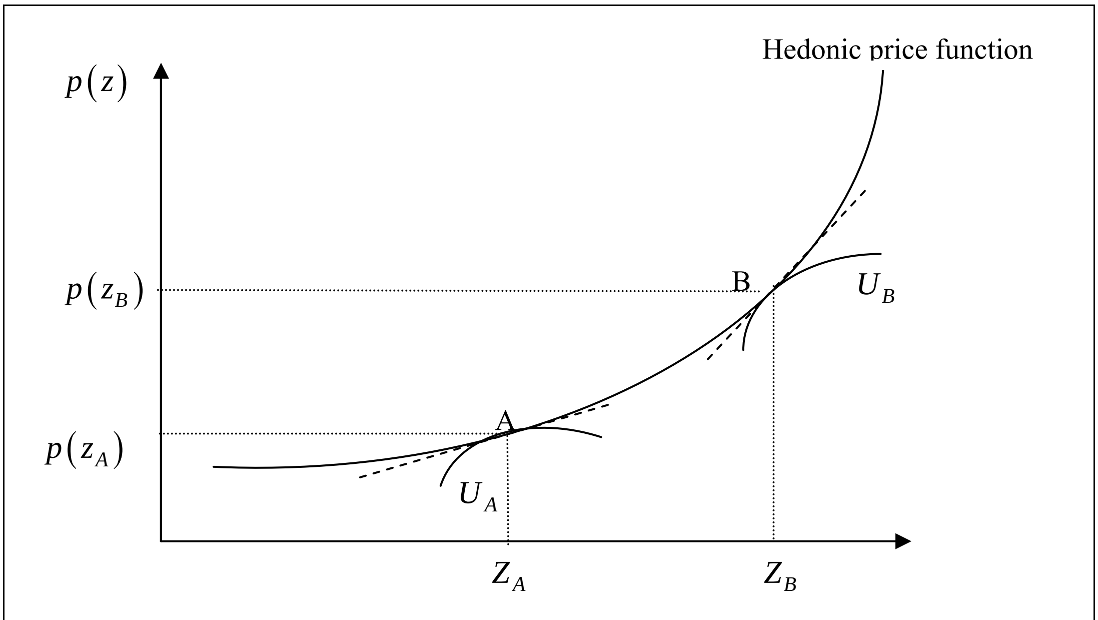
Estimation of implicit prices

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:

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

The derivatives, $\partial z_j / \partial q_i$ are the elements in the technology matrix a_{ij}. The marginal utility of the budget $\lambda = \partial u / \partial x$ 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 $\left(\partial u / \partial z_j\right) \lambda^{-1} = \left(\partial u / \partial z_j\right) / \left(\partial u / \partial x\right)$ is equal to the marginal rate of substitution between expenditure and the characteristics $\left(\partial x / \partial z_j\right)$, this is equivalent to the marginal implicit price π_j of each of the characteristics. 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 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 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), Eppel (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.

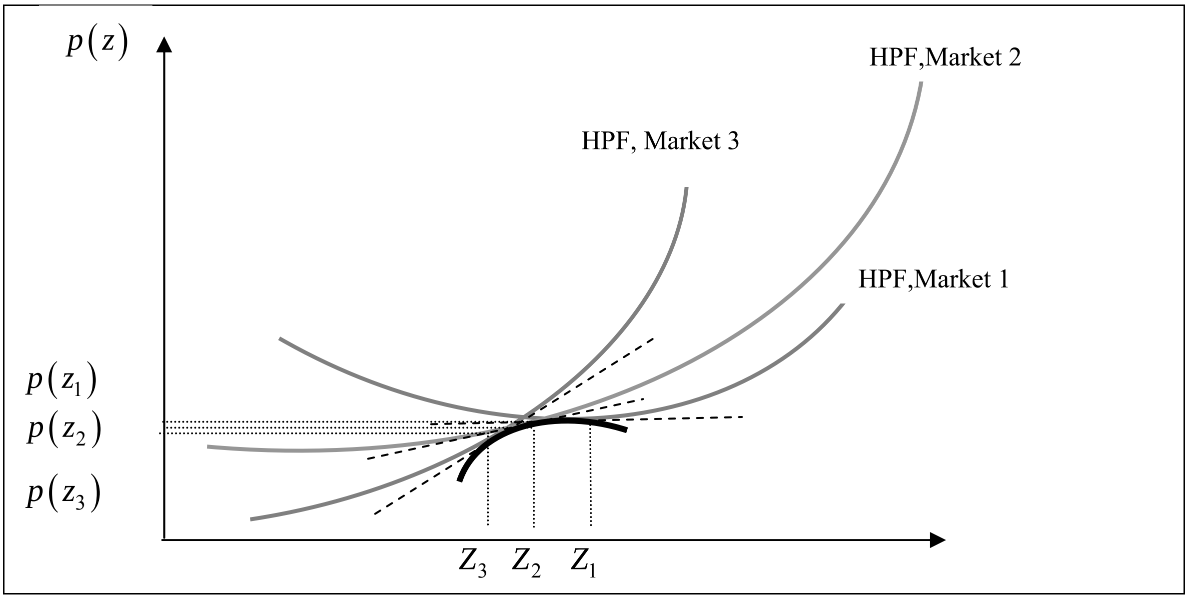
Figure 3 Simultaneous choice of price and quantity in the hedonic model*



*Adapted from Epple, 1987.

The identification of preferences from variation in the hedonic price functions are illustrated in Figure 4. 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 Illustration of identification in the multi-market case



HPF = hedonic price function.

4.3. 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.[3] 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.[4] 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 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

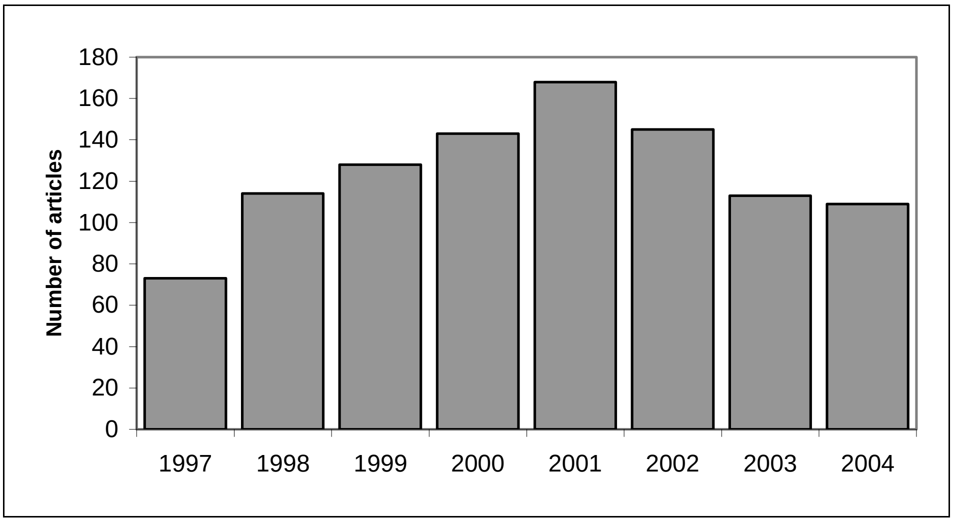
Consumers receive information about the connection between health and the intake of fat through various channels. This includes the internet, face-to-face conversations, television and newspapers. As it is not possible to capture all these diverse types of information most studies incorporating the effect 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

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

[4] 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.

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.[5] 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 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 5 Absolute number of hits in newspapers about the link between consumption of fat and health



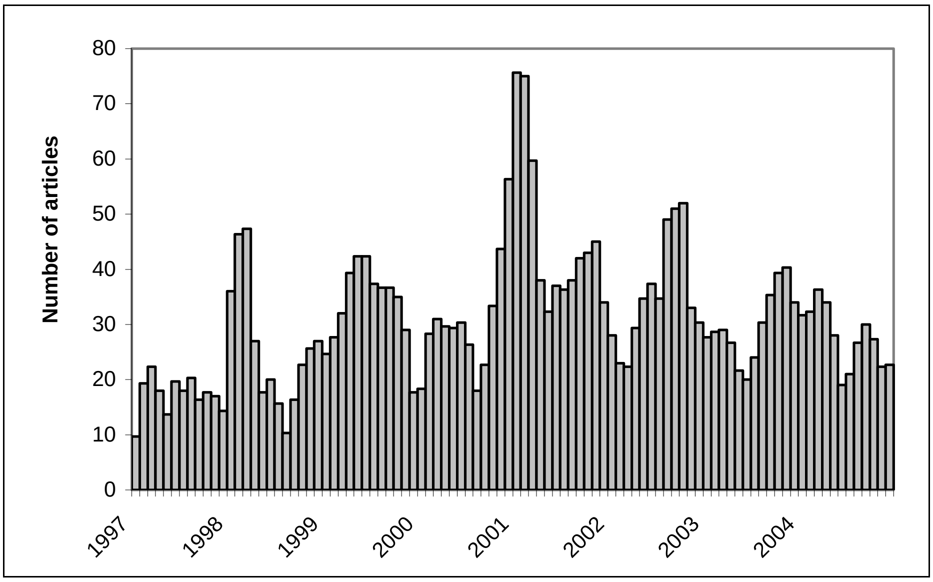
The articles are aggregated over newspapers independently of the size or location of the article. Several of the indices introduced in the literature use a lag structure, as they find that press coverage has a cumulative effect. This includes simple cumulative indices as in McGuirk et al. (1995) and Schmidt 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.[6] 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

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

[6] We have also tried a cumulative 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.

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 6.

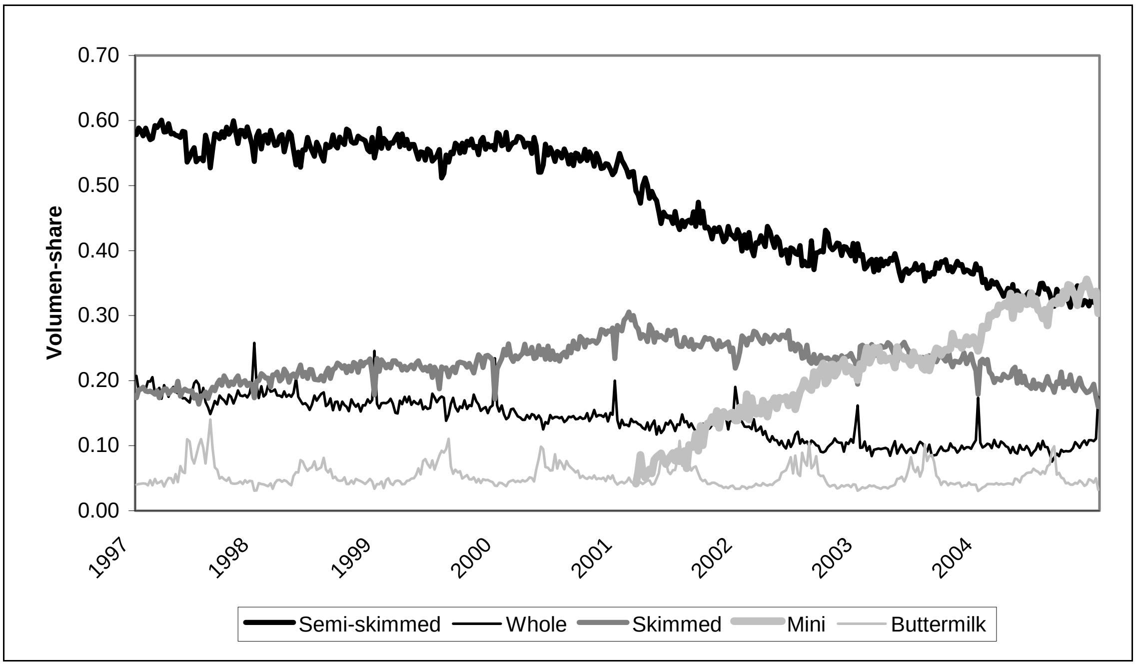
Figure 6 Number of hits in newspapers per month, three months floating index



The milk market

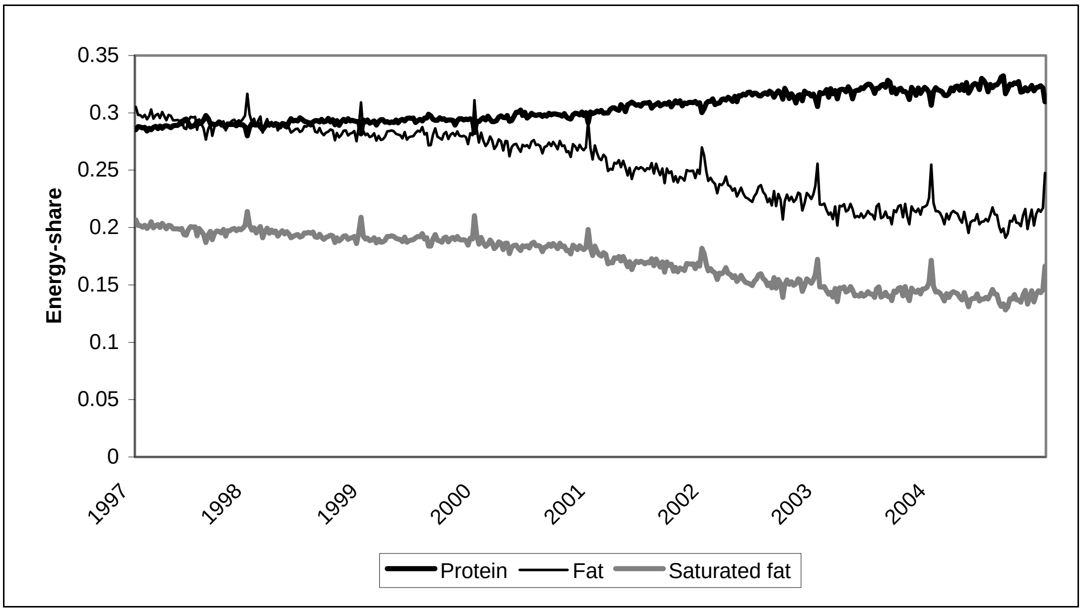
Until February 2001, there were four major types of milk on the Danish market: Whole milk, semi-skimmed milk, skimmed milk and buttermilk. Whole milk has a fat content of 3.5 per cent, semi-skimmed milk of 1.5 per cent, skimmed milk and buttermilk has a fat content of 0.1 per cent. Furthermore, buttermilk 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 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 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 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 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 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 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 were 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 1 Price elasticities before and after the introduction of mini milk

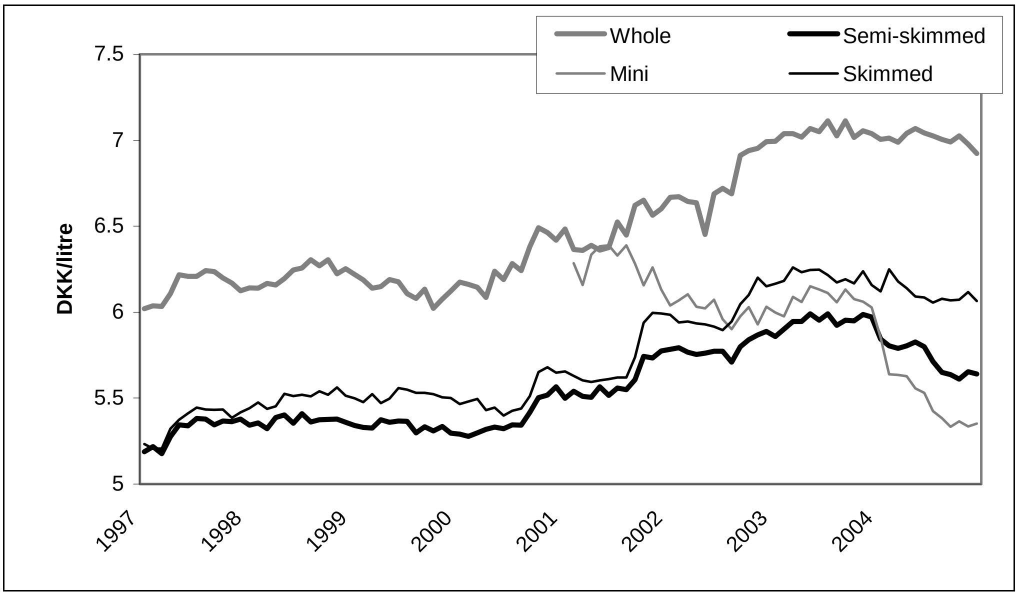
	Whole milk	Semi-skimmed milk	Skimmed milk	Mini milk
January 1997 to February 2001				
Whole milk	-1.45	0.12	0.00	-
Semi-skimmed milk	0.30	-1.16	0.36	-
Skimmed milk	0.00	0.16	-1.00	-
September 2001 to September 2002				
Whole milk	-1.44	0.32	0.06	0.06
Semi-skimmed milk	0.78	-1.68	0.03	0.74
Skimmed milk	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 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 9 Development in average milk prices



In the following figures the consumption of fat in milk for different types of households is described. Figure 10 illustrates the development in average grams of fat per litre of milk for households where the head has different level of education.[7] 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

[7] 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.

it is those with no further education and the longest educated who consume milk with a high fat content.

Figure 10 Fat per litre of milk for households with different education

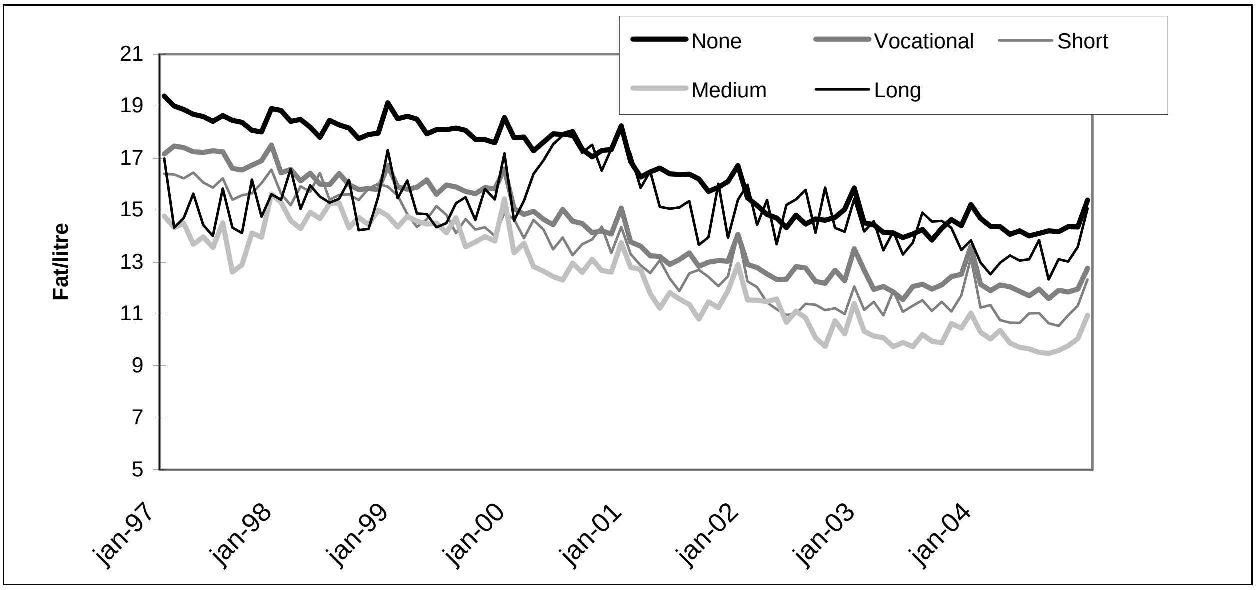


Figure 11 illustrates the development in average grams of fat per litre of milk for different family types. In 1997 households with children between the age of 0 and 3 distinguish themselves by consuming milk with more fat per litre than other households. Families with older children seem to prefer a 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.

Figure 11 Fat per litre of milk for different family types

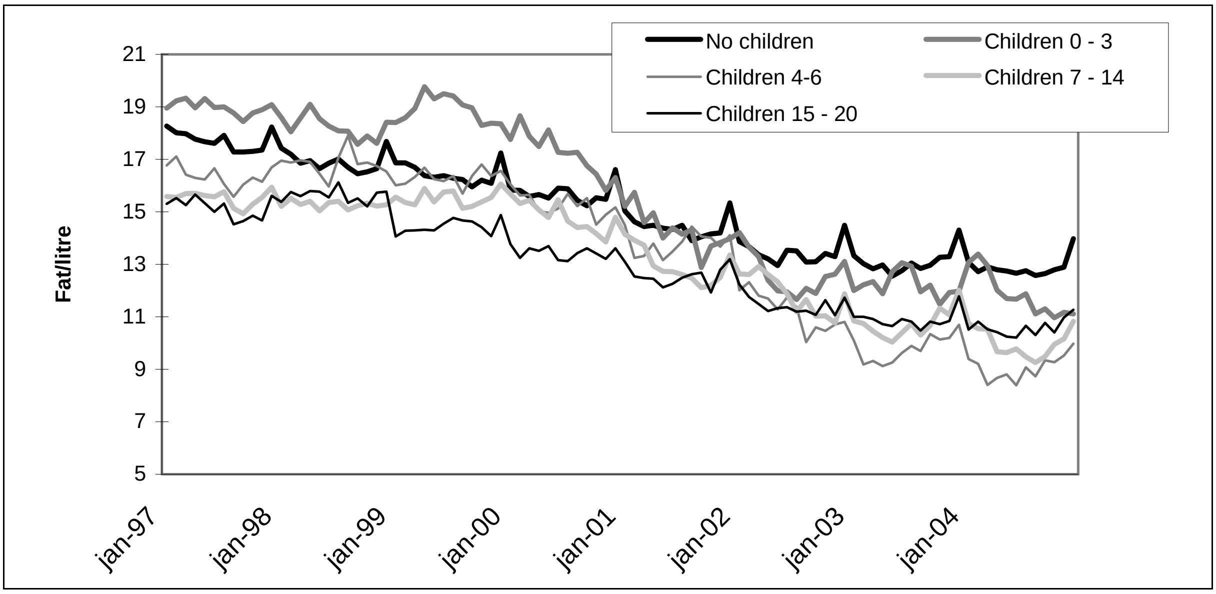
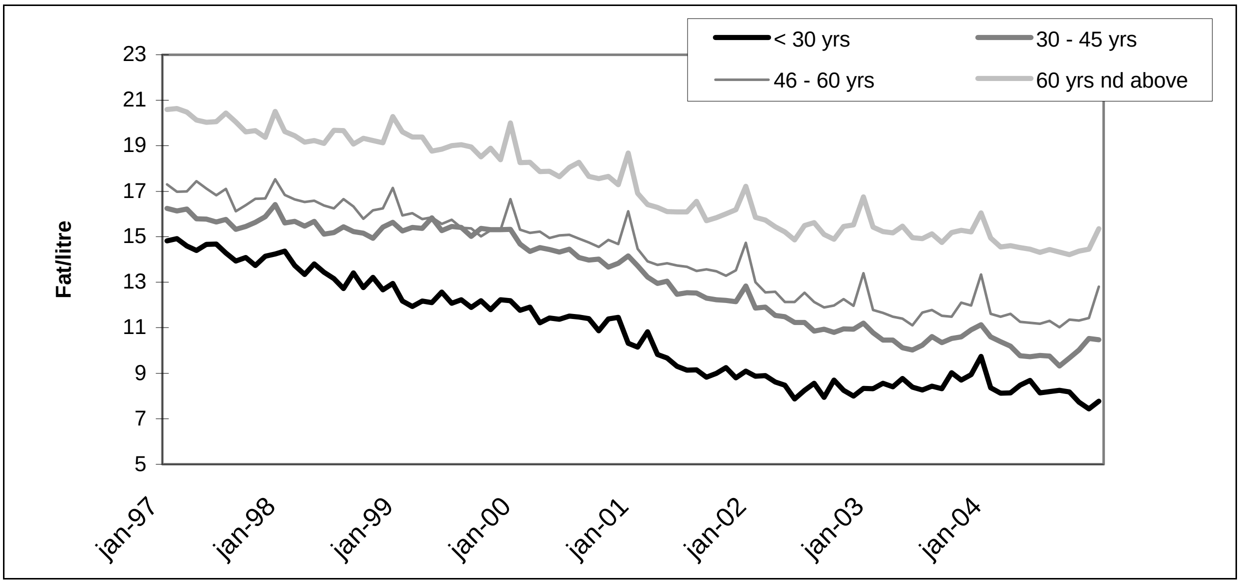


Figure 12 illustrates the development in average grams of fat per litre of milk for households in different age groups. In general, older households consume more fatty milk than other households. Younger people below the age of 30 consume milk with the lowest fat content. As they have a moderate consumption 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 12 Fat per litre of milk for households in different age groups



4.4. 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. Individual prices is then estimated for the households assuming that the household visits several markets i.e. goes 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, super markets and other stores.[8] 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 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.

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

Figure 13 Share of milk bought in each kind of store in each region

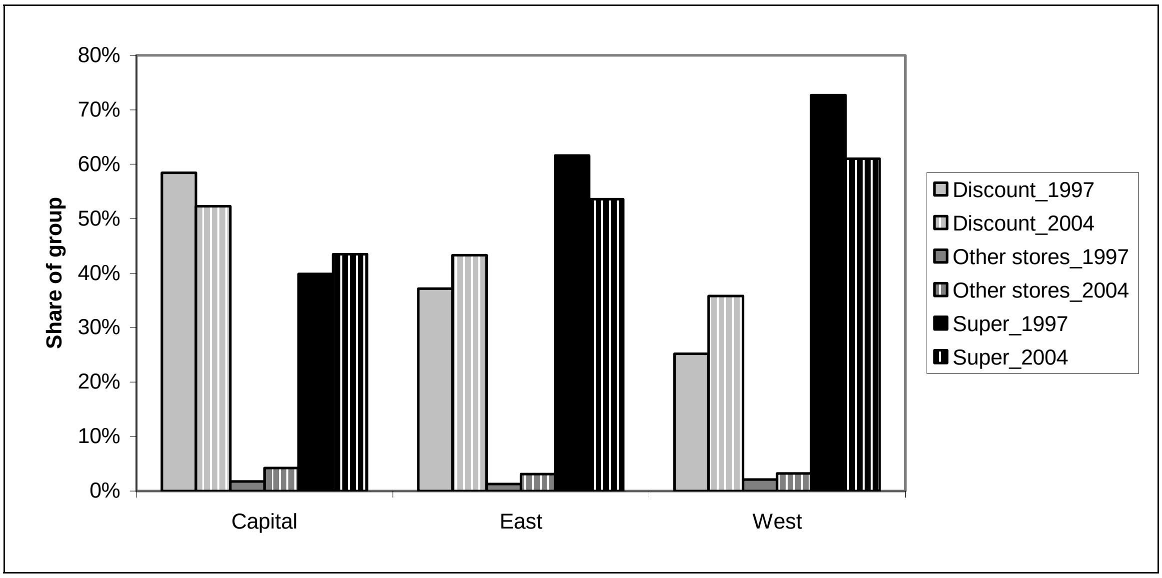
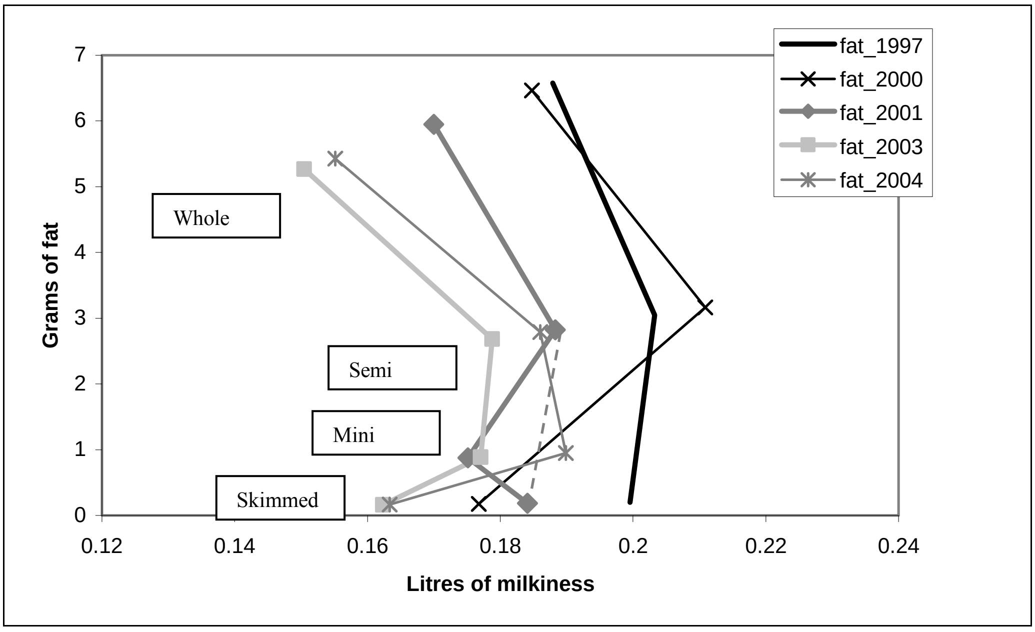


Figure 14 The empirical consumption set, capital, discount, conventional, standard dairy show how much of each of the characteristics fat and milkiness you get if you use one DKK on a particular type of milk, i.e. this is the empirical version of the theoretical Figure 1. In 1997 one DKK used on skimmed 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, 1999 and 2002 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.

Figure 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 dashed 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 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.[9] 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 15 The average price for milk in various stores and regions, standard dairy

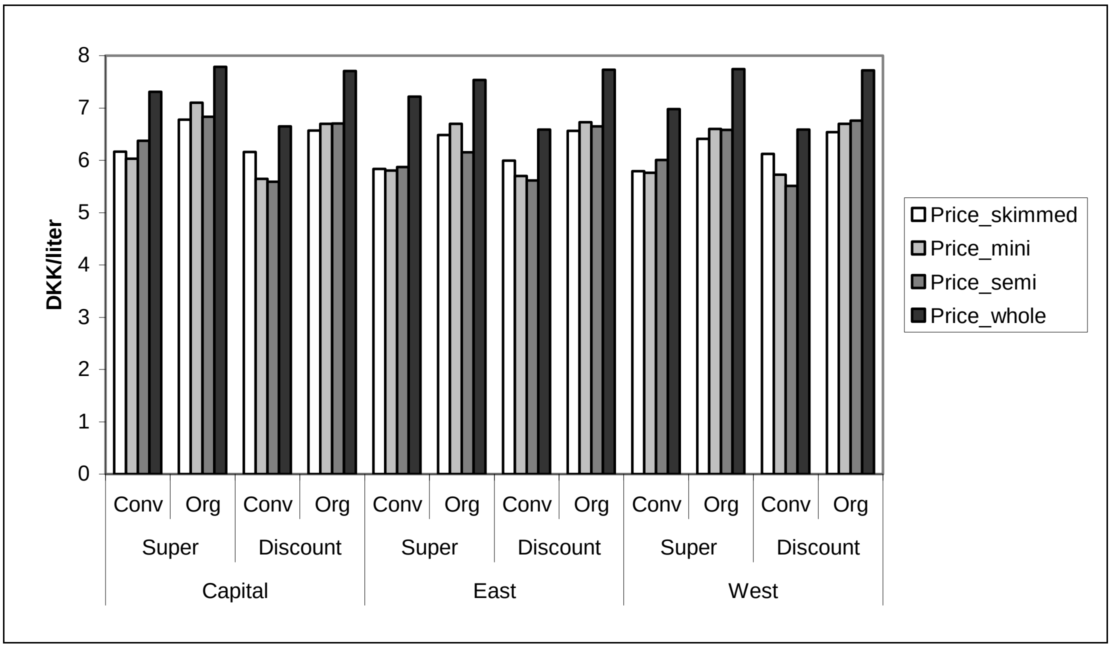
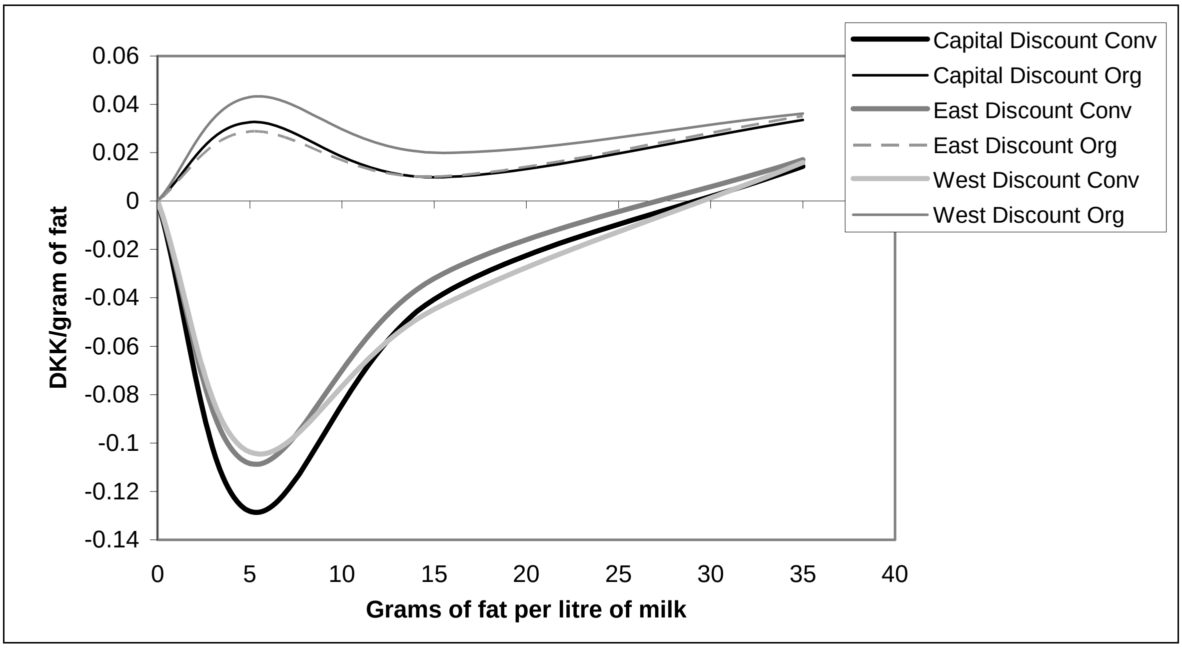


Figure 16 is a crude illustration of the hedonic price function for fat illustrated for selected markets. The figure is used to illustrate the motive behind choosing a quadratic form for the hedonic price function and separate markets for organic and conventional. The figure is crude in the sense that 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

[9] 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.

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 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 16 A crude empirical hedonic price function for fat, year 2003, standard dairy

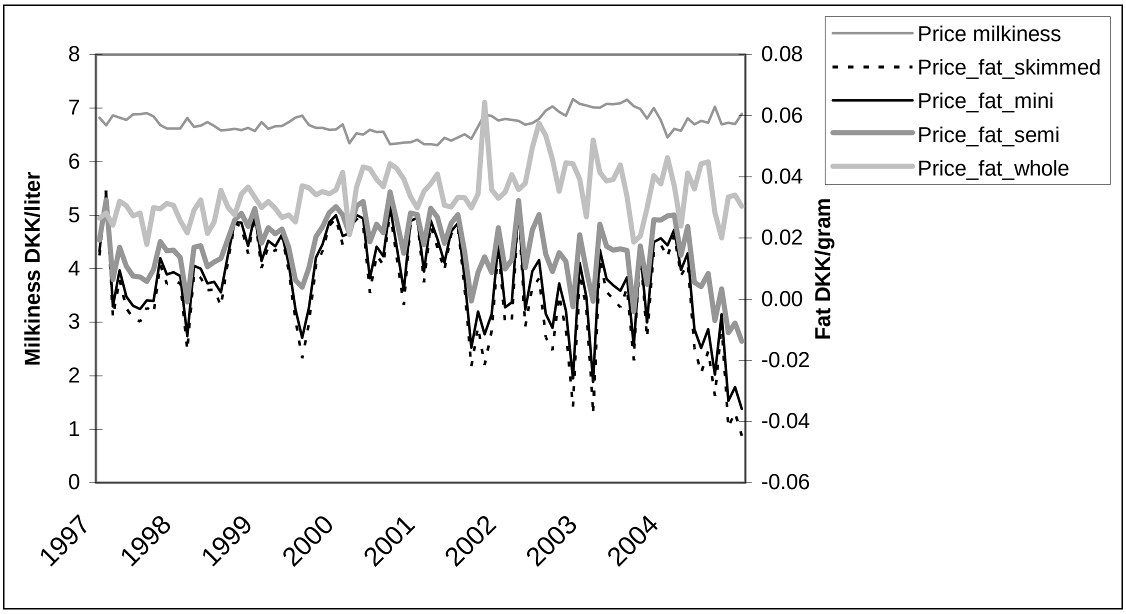


In the demand model we treat preferences for milk as separable from all other food, which of course is questionable as is all separability assumptions. Furthermore, we treat preferences for milkiness and fat as separable from the mode of produce (organic or conventional) and dairy (standard, discount or luxury dairy). As it appears from Figure 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.

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

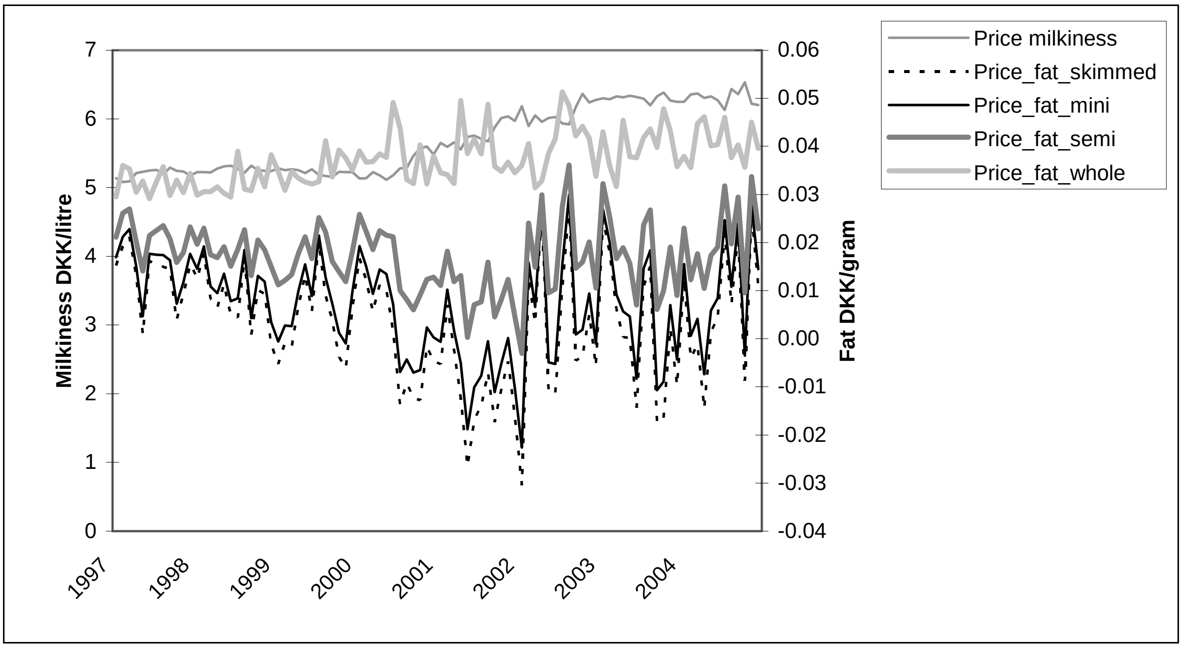
The constant accounts for the price of one litre of "milkiness", D_s and D_d, are dummies accounting for a luxury and discount dairy, respectively,[10] z_{fat} accounts for the content of fat in grams. The polynomial of second order implies that the price of fat varies with the type of milk; as illustrated in Figure 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 set of monthly implicit prices of characteristics, one for each market, equivalent to the two shown in Figure 17 and Figure 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 17 Hedonic prices for organic milkiness and fat, supermarkets in the capital



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

Figure 18 Hedonic prices for conventional milkiness and fat, supermarket in the capital



To construct individual 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.[11] 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 stores. The weighted price per gram of fat will then be calculated as:

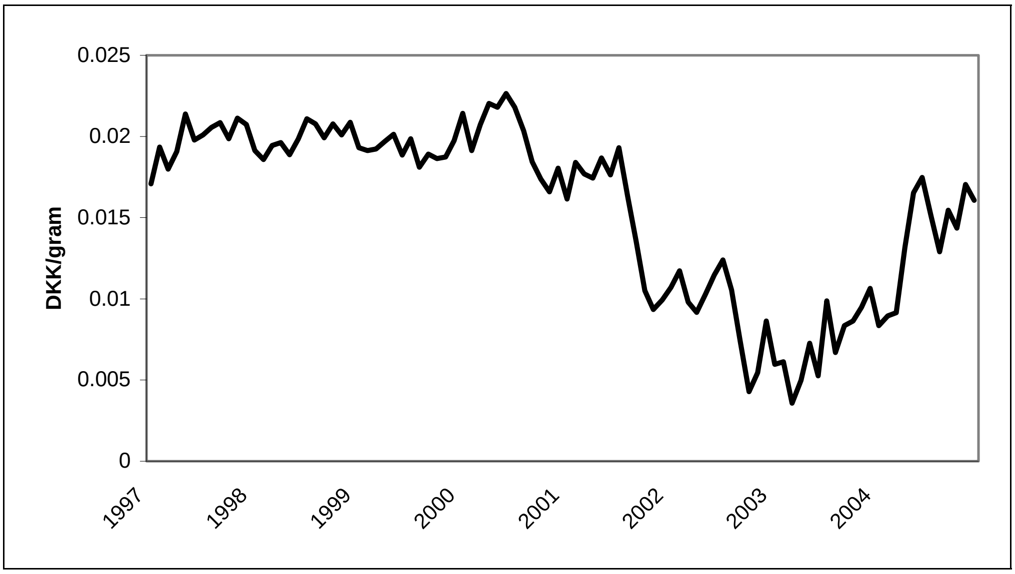
$$p_{fat}^{h} = \frac{fat_{semi,super}^{h}\, p_{semi\_fat,org,super} + fat_{semi,disc}^{h}\, p_{semi\_fat,org,discount} + fat_{whole,conv}^{h}\, p_{whole\_fat,other}}{totfat^{h}} \qquad (4.10)$$

The weighted price paid for fat over time averaged over households is shown in Figure 19. 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 9. The changes in prices have been accompanied by a general movement towards leaner types of milk (see Figure 7) which also adds to the lower price paid for fat. The price of fat is much higher for fat from whole milk (Figure 17 and Figure 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

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

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 19 Mean price paid for fat per gram



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

We assume a quadratic utility function. 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.

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

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 A for derivation).[12]

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

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

This result has a fine intuitive interpretation. Note that:

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

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).

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 in the utility function 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 are 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:

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

which means that in optimum we have:

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

rearranging leads to:

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

which can be further reduced to:

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

If we use the fact that

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

and substitute z_1 into equation (4.20) it becomes:

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

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:

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

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:

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

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:

$$0 = \left(\alpha_2^h + \tau_2^h \ln(t)\right)\frac{\pi_{1t}^h}{\pi_{2t}^h} - \left(1-\alpha_2^h + \tau_1^h \ln(t)\right) + \beta_1^h \frac{x_t^h}{\pi_{1t}^h} - \beta_1^h \frac{\pi_{2t}^h}{\pi_{1t}^h}\left(z_{2t}^h + \xi_{2t}^h\right) - \beta_2^h \frac{\pi_{1t}^h}{\pi_{2t}^h}\left(z_{2t}^h + \xi_{2t}^h\right) \tag{4.26}$$

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

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

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 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:

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

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:

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

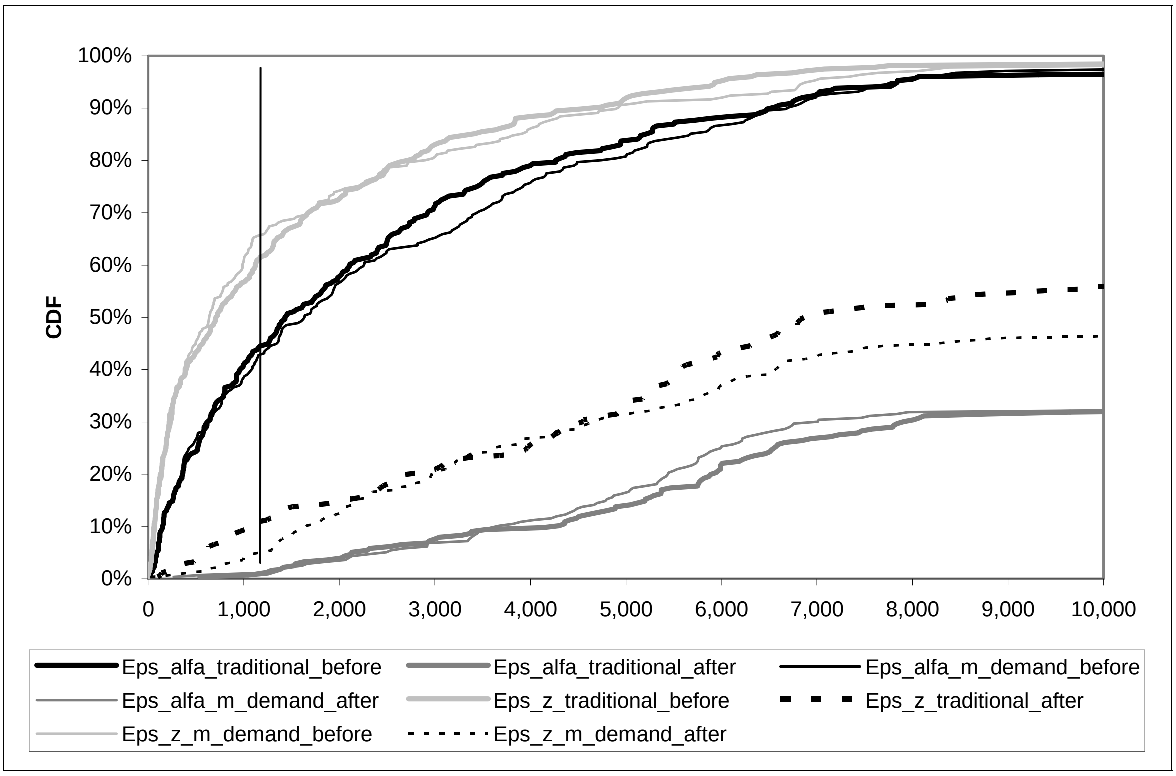
The other versions of the demand functions change in the same way due to instrumentation.

4.5. 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.[13] 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 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 2) and only estimate on households that are not censored. 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 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.

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

Figure 20 Mean squared percentage error on fat, random utility model and measurement error model

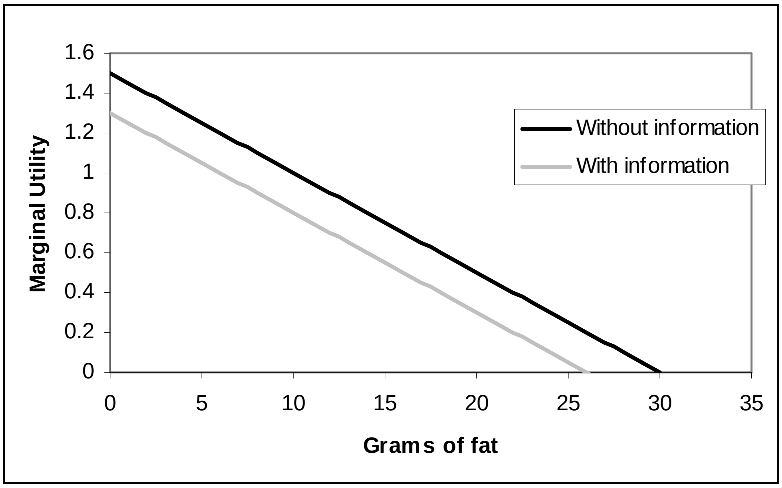


Based on the above realisation of the model we choose to estimate a classical model with measurement errors. This allows us to estimate the linear *m*-demand with measurement error as a two-sided censored Tobit model. Furthermore, we include exogenous information to account for changes in preferences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

4.6. Final model formulation: Tobit estimation, censoring and information

We model the influence of 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 21.

Figure 21 The way information influences the marginal utility for fat



This means that we get at utility function of the form:[14]

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

We do not include the trend and the information in the normalisation $\left(\alpha_1 + \alpha_2 = 1\right)$.

The *m*-demand from (4.27) becomes:

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

which can be rearranged to:

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

where

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

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

[14] 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.

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

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 (wholemilk). 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:

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

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

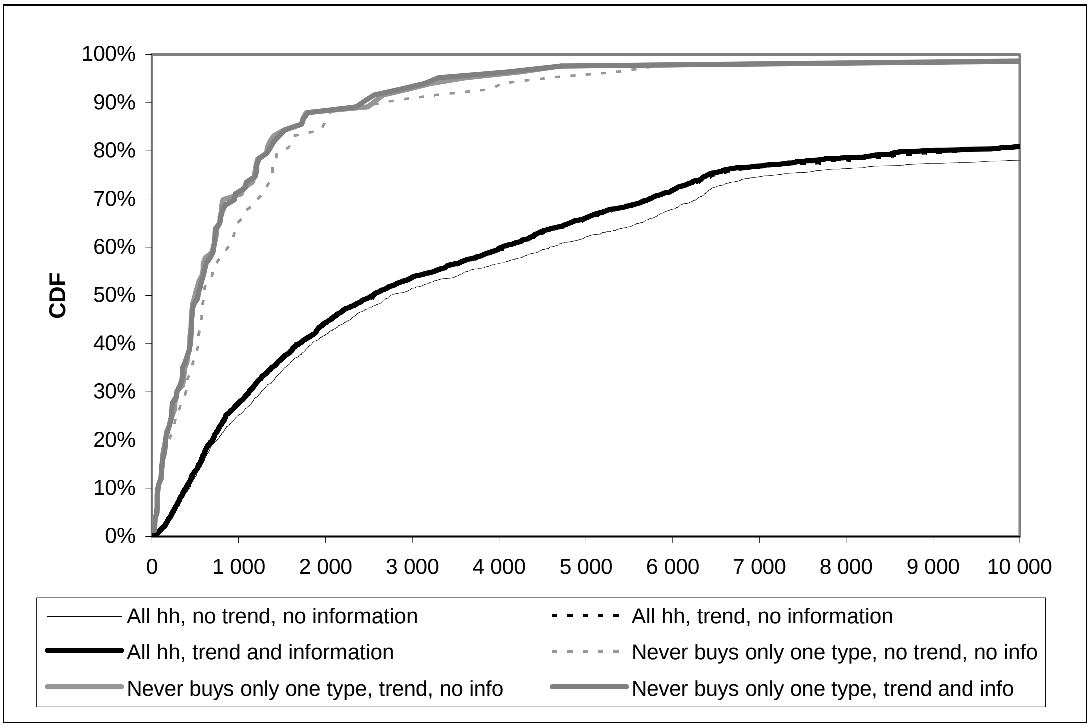
$$\hat{z}_{2t}^h = \omega_1^h + \omega_2^h \ln(t) + \omega_3^h I_t + \omega_4^h \frac{\pi_{2t}^h}{\pi_{1t}^h} + \omega_5^h \frac{\pi_{2t}^h}{\pi_{1t}^h} z_{1t}^h \tag{4.36}$$

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 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 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 20 only includes households who never buy only one type of milk (n=275). Figure 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.

4.7. 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 share 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 share 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 elasticites. The derivation of the own price elasticities for milkiness and cross-price elasticities between milkiness and fat are shown in Appendix B.

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 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 is found to be rather price elastic.

Figure 23 Predicted optimal fat share compared with actual purchases of milk in October 2000

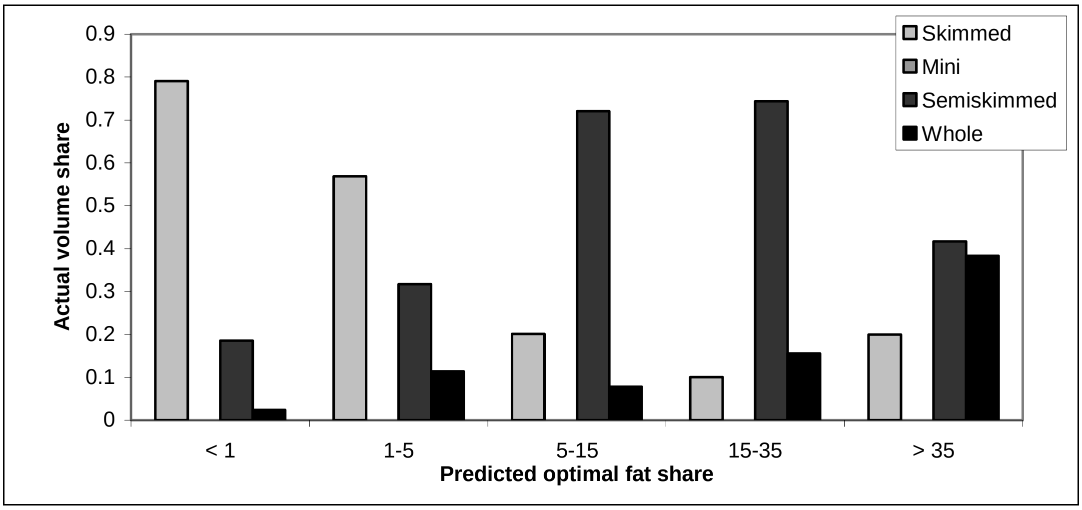


Figure 24 shows actual volume shares of different types of milk in October 2001 ten months after the entrance of mini milk separated by predicted optimal fat share based on parameters estimated in the period before the entrance of mini milk i.e. predicted optimal fat share is based on estimations in 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 23.

Figure 24 Predicted optimal fat share compared with actual purchases of milk in October 2001

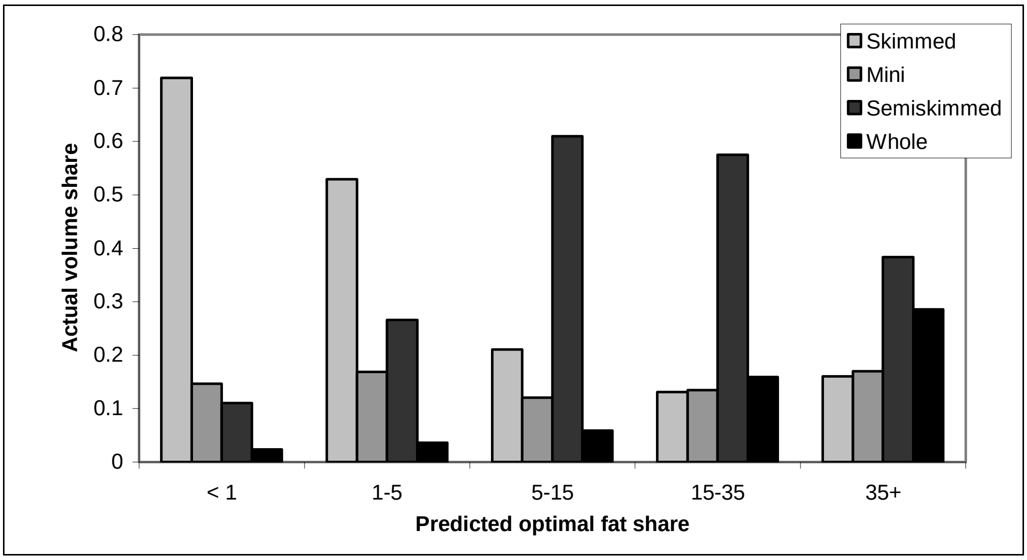
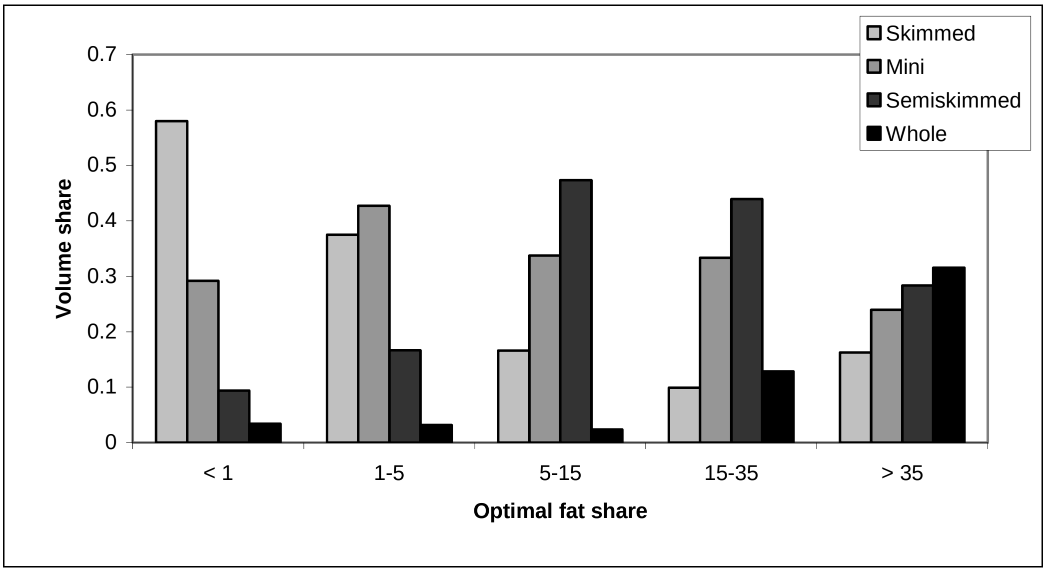


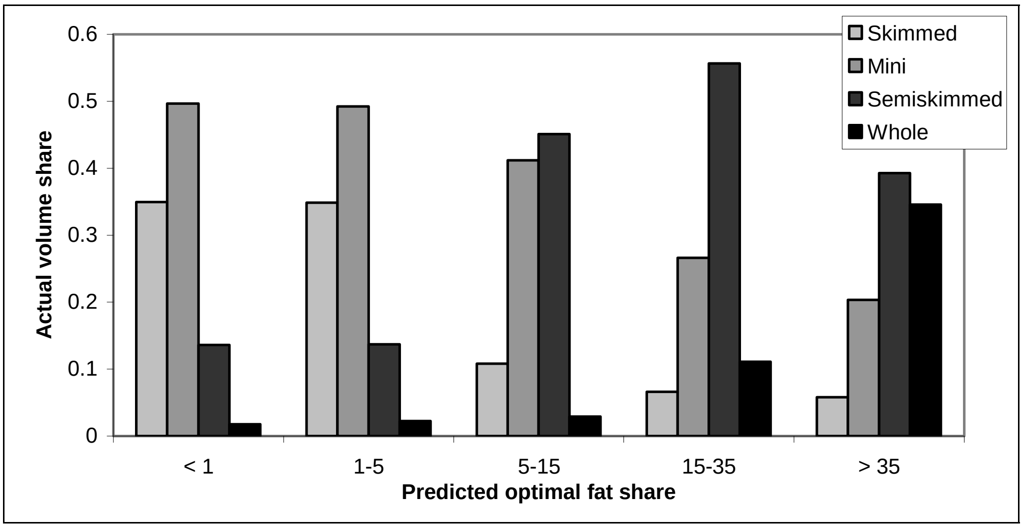
Figure 25 shows predicted optimal fat share based on estimated parameters in the period before the entrance of mini milk and actual purchase of milk in October 2004. This means that we are four years out of the estimation period. At this point mini milk has gained an almost stable volume share and prices 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 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 and thereby lost information in the estimations. The last Figure 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 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 14.

Figure 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 over time and for various social and demographic groups

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 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 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%	8.9%	91.1%
1998	3.2%	7.7%	6.2%	83.0%	8.5%	91.5%
1999	3.0%	6.1%	6.3%	84.6%	6.7%	93.3%
2000	2.6%	6.5%	6.8%	84.1%	7.2%	92.8%
2001	3.7%	14.5%	6.3%	75.5%	16.1%	83.9%
2002	2.6%	7.9%	7.5%	81.9%	8.8%	91.2%
2003	2.4%	9.1%	8.5%	80.1%	10.2%	89.8%
2004	2.5%	9.7%	7.6%	80.3%	10.8%	89.2%

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

Figure 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. To the left and the right of the grey lines in the figure are the areas 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 27 The density function for the optimal fat share over time

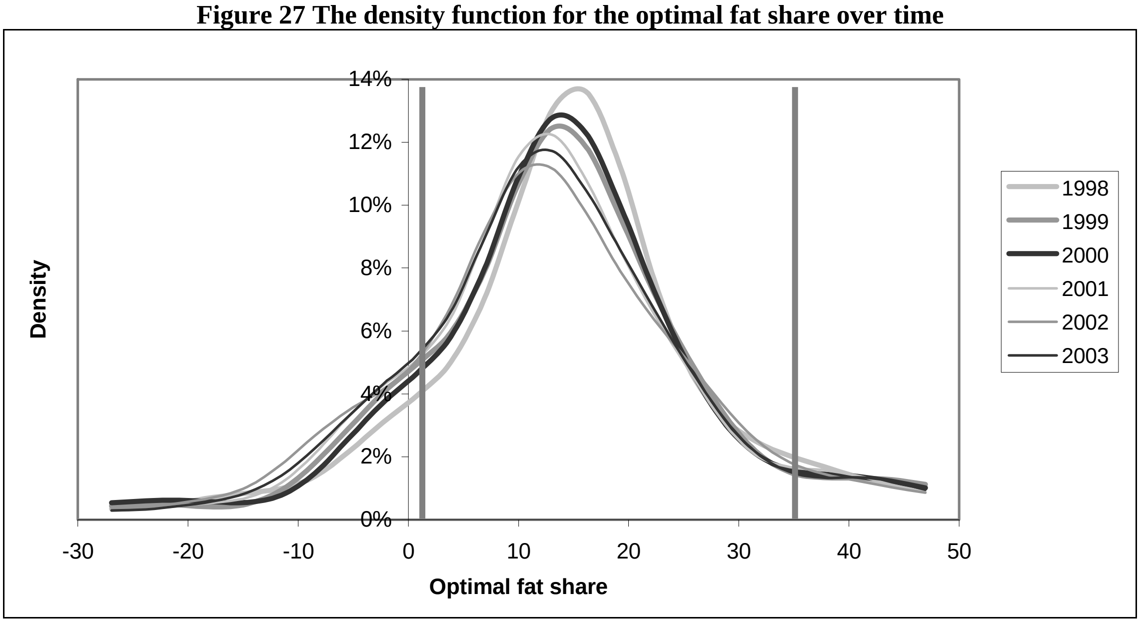


Figure 28 and Figure 29 show the optimal milkiness consumption together with optimal fat share. The milkiness haters are left out of the figures due to the definition of the optimal fat share. All columns in the figure sum to one. Many households, 40 per cent of the panel, have a moderate optimal milkiness 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 28) with 2004 (Figure 29).

Figure 28 Distribution of the panel over different optimal fat share and milkiness in 1997

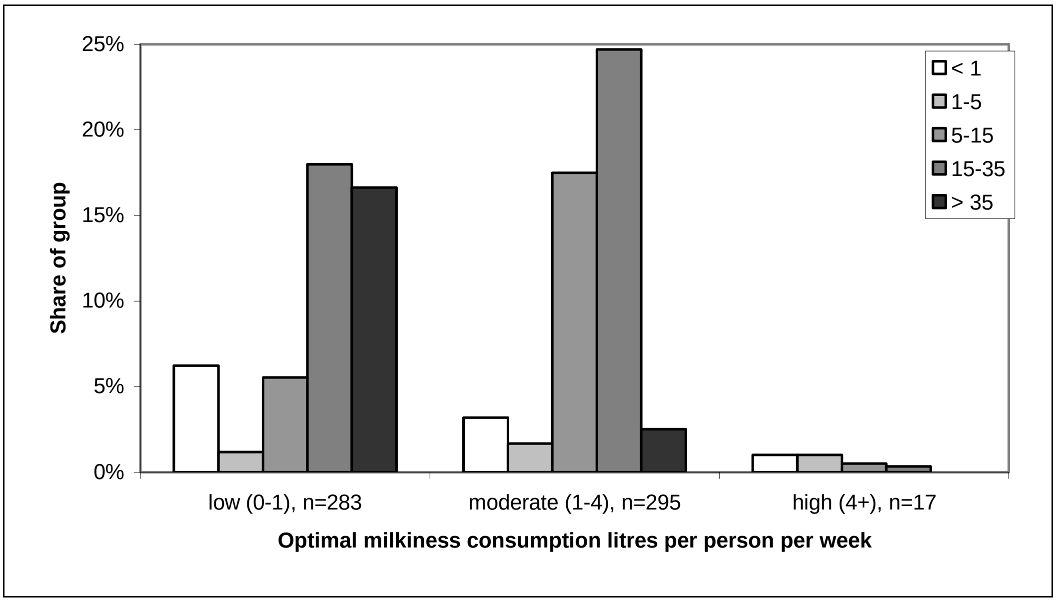


Figure 29 Distribution of the panel over different optimal fat share and milkiness in 2004

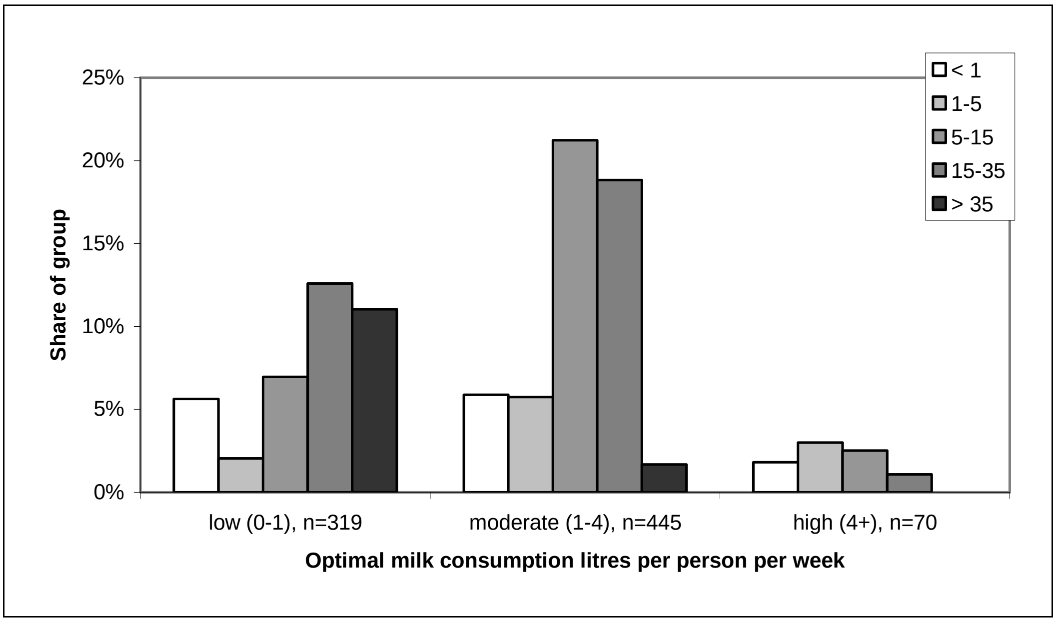


Figure 30 shows the optimal fat share for households with different level of education. There is not much difference between households with no or vocational education, while households with a longer or medium further education[15] prefer a lower fat content. Households with a short education show a distribution with two bulks, one around 12 and another around 32 grams of fat per litre of milk.

Figure 30 Distribution of optimal fat share for households in different educational groups, 1997

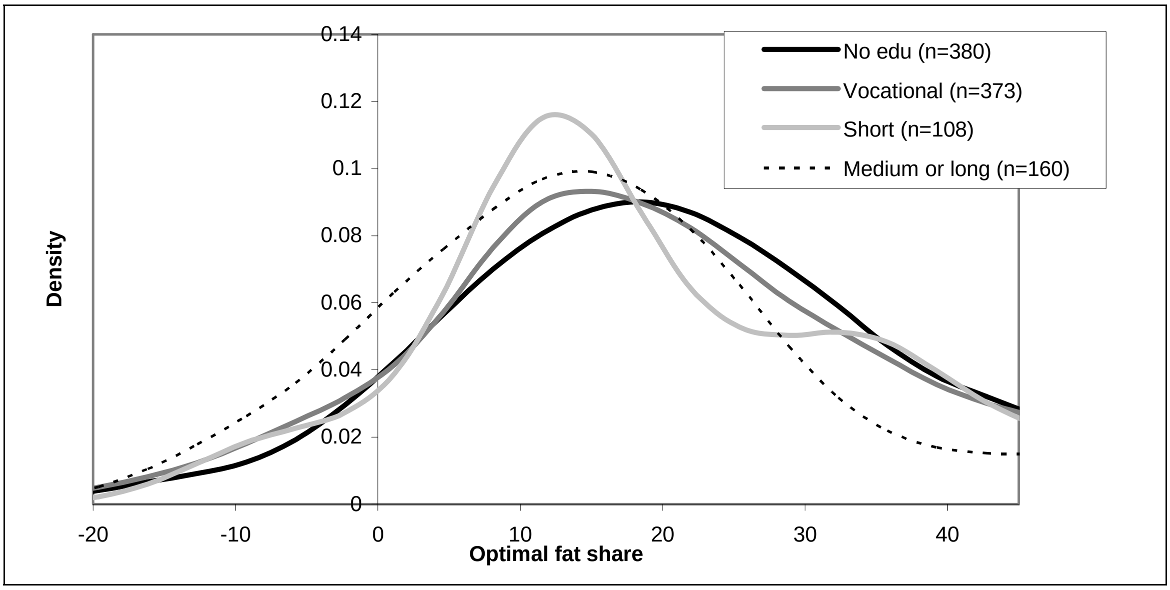
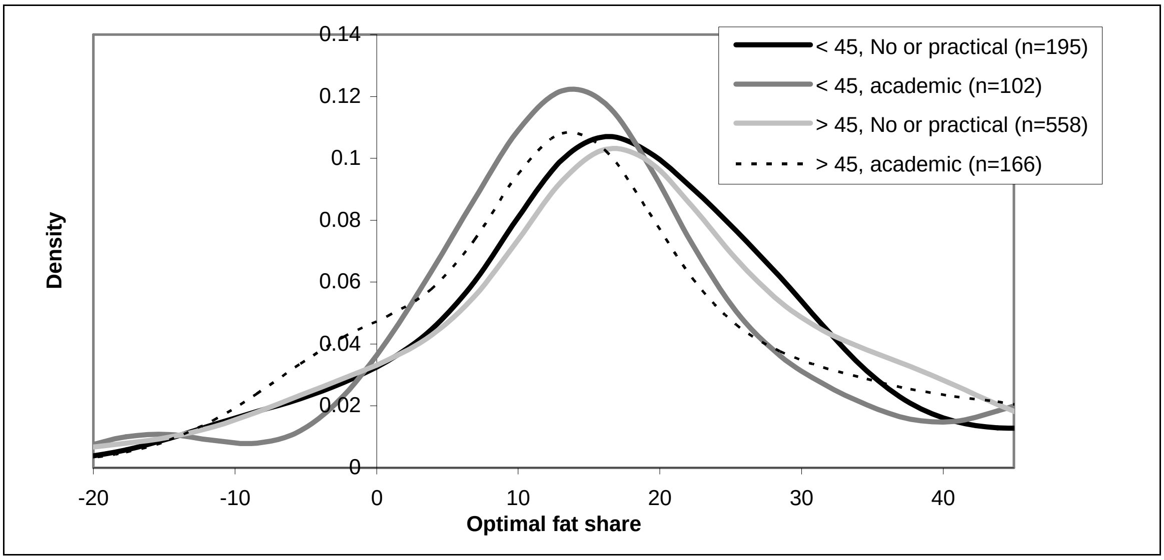


Figure 31 shows the distribution over fat share for a combination of education and age, note that the educational definitions here are slightly different, namely divided into either practical or no education versus theoretical education. For each of the age groups the theoretically educated prefer milk with lower fat content.

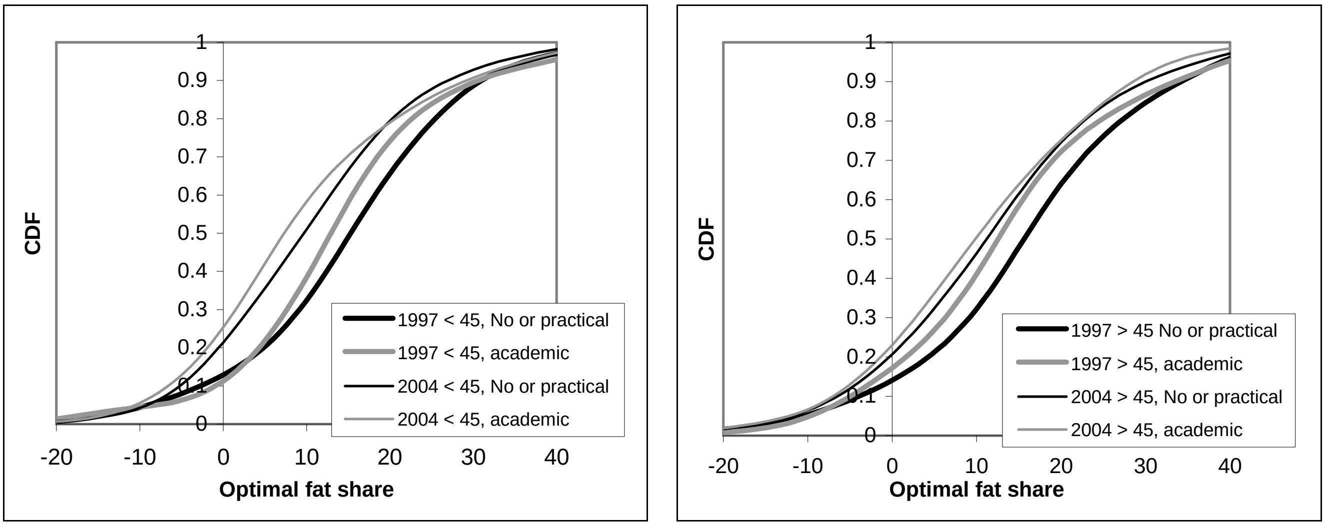
Figure 31 Distribution of fat share for a combination of education and age, 1997



[15] For a detailed description of the educational groups see chapter 2 in Smed (2008).

Figure 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 theoretically educated younger households while the practical or no educated older decrease their optimal fat share more than the theoretically educated older.

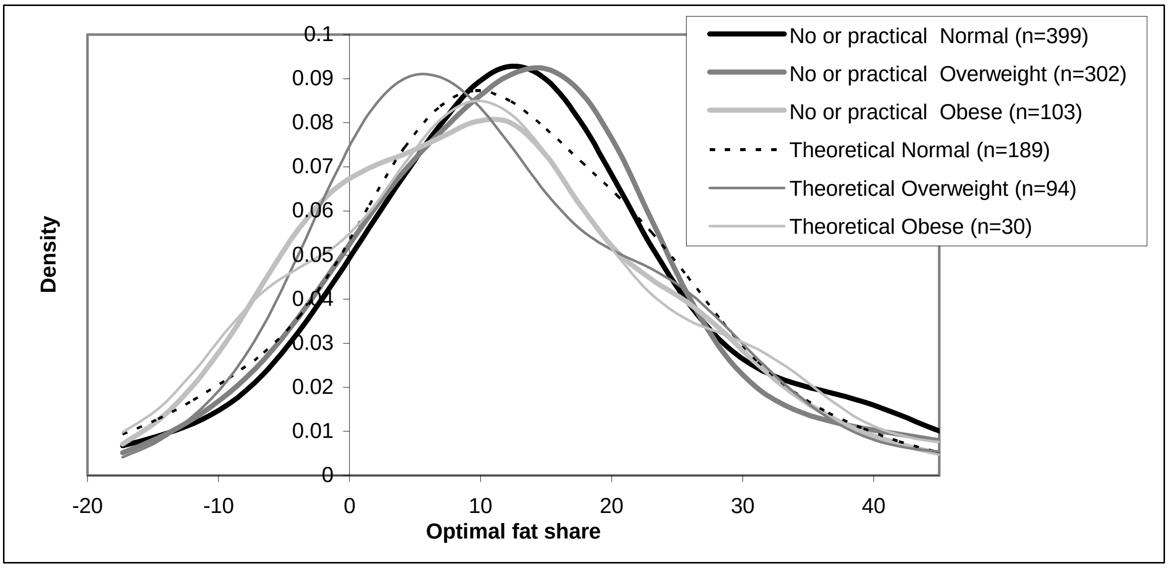
Figure 32 Change in CDF of optimal fat share for combinations of age and education



Finally, Figure 33 shows the distribution over optimal fat share in 2004 for combinations of BMI[16] 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.

[16] Questions of height and weight for each individual in the household are only posed in 2004. BMI is calculated as: $BMI = weight(kg)/(height(m))^2$. Overweight is then defined as a BMI above 25, but below 30, while obesity is defined as having a BMI above 30.

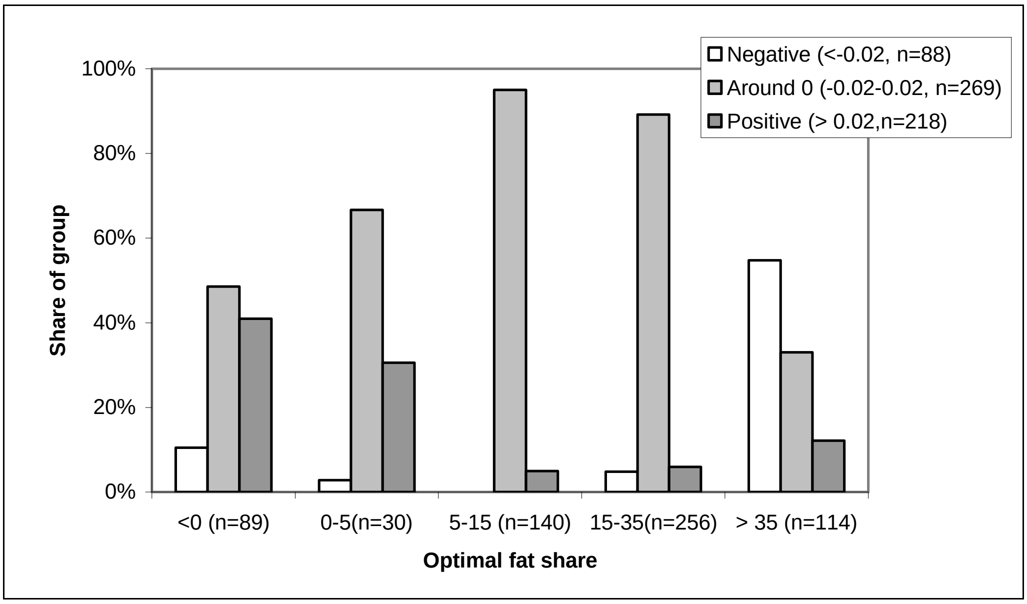
Figure 33 Distribution of optimal fat share for combinations of BMI and education



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 the 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 health. On average, 57 per cent of the households have a negative trend parameter. Figure 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 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 (defined as having 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 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 35 Optimal fat share and the sign of the information parameter in 1997

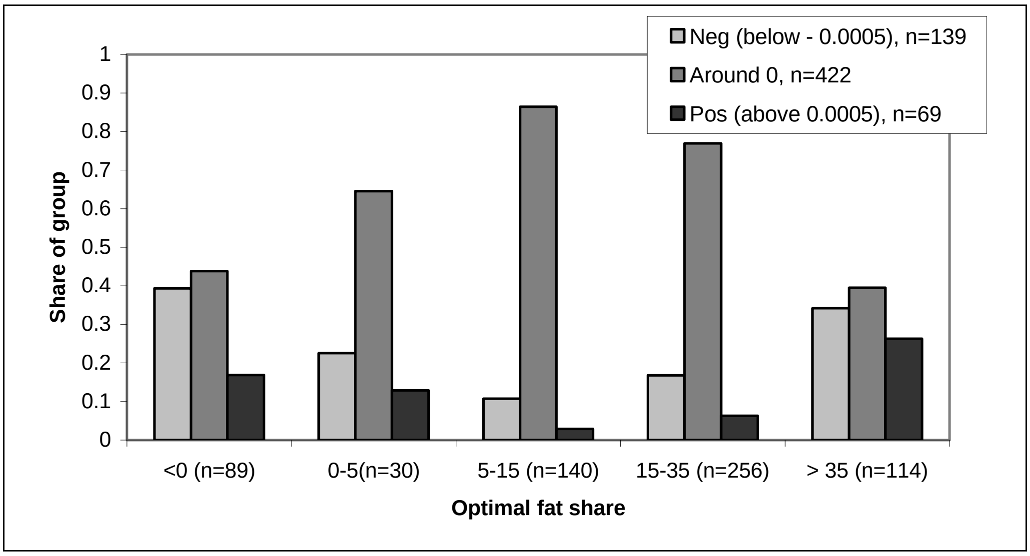
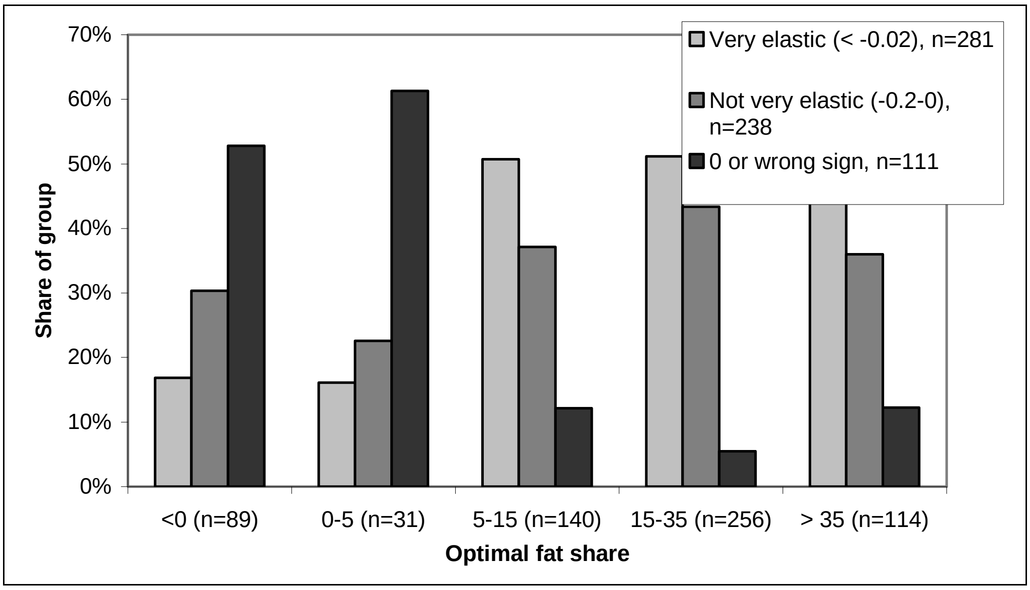


Figure 36 shows the price elasticity separated by optimal fat share (columns in each group sum to one). Most households have a negative own price elasticity for fat (17 per cent have an own price elasticity of 0 or with wrong sign). As much as 45 per cent are rather price elastic with an own price elasticity 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 seems natural as these households are on the edge of the consumption set, and the closest they can come to having their preferences fulfilled is to consume 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 36 Optimal fat share and mean own price elasticity in 1997



4.8. 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.
- Blow, L., M. Browning and I. Crawford (2005): Revealed preference analysis of characteristics models. *CAM working paper no 4*, CAM, University of Copenhagen.
- Blundell R.W. and A. Duncan (1998): Kernel regression in empirical microeconomics. *Journal of Human Resources*, **33** (1): 62-87.
- 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: Theory and Applications, Eighth World Congress*, Vol. 2, Chapter 8: 312-357. Cambridge University Press, Cambridge.
- 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.
- 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): 288-303.
- 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** (1): 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. 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 reaction to produced characteristics. *American Journal of Agricultural Economics*, **58**: 504-510.
- Lancaster K. (1966): A new approach to consumer theory. *Journal of Political Economy*, **4** (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.
- Piggot N.E. and T. 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.
- Schmidt 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.
- Smed S. (2005): Demand structure and willingness to pay for organic dairy products. *Cahiers Options Méditerraneées*, **64**: 33-44.

- 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.
- Smed, S. (2008): Empirical studies on "Health, Information and Consumer Behaviour". PhD thesis, University of Copenhagen.
- 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**: 513-520.
- 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 A: Derivation of demand function given quadratic utility

Assume the utility function:

$$U\left(z_t\right)=\left(\alpha+\varepsilon_t\right)' z_t - 0.5 z_t' B z_t,\ \varepsilon \sim N\left(0,\Sigma\right) \tag{41.1}$$

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:

$$\begin{aligned} &\max_z U\left(z\right) \\ &\text{s.t.} \quad x=\pi' z \end{aligned} \tag{41.2}$$

The Lagrange equation becomes:

$$\begin{aligned} L\left(z,\lambda\right) &= U\left(z\right)-\lambda\left(\pi' z - x\right) \\ &= -0.5 z'\beta z + \left(\alpha+\varepsilon\right)' z - \lambda\left(\pi' z - x\right) \end{aligned} \tag{41.3}$$

and the first-order conditions become:

$$\frac{\nabla L}{\nabla z} = -\beta z + \left(\alpha+\varepsilon\right) - \lambda\pi = 0 \tag{41.4}$$

$$\frac{\nabla L}{\nabla \lambda} = \pi' z - x = 0 \tag{41.5}$$

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

$$z = \beta^{-1}\left(\left(\alpha+\varepsilon\right)-\lambda\pi\right) \tag{41.6}$$

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

$$
\begin{aligned}
0 &= z'\left(\beta^{-1}\left(\left(\alpha+\varepsilon\right)-\lambda\pi\right)\right)-x \\
&= z'\beta^{-1}\left(\alpha+\varepsilon\right)-\lambda\pi'\beta^{-1}\pi-x \\
&\Updownarrow \\
\lambda &= \left(\pi'\beta^{-1}\pi\right)^{-1}\left(\pi'\beta^{-1}\left(\alpha+\varepsilon\right)-x\right) \\
&= \left(\underbrace{\underbrace{\pi'}_{1\times J}\underbrace{\beta^{-1}}_{J\times J}\underbrace{\pi}_{J\times 1}}_{1\times 1}\right)^{-1}\left(\underbrace{\underbrace{\pi'}_{1\times J}\underbrace{\beta^{-1}}_{J\times J}\underbrace{\left(\alpha+\varepsilon\right)}_{J\times 1}-\underbrace{x}_{1\times 1}}_{1\times 1}\right)
\end{aligned}
\tag{41.7}
$$

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

$$
\begin{aligned}
z &= \beta^{-1}\left(\left(\alpha+\varepsilon\right)-\lambda\pi\right) \\
&= \beta^{-1}\left(\left(\alpha+\varepsilon\right)-\left(\pi'\beta^{-1}\pi\right)^{-1}\left(\pi'\beta^{-1}\left(\alpha+\varepsilon\right)-x\right)\pi\right)
\end{aligned}
\tag{41.8}
$$

Rearranging (A.8) leads to:

$$
z = \beta^{-1}\left(\alpha+\varepsilon\right)-\left(\beta^{-1}\pi\left(\pi'\beta^{-1}\pi\right)^{-1}\right)\left(\pi'\beta^{-1}\left(\alpha+\varepsilon\right)-x\right)
\tag{41.9}
$$

with the dimensions:

$$
\underbrace{z}_{J\times 1} = \underbrace{\underbrace{\beta^{-1}}_{J\times J}\underbrace{\left(\underbrace{\alpha}_{J\times 1}+\underbrace{\varepsilon}_{J\times 1}\right)}_{J\times 1}}_{J\times 1}-\underbrace{\left(\underbrace{\beta^{-1}}_{J\times J}\underbrace{\pi}_{J\times 1}\underbrace{\left(\underbrace{\pi'}_{1\times J}\underbrace{\beta^{-1}}_{J\times J}\underbrace{\pi}_{J\times 1}\right)^{-1}}_{1\times 1}\right)}_{J\times 1}\underbrace{\left(\underbrace{\pi'}_{1\times J}\underbrace{\beta^{-1}}_{J\times J}\underbrace{\left(\underbrace{\alpha}_{J\times 1}+\underbrace{\varepsilon}_{J\times 1}\right)}_{J\times 1}-\underbrace{x}_{1\times 1}\right)}_{1\times 1}
\tag{41.10}
$$

Appendix B: Derivation of elasticities in the Tobit model

Fat:

The predicted demand for fat is given by:

$$z_2 = \omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} + \omega_5 \frac{\pi_2}{\pi_1} z_1 \tag{42.1}$$

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

$$z_1 = \frac{x - \pi_2 z_2}{\pi_1}, \quad z_2 = \frac{x - \pi_1 z_1}{\pi_2} \tag{42.2}$$

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

$$z_2 = \left(\frac{\pi_1^2}{\pi_1^2 + \omega_5 \pi_2^2} \right) \left(\omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} + \omega_5 \frac{\pi_2}{\pi_1} \frac{x}{\pi_1} \right) \tag{42.3}$$

this can be translated into:

$$\begin{gathered} z_2 \equiv \frac{f_2}{g_2} h_2, \\ f_2 \equiv \pi_1^2, \quad g_2 \equiv \pi_1^2 + \omega_5 \pi_2^2, \quad h_2 \equiv \omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} + \omega_5 \frac{\pi_2}{\pi_1} \frac{x}{\pi_1} \end{gathered} \tag{42.4}$$

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

$$\frac{\partial \left(\frac{f}{g} h \right)}{\partial y} = \frac{g \frac{\partial f}{\partial y} - f \frac{\partial g}{\partial y}}{g^2} h + \frac{f}{g} \frac{\partial h}{\partial y} \tag{42.5}$$

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

$$\begin{gathered} \frac{\partial f_2}{\partial \pi_1} = 2\pi_1, \quad \frac{\partial f_2}{\partial \pi_2} = 0, \quad \frac{\partial f_2}{\partial x} = 0, \quad \frac{\partial g_2}{\partial \pi_1} = 2\pi_1, \quad \frac{\partial g_2}{\partial \pi_2} = 2\omega_5 \pi_2, \quad \frac{\partial g_2}{\partial x} = 0 \\ \frac{\partial h_2}{\partial \pi_1} = -\omega_4 \frac{\pi_2}{\pi_1^2} - 2\omega_5 \frac{\pi_2 x}{\pi_1^3}, \quad \frac{\partial h_2}{\partial \pi_2} = \frac{\omega_4}{\pi_1} + \omega_5 \frac{x}{\pi_1^2}, \quad \frac{\partial h_2}{\partial x} = \omega_5 \frac{\pi_2}{\pi_1^2} \end{gathered} \tag{42.6}$$

Define

$$D \equiv \pi_1^2 + \omega_5 \pi_2^2$$
$$C \equiv \omega_1 + \omega_2 \ln(t) + \omega_3 I + \omega_4 \frac{\pi_2}{\pi_1} \tag{42.7}$$

then the elasticities become:

$$\frac{\partial z_2}{\partial \pi_1}\frac{\pi_1}{z_2} = \frac{\partial\left(\frac{f_2}{g_2}h_2\right)}{\partial \pi_1}\frac{\pi_1}{z_2} = \left(\frac{2\omega_5\pi_1\pi_2^2}{D^2}\left(C + \omega_5\frac{\pi_2}{\pi_1}\frac{x}{\pi_1}\right) + \frac{\pi_2}{D}\left(-\omega_4 - 2\omega_5\frac{x}{\pi_1}\right)\right)\frac{\pi_1}{z_2} \tag{42.8}$$

$$\frac{\partial z_2}{\partial \pi_2}\frac{\pi_2}{z_2} = \frac{\partial\left(\frac{f_2}{g_2}h_2\right)}{\partial \pi_2}\frac{\pi_2}{z_2} = \left(\frac{-2\omega_5\pi_1^2\pi_2}{D^2}\left(C + \omega_5\frac{\pi_2}{\pi_1}\frac{x}{\pi_1}\right) + \frac{\pi_1}{D}\left(\omega_4 + \omega_5\frac{x}{\pi_1}\right)\right)\frac{\pi_2}{z_2} \tag{42.9}$$

$$\frac{\partial z_2}{\partial x}\frac{x}{z_2} = \frac{\partial\left(\frac{f_2}{g_2}h_2\right)}{\partial x}\frac{x}{z_2} = \frac{\omega_5\pi_2}{D}\frac{x}{z_2} \tag{42.10}$$

Milk:

From the equations (42.1) and (42.2) it is also possible to calculate z_1 as a function of the budget:

$$z_1 = \left(\frac{\pi_1\pi_2}{\omega_5\pi_2^2 + \pi_1^2}\right)\left(-\omega_1 - \omega_2\ln(t) - \omega_3 I - \omega_4\frac{\pi_2}{\pi_1} + \frac{x}{\pi_2}\right) \tag{42.11}$$

Checking that the cost of z_1 and z_2 ad up to the budget:

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 (42.3) and (42.11)):

$$z_1 = \left(\frac{\pi_1\pi_2}{D}\right)\left(\frac{x}{\pi_2} - C\right), \quad z_2 = \left(\frac{\pi_1^2}{D}\right)\left(C + \omega_5\frac{\pi_2}{\pi_1}\frac{x}{\pi_1}\right) \tag{42.12}$$

and the price of the choices is

$$\pi_1 z_1 + \pi_2 z_2 = \pi_1 \left(\frac{\pi_1 \pi_2}{D} \right) \left(\frac{x}{\pi_2} - C \right) + \pi_2 \left(\frac{\pi_1^2}{D} \right) \left(C + \omega_5 \frac{\pi_2}{\pi_1} \frac{x}{\pi_1} \right) = x \tag{42.13}$$

as desired.

In order to calculate the elasticities we reformulate z_1 in the same way as we reformulated z_2 in (42.4):

$$\begin{gathered} z_1 \equiv \frac{f_1}{g_1} h_1, \\ f_1 \equiv \pi_1 \pi_2, \quad g_1 \equiv \pi_1^2 + \omega_5 \pi_2^2, \quad h_1 \equiv -\omega_1 - \omega_2 \ln(t) - \omega_3 I - \omega_4 \frac{\pi_2}{\pi_1} + \frac{x}{\pi_2} \end{gathered} \tag{42.14}$$

again we calculate the derivatives:

$$\begin{gathered} \frac{\partial f_1}{\partial \pi_1} = \pi_2, \quad \frac{\partial f_1}{\partial \pi_2} = \pi_1, \quad \frac{\partial f_1}{\partial x} = 0, \quad \frac{\partial g_1}{\partial \pi_1} = 2\pi_1, \quad \frac{\partial g_1}{\partial \pi_2} = 2\omega_5 \pi_2, \quad \frac{\partial g_1}{\partial x} = 0 \\ \frac{\partial h_1}{\partial \pi_1} = \omega_4 \frac{\pi_2}{\pi_1^2}, \quad \frac{\partial h_1}{\partial \pi_2} = -\frac{\omega_4}{\pi_1} - \frac{x}{\pi_2^2}, \quad \frac{\partial h_1}{\partial x} = \frac{1}{\pi_2} \end{gathered} \tag{42.15}$$

and again this leads to a set of elasticities:

$$\frac{\partial z_1}{\partial \pi_1} \frac{\pi_1}{z_1} = \frac{\partial \left(\frac{f_1}{g_1} h_1 \right)}{\partial \pi_1} \frac{\pi_1}{z_1} = \left(\frac{\omega_5 \pi_2^3 - \pi_1^2 \pi_2}{D^2} \left(-C + \frac{x}{\pi_2} \right) + \frac{\pi_1 \pi_2}{D} \left(\omega_4 \frac{\pi_2}{\pi_1^2} \right) \right) \frac{\pi_1}{z_1} \tag{42.16}$$

$$\frac{\partial z_1}{\partial \pi_2} \frac{\pi_2}{z_1} = \frac{\partial \left(\frac{f_1}{g_1} h_1 \right)}{\partial \pi_2} \frac{\pi_2}{z_1} = \left(\frac{\pi_1 \left(\pi_1^2 - \omega_5 \pi_2^2 \right)}{D^2} \left(-C + \frac{x}{\pi_2} \right) + \frac{\pi_1 \pi_2}{D} \left(-\frac{\omega_4}{\pi_1} - \frac{x}{\pi_2^2} \right) \right) \frac{\pi_2}{z_1} \tag{42.17}$$

$$\frac{\partial z_1}{\partial x} \frac{x}{z_1} = \frac{\partial \left(\frac{f_1}{g_1} h_1 \right)}{\partial x} \frac{x}{z_1} = \frac{\pi_1}{D} \frac{x}{z_1} \tag{42.18}$$

□