December 2006   J.nr. 2612

☎ 33145949 ♪ 151 e-mail lma@akf.dk

**Akf working paper**

## Consumer Evaluation of Environmental and Animal Welfare Labelling

*Af Laura Mørch Andersen*

## Contents

*Preface:*

This working paper is a reprint of a working paper originally published on the homepage of the project "Consumer Demand for organic Foods – Domestic and Foreign Market Perspectives". The overall goal of the project was to assess the long-term potential of demand for organic food. Part of the project focused on identification and assessment of valued product attributes and the evaluation of labelling and other information strategies.

The project was headed by AKF, Institute of Local Government Studies – Denmark. Partners in the project were GfK Denmark, University of Copenhagen, CIRED – France, and Aalborg University. The working paper has been worked out by Laura Mørch Andersen, and commented by Mette Ejrnæs (University of Copenhagen), Thomas Bue Bjørner, Lars Gårn Hansen and Mette Wier, all from AKF.

The project was funded by the Danish Ministry of Food, Agriculture and Fisheries.

AKF, Copenhagen, December 2006,

Laura Mørch Andersen

*Title: Consumer Evaluation of Environmental and Animal Welfare Labelling*

Abstract:

The purpose of this working paper is to estimate marginal willingness to pay for eggs carrying different labels. Among other things these labels indicate environmental features and different levels of animal welfare for the hens that produce the eggs. The data on eggs are part of a very comprehensive panel data set covering household purchases of non-durables during a five year period. For each family a wide range of background characteristics are available. Detailed data on eggs are available during the one year period from July 1999 to June 2000 used in this working paper.

Compared to simple statistics, such as the average market share of different egg types, econometric estimations make it possible to disentangle the effect of labels from the effects of e.g. differences in prices. Discrete models such as the multinomial logit make it particularly simple to estimate the willingness to pay for different characteristics of goods, in this case different labels. Had the purpose been to estimate substitution effects, a continuous model would have been used instead.

The estimations are conducted using the new and flexible Mixed Multinomial Logit model (MMNL) also known as Random Parameter Logit (RPL). Mixed multinomial logit allows heterogeneity among households by letting the parameters of the household utility functions be drawn from a common distribution instead of restricting them to be identical for all households. Estimating the parameters of the *distribution* of the parameters of the utility functions yields not only a measure of the marginal willingness to pay for different types of eggs, but also a measure of the degree of heterogeneity among the households.

The eggs are divided into battery eggs (‘buræg’), barn eggs (‘skrabe æg’), free-range eggs (‘fritgående’) and organic eggs (‘økologiske’) and the marginal willingness to pay for the three last types relative to battery eggs are estimated. The marginal willingness to pay for different types of eggs turns out to vary whit the chain of stores in which the purchase is made. Econometric estimations using store-level data reveals that customers in some stores (e.g. Superbrugsen) are generally willing to pay for labels indicating environmental and animal friendly production methods, while customers in other stores (e.g. Bilka) are reluctant to do so. Combining data from many different stores leads to contra-intuitive results caused by the high level of heterogeneity among customers, prices and variety in the different stores.

It is found that models allowing the consumers’ evaluation of the different labels to vary with background characteristics, such as the geographical location of the household residence, are significantly better than models ignoring background variables completely. Models allowing the evaluation to vary with attitudes, such as attitude to branded goods, are also significantly better than the model ignoring these effects. The effect of the age of the main buyer is, in most cases, not significant.

It is reasonable to expect the value of different labels to vary between households. Animal welfare may be very important to some households, but have little or no value in other households. These differences are perceived as ‘heterogeneity of preferences’ in the econometric model. The labels ‘barn eggs’ and ‘free-range eggs’ mainly indicates increased animal welfare, whereas the ‘organic’ label indicates a more environmentally friendly production as well as a higher level of animal welfare. Some households may also perceive the organic eggs as being healthier than other egg types because the hens are fed with organic feed. The heterogeneity of marginal willingness to pay for organic eggs can therefore be induced by differences in the perception and evaluation of at least three different attributes, whereas the heterogeneity of marginal willingness to pay for barn and free-range eggs is expected to arise only from differences in perception and evaluation of animal welfare. Data supports this hypothesis as the estimated heterogeneity of marginal willingness to pay is generally higher for organic eggs than for the two other egg types.

Several practical problems must be addressed when estimating marginal willingness to pay for different egg-labels. In the data used in this working paper only prices of purchased goods are recorded. This means that the prices of goods that were available but not purchased must be imputed. Another problem is to reveal which types of eggs were available in a given purchase situation. If a given type of eggs is frequently subject to shortage of supply the marginal willingness to pay for this type of eggs may systematically be underestimated. It is therefore important to include shortage of supply in the model. This working paper proposes and utilises solutions to both the imputed prices and the shortage of supply issue, and discuss possible refinements of the methods.

# 1 Introduction

In environmental economics, non-market goods have been valued for decades. According to Hanley et al. (1997) the environmental valuation has particularly been intensified since the Exxon Valdez oil spill off the coast of Alaska in 1989. Environmental goods and other public goods are, by nature, hardly ever exchanged on a real market with observable prices. This problem is overcome by either asking people to state their marginal willingness to pay for the environmental good directly, or by using consumption of market goods to reveal preferences for related non-market goods. Variations of the multinomial logit model have been used extensively in the latter type of valuations.

Achieving the first type of information is expensive and time consuming and the hypothetical questions needed require the construction of a realistic hypothetical market scenario, something that is often very difficult. This study uses the second approach and observations of approximately 24,000 actual purchases made by almost 2,000 Danish families during an entire year, thereby overcoming the problems arising from the hypothetical nature of direct questions. The data is provided by GfK Denmark as part of a large panel data set covering purchases of all types of non-durable goods.

## 1.1 The purpose of the study

The main purpose of this working paper is to estimate marginal willingness to pay for eggs carrying different labels. Among other things these labels indicate environmental features and different levels of animal welfare for the hens that produce the eggs.

The second purpose of the study is to examine the potential and the practical difficulties in exploiting the new and comprehensive data set from GfK to value non-market goods such as animal welfare or environmentally friendly production related to different egg types. Finally, the third purpose is to apply the new and more flexible econometric method called *mixed multinomial logit* on a large panel data set.

## 1.2 The data material

As mentioned above, the data was provided by GfK Denmark. The data on egg purchases is only a small part of a very large data set. The original data set consists of more than five million observations of purchases of over fifty different types of food, made by 3,278 families. Each family has reported to the panel for some period within the four-year period covered by data. For estimation purposes, the data is divided into several subsamples. The data can be divided into subsamples where all purchases are made in the same store aggregate (group of stores), but it can also be divided into nested subsamples consisting of more and more heterogeneous store aggregates. Both types of subsamples will be used in the study. Estimating separately on data from different stores allows comparison of the general behaviour of consumers in different stores.

## 1.3 Econometric method

The mixed multinomial logit allows heterogeneity in the panel by letting the parameters of the individual utility functions be drawn from a common distribution instead of restricting them to be identical for all individuals. Estimating the parameters of the *distribution* of the parameters of the utility functions yields not only a measure of the marginal willingness to pay for different types of eggs, but also a measure of the degree of heterogeneity in the population. In addition, it is less restrictive in the underlying assumptions about substitution behaviour than the conventional multinomial logit, which typically has been used to analyse data with the same structure as the data used in this study. In the recent years the mixed multinomial logit has gained popularity, and is now a serious rival to the conventional multinomial logit. According to Hensher and Greene (2001), page 1, 'The mixed multinomial logit model is considered to be the most promising state of the art discrete model currently available'.

## 1.4 Composition

This study consists of a theoretical part and an empirical part. The theoretical part is divided into a chapter about the economic theory underpinning the random utility model and, not least, about the economic theory behind the concept of 'marginal willingness to pay' that is often applied in empirical valuation studies, and also will be used in this study (chapter 2). This is followed by a chapter about the statistical theory behind the mixed multinomial logit and an introduction to the simulation of the model (chapter 3). The empirical part starts with a presentation of data in chapter 4, and continues with chapter 5, which describes how the economic theory in chapter 2 is put into practice. The solutions presented in chapter 5 are used in the actual estimations. These are presented and discussed in detail for SuperBrugsen in chapter 6 and in chapter 7 results from different chains of stores are compared. Finally, chapter 8 sums up the entire working paper.

## 1.5 Main results

The estimations show that heterogeneity is indeed present in the panel and that the mixed multinomial logit provides a solution to this that is both feasible to estimate and yields interesting interpretations. Estimations also show that part of the heterogeneity must be captured by estimating separately on different subsamples, since estimations using data from different stores lead to very different results. As an example, 60 percent of the customers in Superbrugsen are willing to pay extra for eggs that include higher animal welfare than battery eggs ('buræg'). In Føtex only half of the customers are willing to pay extra, and at the other end of the scale, customers in Bilka generally refuse to pay extra (only 36 percent will pay more).

Summing up, this study combines a new data set that provides huge amounts of complex information, with a new and promising econometric technique. The resulting estimations yield interesting new information about the heterogeneity of preferences for animal welfare related to eggs. This study also considers the practical problems of using market data and proposes solutions to these problems.

# 2 Deriving marginal willingness to pay

Assuming that individuals have rational and continuous preferences means that they can be perceived as utility maximising consumers whose preferences can be represented by a utility function (cf. e.g. Deaton and Muellbauer 1980). The utility function is only defined up to a monotone increasing transformation, which means that the difference between the value of the utility function given that a certain type of egg is chosen and the utility function given that another type of egg was chosen is not a meaningful measure of the utility gained by purchasing one type of egg instead of another. The difference can be used to rank the utility of different eggs (by the ordinal nature of the utility function) but not to say if one egg is twice as good as another. If the utility of eggs is to be compared with utility of other goods, it is necessary to have some kind of cardinal measure of utility. This is usually obtained by combining the utility of, e.g., an egg with the utility of money, leading to utility measured as a monetary value. This measure is called marginal willingness to pay ('wtp').[1]

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 to buy. 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* marginal willingness to pay', but it is often simply referred to as 'marginal willingness to pay', 'willingness to pay' or 'wtp'.

## 2.1 Hanemann's original method for deriving marginal willingness to pay

The most commonly used method for deriving marginal willingness to pay from the utility function was developed by Hanemann (1984). Hanemann uses the Random Utility Model (RUM) in which the individual utility is not perfectly observed by the econometrician. In his article Hanemann wishes to find marginal willingness to pay for the availability of hunting (a hunting permit), the purpose of this study, however, is to estimate marginal willingness to pay for different types of eggs. Eggs will therefore be used in examples in the following, but otherwise the presentation will follow Hanemann (1984).

The individual is assumed to gain utility from a given type of egg. If the egg is not purchased $b$ (for '*b*uy') is 0 and if the egg is purchased $b$ is 1. Hanemann also allows the

[1] If marginal willingness to pay is to be aggregated over consumers it is important to deal with the issue of utility of money. If the utility of money is allowed to vary between consumers marginal willingness to pay can not be aggregated over consumers without taking these differences into account.

utility to depend on the income, $y$, of the individual, $i$, and on other characteristics, $s$, of the individual. (For ease of notation, the subscript $i$ is dropped).

The assumption of weak separability of preferences means that consumption of eggs can be modelled without information about consumption of other goods, and combined with the assumption of homothetic preferences for the different egg types it means that the choice of egg type can be modelled without including the total quantity of eggs purchased. In Hanemann's case he has no information about consumption of any goods except the hunting permit, so he is forced to assume separability, and since the choice is always between one or zero units of the hunting permit the quantity issue does not occur.

If individual $i$ chooses to buy the egg his utility is $u_1 = u(1, y; s)$, if he chooses not to buy the egg his utility is $u_0 = u(0, y; s)$. The crucial assumption in RUM is that the individual knows his own utility function with certainty, but it contains some components that are unobservable to the econometrician, and are treated as stochastic when modelling the utility. These components generate the stochastic structure of the statistical response model. The un-observables could be attributes of the goods from which consumption is chosen, or characteristics of the individual, or a combination of these.

The difference between individual utility functions can be perceived as arising from differences in the (observable as well as unobservable) characteristics of the consumer. In Hanemann (1984), this assumption is expressed by allowing the observable characteristics to be an argument of the utility function.

RUM means that the econometrician assumes that the true utility function $u(b, y; s)$ is a random variable with mean $v(b, y; s)$ which depends on the observable characteristics of the individual through a given parametric function. The true utility can be written as

$$u(b, y; s) = v(b, y; s) + \varepsilon_b, \quad b \in \{0,1\} \tag{2.1}$$

where $\varepsilon_0$ and $\varepsilon_1$ are i.i.d. random variables with zero mean.

Hanemann proposes two examples of functional form of the utility function:

$$v(b, y; s) = \alpha_b + \beta y, \quad \beta > 0, b \in \{0,1\} \tag{2.2}$$

and

$$v(b, y; s) = \alpha_b + \beta \ln(y), \quad \beta > 0, b \in \{0,1\} \tag{2.3}$$

In both cases $\alpha_0, \alpha_1$ and $\beta$ are functions of the observed characteristics, $s$, of the individual, even though this is suppressed in the notation. In this presentation of Hanemann's results the simple version (2.2) is chosen as the utility function.

If the price of one egg is $\pi$, the individual will choose to buy the egg if

$$v(1, y-\pi; s)+\varepsilon_1 > v(0, y; s)+\varepsilon_0 \tag{2.4}$$

and refuse otherwise. Therefore the probability of buying the egg given $\pi$ is

$$\begin{aligned}&\Pr\{i \text{ being willing to pay}|\pi\} = \\ &\Pr\{v(1, y-\pi; s)+\varepsilon_1 > v(0, y; s)+\varepsilon_0\} = \\ &\Pr\{v(1, y-\pi; s)-v(0, y; s) > \varepsilon_0-\varepsilon_1\}\end{aligned} \tag{2.5}$$

Given the utility function in (2.2) the probability of individual *i* being willing to pay $\pi$ for the egg is

$$\begin{aligned}&\Pr\{v(1, y-\pi; s)-v(0, y; s) > \varepsilon_0-\varepsilon_1\} = \\ &\Pr\{\alpha_1+\beta(y-\pi)-(\alpha_0+\beta y) > \varepsilon_0-\varepsilon_1\} = \\ &\Pr\{\alpha_1-\alpha_0-\beta\pi > \varepsilon_0-\varepsilon_1\}\end{aligned} \tag{2.6}$$

The point $\pi^*$ where the probability in (2.6) equals 0.5 (so that the probability of buying equals the probability of not buying), can be perceived as the maximum marginal willingness to pay. At exactly this point the consumer is indifferent between the two options.

Since the utility is random, the marginal willingness to pay is also randomly distributed. The point $\pi^*$ is the median of this distribution. The mean of the distribution is more difficult to calculate and according to Hanemann (1984), page 339, it is also 'very sensitive to slight changes in the distribution resulting from different estimation methods or outliers in the data' while the median $\pi^*$ is 'relatively robust'. The median $\pi^*$ is therefore used in this study.

If the distribution of $\varepsilon_0-\varepsilon_1$ is symmetric around the mean of zero the maximum marginal willingness to pay $\pi^*$ is the point that satisfies

$$\begin{gathered}\alpha_1-\alpha_0-\beta\pi^* = 0, \quad \beta>0 \\ \Updownarrow \\ \pi^* = \frac{\alpha_1-\alpha_0}{\beta}\end{gathered} \tag{2.7}$$

Note that $y$ does not enter the expression for marginal willingness to pay using this particular utility function, whereas it would enter if the functional form in (2.3) was used instead.[2]

In most applications of the random utility model, the functional form of the utility function is assumed to be such that the income cancels out of the expression for marginal willingness to pay. Often it is useful because information about income is unavailable, but in this particular case income could actually be measured either by the discrete value of income that is reported once a year (total household income in classes of 50,000 DKK) or by the total consumption of non-durables that can be derived from the purchase data. In the present study the conventional method is used and the effect of income is left out, but in another study it could be interesting to investigate the functional form of the utility function further, and in that connection it would be natural to examine whether income ought to enter the expression for marginal willingness to pay.

The important result of Hanemann's analysis is that assuming separability and restricting the utility function to having a very simple functional form makes it possible to estimate one single number (or as will be seen later, a distribution from which this number is drawn) that has an intuitively easy interpretation, and that is also relevant in the usual theoretical economic framework.

Both assumptions can of course be argued against, but Hanemann (1984) does not exclude the possibility of a more complicated utility function, his analysis simply shows that the derived marginal willingness to pay becomes less complicated and less difficult to interpret intuitively when using this particular functional form.

Hanemann's model can be extended to cover more complicated cases; a few of these will be discussed in the following section.

## 2.2 Extensions of the model

First, the alternative to buying a special type of egg or buying the hunting permit may not be to buy nothing at the price of zero. Most choice models are conditioned on the fact that a purchase is made. An example is travel cost models, where the expenses (in time as well as money) are used as an expression of preferences for different attributes of the sites in the choice set from which the individual chooses a destination, conditioned on the fact that a trip is taken. Discrete choice models are, in general, good at describing choices given a well-defined choice-set. Modelling consumption as such (including the decision to take a trip to the beach or make a dinner that requires eggs) means that the discrete choice models must be integrated with a completely different

---

[2] If the utility function has the form $v(b,y;s)=\alpha_b+\beta\ln(y),\ \ \beta>0, b\in\{0,1\}$, the derived marginal willingness to pay becomes $\alpha_1-\alpha_0-\beta\ln\left(1-\frac{\pi^*}{y}\right)\approx\alpha_1-\alpha_0-\beta\frac{\pi^*}{y}=0,\quad \beta>0\ \Leftrightarrow \pi^*=\frac{\alpha_1-\alpha_0}{\beta}y$, which depends on the income, $y$.

model, something that is hardly ever done because modelling the choice itself can be troublesome enough.

According to Bockstael (1995), Morey et al. (1991) have proposed a model that allows joint estimation of the number of fishing trips and the choice of site. The model depends on the assumption that a season can be divided into *T* decision periods within which at most one trip can be taken. The model combines a binomial distribution on number of 'successes' in *T* trials with a multinomial logit model of site choice. The model is developed for the case where the number of trips (successes) is known, but the destinations of only a subset of trips is available.

When modelling the choice behaviour conditional on the fact that a purchase of some kind is made, the price of the alternative is not zero as in the model of Hanemann (1984). If the choice is between two goods with costs $\pi_1$ and $\pi_2$, this changes (2.6) to

$$\begin{aligned}
&\Pr\left\{v(1, y-\pi_2; s) - v(0, y-\pi_1; s) > \varepsilon_1 - \varepsilon_2\right\} = \\
&\Pr\left\{\alpha_2 + \beta(y-\pi_2) - (\alpha_1 + \beta(y-\pi_1)) > \varepsilon_1 - \varepsilon_2\right\} = \\
&\Pr\left\{\alpha_2 - \alpha_1 - \beta(\pi_2 - \pi_1) > \varepsilon_1 - \varepsilon_2\right\}
\end{aligned} \tag{2.8}$$

and (2.7) to

$$\begin{gathered}
\alpha_2 - \alpha_1 - \beta(\pi_2 - \pi_1)^* = 0, \quad \beta > 0 \\
\Updownarrow \\
(\pi_2 - \pi_1)^* = \frac{\alpha_2 - \alpha_1}{\beta}
\end{gathered} \tag{2.9}$$

This means that marginal willingness to pay measures the amount that a person is willing to pay extra for good 2 compared to good 1.

Hanemann's model can also be extended by allowing repeated choices. Hanemann investigates the case in which the decision about the permit is made only once. (Of course it is probably made once a year in real life, but only once in the data used by Hanemann). If each individual makes repeated choices it becomes a key assumption that the utility function (represented by the $\alpha$'s and the $\beta$) is constant over time or choice occasion. The utility of the different types of eggs should, therefore, be the same every time an egg is purchased, and the utility of one monetary unit should also be the same.

The expression for marginal willingness to pay in (2.7) and (2.9) has a very nice intuitive interpretation. $\alpha_2 - \alpha_1$ is the utility gained by buying one egg of type 2 instead of one egg of type 1, and $\beta$ is the utility of money. Dividing $\alpha_2 - \alpha_1$ by $\beta$ therefore converts the utility measure to the monetary measure mentioned earlier in this chapter. This is necessary to be able to compare marginal willingness to pay for different persons and for different goods.

## 2.3 Conclusion

As can be seen from the examples in this section, Hanemann's method for deriving the median of the marginal willingness to pay can be used in many different settings, and it is the method used in e.g. Layerton and Brown (2000), Revelt and Train (1998) and Rouwendal and Meijer (2001). It will also be the method used in the present study.

# 3 The statistical model

Compared to simple statistics, such as the average market share of different egg types, econometric estimations make it possible to disentangle the effect of labels from the effects of e.g. differences in prices. Discrete models such as the multinomial logit make it particularly simple to estimate the marginal willingness to pay for different characteristics of goods, in this case different labels. Had the purpose been to estimate substitution effects, a continuous model would have been used instead.

The estimations are conducted using the new and flexible Mixed Multinomial Logit model (MMNL) also known as Random Parameter Logit (RPL). Mixed multinomial logit allows heterogeneity among households by letting the parameters of the household utility functions be drawn from a common distribution instead of restricting them to be identical for all households. Estimating the parameters of the *distribution* of the parameters of the utility functions yields not only a measure of the marginal willingness to pay for different types of eggs, but also a measure of the degree of heterogeneity among the households.

This chapter describes the statistical model that will be used to estimate marginal willingness to pay as introduced in chapter 2. First the conventional multinomial logit is presented briefly in section 3.1. Then section 3.2 introduces the main problem (which in some cases is an advantage) of the logit model, namely the Independence of Irrelevant Alternatives (IIA) that is a fundamental part of logit models. One way of solving this problem is to use a *nested logit*; a concept that is explained and exemplified in section 3.3. It is concluded that nesting is not the solution in this study and the mixed multinomial logit model is therefore introduced in section 3.4.

Section 3.4.1 discusses the implications of the assumptions Train (1998) makes about the behaviour of the individual in the model. The mixed multinomial logit model is generally defined in section 3.4.2 and compared with the conventional multinomial logit model in section 3.4.3.

In this study a program, developed at University of California, Berkeley (Train et al. 1999) is used to conduct the estimation of the MMNL model. The program may yield a better understanding of the model than the general definition in section 3.4.2, and a brief introduction to Train's version of the model is therefore given in section 3.5, with special focus on the likelihood function, and the way it is simulated (section 3.5.2).

Section 3.6 discusses the implications of the fact that individual parameters in MMNL are drawn from a common distribution, and concludes that it can be used to make interesting statements about the size of different segments of the population (in this case the households in the panel).

The difference between logit and probit and between mixed and conventional models are briefly summarised in section 3.7, and section 3.8 emphasises that in this study the most important difference between a conventional logit and a mixed logit is the fact that estimated marginal willingness to pay is not assumed to be the same for all individuals when using the mixed model, but rather to follow a distribution defined by the econometrician.

Section 3.9 sums up this chapter and concludes that the mixed multinomial logit will allow estimation of more interesting features of marginal willingness to pay than conventional discrete models.

## 3.1 The conventional multinomial logit

The data used in this study come from an unbalanced panel of families who are followed over time. Panel data combines time series (being able to follow the consumption of an individual/a group of individuals over time) and cross section data (knowing the purchases of many individuals at a given time). In this application there are therefore three dimensions: Individual $i$, $i$=1,...,$N$, time $t$, $t$=1,...,$T_i$ and egg type $j$, $j$=1,...,$J$. Many panel models only include two dimensions: Individual and time, and the type of egg chosen would simply be the observed value of egg type for individual $i$ at time $t$. But in this case for each variable, $x$, information is needed about the value of $x$ for individual $i$ at time $t$, *for each possible choice of egg type j*. Assuming that there are $N$ individuals with $J$ possible choices in each of the $T_i$ observations, $x$ does not consist of $\sum_{i=1}^{N} T_i$ numbers (or $N \cdot T$ if all individuals had the same $T_i$) but of $J \cdot \sum_{i=1}^{N} T_i$ numbers.

The Random Utility Model (RUM) mentioned in chapter 2, is the base assumption in the logit model. In RUM the utility of choice $j$ is assumed to be $U_j^i = \beta_j' x_{ijt} + \varepsilon_{ijt}$. This means that the probability of $y_{it} = j$ is equal to

$$\begin{aligned} prob\left(y_{it} = j\right) &= prob\left(U_j^i > U_k^i, \forall k \neq j\right) \\ &= prob\left(\beta_j' x_{ijt} + \varepsilon_{ijt} > \beta_k' x_{ijt} + \varepsilon_{ikt}, \forall k \neq j\right) \\ &= prob\left(\left(\beta_j' - \beta_k'\right) x_{ijt} > \varepsilon_{ikt} - \varepsilon_{ijt}, \forall k \neq j\right) \end{aligned} \tag{3.1}$$

disregarding the case of $U_j^i = U_k^i$.

McFadden (1973) proved that letting $\varepsilon_{ijt}$ follow a Weibull distribution:[1]

$$prob(\varepsilon_{ijt} \leq \kappa) = \exp\left(-\exp\left(-\kappa\right)\right) \tag{3.2}$$

leads to the simple result

[1] Ben-Akiva and Lerman (1985) describe this as the *Gumbel* distribution. If $\varepsilon$ is Gumbel distributed with location parameter $\eta$ and scale parameter $\mu > 0$, then the cumulative density function is $F(\varepsilon) = \exp(-\exp[-\mu(\varepsilon-\eta)])$ and the density function is $f(\varepsilon) = \mu \cdot \exp[-\mu(\varepsilon-\eta)] \cdot \exp(-\exp[-\mu(\varepsilon-\eta)])$. This distribution is often referred to as the *extreme value* distribution.

$$prob\left(y_{it} = j\right) = \frac{\exp\left(\beta_j' x_{ijt}\right)}{\sum_{k=1}^{J} \exp\left(\beta_k' x_{ijt}\right)} \tag{3.3}$$

This is the general formula for the multinomial logit model.

In some cases the value of $x_{ijt}$ is the same for all $j$ (e.g. for information about the individual) but in other cases the value of $j$ is very important (e.g. for information about price). This has important implications for the logit modelling. To illustrate this, let $x_{ijt} = [z_{ijt}, w_{it}]$, where $z_{ijt}$ varies between choices, and possibly also between individuals and time. An example could be prices that vary between egg types, but could also vary between individuals and over time. $z_{ijt}$ is called the *attributes* of the choices. $w_{it}$ varies only across individuals (and possibly over time), and is called the *characteristics* of the individual. Examples could be gender, age or geographical location of residence. This implies that it is also necessary to split up $\beta_j$ so that $\beta_j' = \left[\beta_j'^z, \beta_j'^w\right]$ and $\beta_j' x_{ijt} = \beta_j'^z z_{ijt} + \beta_j'^w w_{it}$.

If all $\beta$'s are expected to be choice specific (depend on $j$) the model is called a *generalised* multinomial logit model.

### *3.1.1 The generalised multinomial logit model*

In the generalised multinomial logit model the probability that individual $i$ chooses $j$ at time $t$ is:

$$prob\left(y_{it} = j\right) = \frac{\exp\left(\beta_j' x_{ijt}\right)}{\sum_{k=1}^{J} \exp\left(\beta_k' x_{ikt}\right)} = \frac{\exp\left(\beta_j'^z z_{ijt} + \beta_j'^w w_{it}\right)}{\sum_{k=1}^{J} \exp\left(\beta_j'^z z_{ikt} + \beta_j'^w w_{it}\right)} \tag{3.4}$$

This model is usually (but not always) used on data where the explanatory variables vary across *individuals*, since the individual specific explanatory variables do not cancel out as long as $\beta$ depends on the different choices.

If on the contrary all $\beta$'s are expected to be choice invariant (independent of $j$) the model changes to:

### *3.1.2 The conditional multinomial logit model*

In the special case where all parameters are choice invariant, the probability that individual $i$ chooses $j$ at time $t$ is:

$$\begin{aligned} prob\left(y_{it} = j\right) &= \frac{\exp\left(\beta' x_{ijt}\right)}{\sum_{k=1}^{J} \exp\left(\beta' x_{ijt}\right)} = \frac{\exp\left(\beta'^z z_{ijt} + \beta'^w w_{it}\right)}{\sum_{k=1}^{J} \exp\left(\beta'^z z_{ikt} + \beta'^w w_{it}\right)} \\ &= \frac{\exp\left(\beta'^w w_{it}\right)\exp\left(\beta'^z z_{ijt}\right)}{\exp\left(\beta'^w w_{it}\right)\sum_{k=1}^{J} \exp\left(\beta'^z z_{ikt}\right)} = \frac{\exp\left(\beta'^z z_{ijt}\right)}{\sum_{k=1}^{J} \exp\left(\beta'^z z_{ikt}\right)} \end{aligned} \tag{3.5}$$

As can be seen in (3.5) the characteristics of the individuals cancel out of the conditional multinomial logit model because the utility is assumed to be linear in $x_{ijt}$, and the $\beta$'s are assumed to be choice invariant. The model is therefore usually (but not always) used on data where the explanatory variables vary between choices (in this application, *egg types*).

As mentioned in Greene (2000), it is, however, possible to estimate effects of individual specific characteristics if these characteristics are multiplied by a set of dummy variables for the choices. A complete set of interaction terms creates a singularity, therefore, one of the interactions must be dropped.

An example could be the geographical location of the individual residence. This variable is clearly independent of choice. If all exogenous variables are crossed with dummies for choices, the reaction to these exogenous variables will be different for each potential choice, leading to the generalised multinomial logit defined above.

If only exogenous variables that do not vary over choices are crossed with dummies for these choices, the effects of these variables will be choice specific while the effects of variables that do vary over choices are choice invariant.

As an example one could look at a model where the choice of egg type is explained by the prices of the different egg types and the geographical location of the individual residence. Then price can enter directly since it varies over egg types, but geography must be crossed with dummies for egg type. The model will result in one parameter for price, but a set of parameters for each type of egg for each geographical category. These parameters will vary over egg types as well as over geography, making them choice specific.

It is also possible to cross an individual specific variable with any other choice specific variable. Dummies for choices are just one possibility (but of course the only one that leads to choice specific parameters). In the example above geography could be crossed with price if it was expected that all individuals respond identically to egg types, but have different reactions to price (perhaps people from Jutland react more strongly to increases in prices than people on Zealand). That would allow geography to enter the model, but would not produce any choice specific parameters.

The generalised and the conditional multinomial logit can therefore be seen as two extreme cases of the multinomial logit model, and often the models estimated will be a combination of the two extremes.

## 3.2 Independence of Irrelevant Alternatives (IIA) in the multinomial logit model

One of the attractive features of the multinomial logit model is that the estimated parameters can easily be used to make statements about the probability of choosing $j$ in stead of $k$. This relationship is measured as the '*odds ratio*':

$$\frac{P_{ijt}}{P_{ikt}} = \frac{prob\left(y_{it} = j\right)}{prob\left(y_{it} = k\right)} \tag{3.6}$$

As will be seen, the multinomial logit model makes it easy to interpret the log of the odds ratio.

Returning to the notation $x_{ijt} = [z_{ijt}, w_{it}]$ defined above, the odds ratio, $P_{ijt}/P_{ikt}$, can, in general, be calculated in the following way:

$$\begin{aligned}
\frac{prob(y_{it} = j)}{prob(y_{it} = k)} &= \frac{\exp\left(\beta_j' x_{ijt}\right) \Big/ \sum_{l=1}^{J} \exp\left(\beta_l' x_{ilt}\right)}{\exp\left(\beta_k' x_{ikt}\right) \Big/ \sum_{l=1}^{J} \exp\left(\beta_l' x_{ilt}\right)} \\
&= \frac{\exp\left(\beta_j' x_{ijt}\right)}{\exp\left(\beta_k' x_{ikt}\right)} \\
&= \frac{\exp\left(\beta_j'^{z} z_{ijt}\right)\exp\left(\beta_j'^{w} w_{it}\right)}{\exp\left(\beta_k'^{z} z_{ikt}\right)\exp\left(\beta_k'^{w} w_{it}\right)} \\
&= \exp\left(\beta_j'^{z} z_{ijt} + \beta_j'^{w} w_{it} - \beta_k'^{z} z_{ikt} - \beta_k'^{w} w_{it}\right) \\
&= \exp\left(\beta_j'^{z} z_{ijt} - \beta_k'^{z} z_{ikt} + \left(\beta_j'^{w} - \beta_k'^{w}\right) w_{it}\right)
\end{aligned} \tag{3.7}$$

In the conditional multinomial logit model the $\beta$'s are assumed to be fixed across choices leading to:

$$\begin{aligned}
\frac{prob(y_{it} = j)}{prob(y_{it} = k)} &= \exp\left(\beta'^{z} z_{ijt} - \beta'^{z} z_{ikt} + \left(\beta'^{w} - \beta'^{w}\right) w_{it}\right) \\
&= \exp\left(\beta'^{z} \left(z_{ijt} - z_{ikt}\right) + 0 \cdot w_{it}\right) \\
&= \exp\left(\beta'^{z} \left(z_{ijt} - z_{ikt}\right)\right)
\end{aligned} \tag{3.8}$$

Again the $w_{it}$'s cancel out in this model as described in the definition of the conditional multinomial logit in 3.1.2.

In the generalised multinomial logit the $\beta$'s are allowed to vary between choices leading to

$$\frac{prob(y_{it} = j)}{prob(y_{it} = k)} = \exp\left(\beta_j'^{z} z_{ijt} - \beta_k'^{z} z_{ikt} + \left(\beta_j'^{w} - \beta_k'^{w}\right) w_{it}\right) \tag{3.9}$$

Often the model is used on data that only vary across individuals, and not across choices. This special case leads to:

$$\frac{prob(y_{it} = j)}{prob(y_{it} = k)} = \exp\left(\left(\beta_j'^{w} - \beta_k'^{w}\right) w_{it}\right) \tag{3.10}$$

In all cases the odds ratio is independent of the other alternatives. This is called Independence of Irrelevant Alternatives (IIA).

The IIA makes it easier to interpret the model since the change in the odds ratio as a result of a change in the exogenous variable, $z_{ijt}$, for choice *j* (e.g., an increase in the price of egg type *j*) becomes

$$\begin{aligned}\frac{\partial\left(\frac{prob(y_{it}=j)}{prob(y_{it}=k)}\right)}{\partial z_{ijt}} &= \frac{\partial \exp\left(\beta_j'^z z_{ijt} - \beta_k'^z z_{ikt} + \left(\beta_j'^w - \beta_k'^w\right) w_{it}\right)}{\partial z_{ijt}} \\ &= \beta_j'^z \exp\left(\beta_j'^z z_{ijt} - \beta_k'^z z_{ikt} + \left(\beta_j'^w - \beta_k'^w\right) w_{it}\right) \\ &= \beta_j'^z \frac{prob(y_{it}=j)}{prob(y_{it}=k)}\end{aligned} \tag{3.11}$$

and the change in the odds ratio as result of a change in the exogenous variable, $z_{ikt}$, for choice *k* (e.g., increase in the price of egg type *k*) becomes

$$\begin{aligned}\frac{\partial\left(\frac{prob(y_{it}=j)}{prob(y_{it}=k)}\right)}{\partial z_{ikt}} &= \frac{\partial \exp\left(\beta_j'^z z_{ijt} - \beta_k'^z z_{ikt} + \left(\beta_j'^w - \beta_k'^w\right) w_{it}\right)}{\partial z_{ijt}} \\ &= -\beta_k'^z \frac{prob(y_{it}=j)}{prob(y_{it}=k)}\end{aligned} \tag{3.12}$$

Note that if $\beta_j'^z \neq \beta_k'^z$ then increasing $z_{ijt}$ by one does not give the same effect as decreasing $z_{ikt}$ by one. In the conditional multinomial logit model the $\beta$'s are independent of the choices, therefore the effects are always symmetric.

The effect of changes in $w_{it}$ is of course only defined in the generalised multinomial logit model:

$$\frac{\partial\left(\frac{prob(y_{it}=j)}{prob(y_{it}=k)}\right)}{\partial w_{it}} = \left(\beta_j'^w - \beta_k'^w\right)\frac{prob(y_{it}=j)}{prob(y_{it}=k)} \tag{3.13}$$

IIA is reasonable in some cases, but more often it causes problems. The classical example is transportation. Imagine that 2/3 of the population drive their own car to work and 1/3 take a bus called ‘bus 1’. The ratio of car drivers versus bus users is therefore 2:1. Now a new bus company enters the market. The natural assumption would, of course, be that the people who drive a car continue to use the car (at least most of them) and a proportion of the bus 1 users switch to the new company. But in the multinomial logit model the odds ratio of car versus bus 1 should still be two, meaning that the percentage of people driving their own car should decrease proportionally to the decrease in the percentage of people using bus 1. If ten percent of the bus 1 users shift to bus 2, so should ten percent of the car users. This seems unrealistic.

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, e.g., organic eggs. This is very unlikely. Imagine that organic eggs leave the market. Then the IIA in the multinomial logit model would imply that the people who used to buy organic eggs would distribute themselves between the rest of the egg types according to the market share of these other types of eggs. But people who buy organic eggs may very well have a higher propensity to buy, e.g., free-range eggs than the population in general and, in particular, have a lower propensity to buy battery eggs. IIA is therefore not reasonable in this case.

One way of avoiding IIA is to use a probit model, but as will be explained further in section 3.7, the probit requires more computation time and the multinomial logit model is therefore often considered more attractive.

## 3.3 Nested multinomial logit

Another way of avoiding IIA is by using nested multinomial logit. Sometimes it is possible to split the decision into several stages. A classical example is (again) transport. If a person wants to cross America from coast to coast it could be argued that he first chooses whether he wants to fly or travel on the ground. If he chooses to fly, he might have an additional choice between different air companies, if he chooses to travel on the ground, he might have to choose between car, bus and train.

This structure can be used in a nested multinomial logit. First a binomial logit is estimated with a choice set $C_1$={air,ground}. In cases with many individuals it is most likely that both elements are chosen by one or more persons. The next step is therefore to estimate two multinomial logit models with choice sets $C_{21}$={air$_1$, air$_2$, air$_3$} and $C_{22}$={car,bus,train}. Now IIA is present within each choice set $C_1$, $C_{21}$ and $C_{22}$ but not between the choice sets.

The problem with this method is that it is often not perfectly clear how to split up the decision process. In this study, IIA could be eliminated by letting the consumer choose between the alternative with least animal welfare and the 'rest' of the egg types leading to

$$\begin{array}{llcl}
C_1: & & & C_1 = \{battery, \{barn,\ free-range, organic\}\} \\
 & & \swarrow & \quad\downarrow \\
C_2: & C_{21} = \{battery\} & & C_{22} = \{barn, \{free-range, organic\}\} \\
 & & \swarrow & \quad\downarrow \\
C_3: & C_{31} = \{barn\} & & C_{32} = \{free-range, organic\} \\
 & & \swarrow & \quad\downarrow \\
C_4: & C_{41} = \{free-range\} & & C_{42} = \{organic\}
\end{array} \tag{3.14}$$

corresponding to a choice experiment where the consumer chooses between a lower and a higher level of animal welfare, with the lowest level of animal welfare increasing for each step until he reaches the optimal level of animal welfare. This experiment could be performed just as well by choosing between the best level of animal welfare and the other types, that is, by decreasing the highest level of animal welfare in each step. This would lead to a different nesting, with $C_1 = \{organic, \{free-range, barn, battery\}\}$ and $C_{41} = \{barn\}, C_{42} = \{battery\}$. The choice of nesting structure is therefore not unambiguous. If the choice set consists of four elements, the number of possible nesting structures[2] is $4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$. It is not feasible in practice to estimate all of these models, and even if it was, there would be no obvious way of choosing the best model. Nesting requires a well-founded nesting structure, based on behavioural arguments, and that is not possible in this case. It is therefore necessary to find another way of avoiding IIA.

### 3.4 The mixed multinomial logit[3]

In the recent years a new multinomial logit model has been introduced. In Hensher and Greene (2001), page 1, it is stated that 'The mixed multinomial logit model is considered to be the most promising state of the art discrete model currently available.' According to Revelt and Train (1998) the Mixed MultiNomial logit (henceforward: MMNL) model was introduced in 1980 by Boyd and Mellman (1980) and Cardell and Dunbar (1980), but these applications did not allow variation in explanatory variables across individuals. Thanks to advances in computer speed and understanding of simulation methods for approximating integrals, models that allow explanatory variables to vary over consumers have been developed: Ben-Akiva et al. (1993), Ben-Akiva and Bolduc (1996), Bhat (1996) and Brownstone and Train (1999). However these models did not allow repeated choices. The version used in the present study allows for repeated choices for different individuals and was used for the first time in Revelt and Train (1998). The model is described thoroughly in McFadden and Train (2000).

Today there are at least two programs available for estimating a MMNL model. Limdep™ version 7.0.2 allows for mixing of parameters in a multinomial logit model, but does not exploit the extra information given by a panel structure. Since one of the advantages of the data available in this study is the panel structure, Limdep is used only as a way of controlling the results obtained using the GAUSS program developed by Train et al. 1999 (section 3.5 will briefly describe the GAUSS program). This is the program that is used in Revelt and Train (1998) and Train (1998), and it allows for the panel structure. It does not make direct use of the time perspective, but treats the observations as repeated choices instead. The number of observations are allowed to

[2] In the special case of one type of egg versus the remaining egg types in each step.

[3] The model is also known as the random parameter logit (RPL) or the random coefficient multinomial logit

vary from individual to individual, which is fortunate since the panel used in this study is highly unbalanced.

### 3.4.1 Behavioural assumptions in the mixed multinomial logit

In Train (1998) the utility function in the random utility model underlying the MMNL model is presented as

$$U_{it}^{j} = \beta_i' x_{ijt} + \varepsilon_{ijt} = \left(b' + \eta_i'\right) x_{ijt} + \varepsilon_{ijt} = b' x_{ijt} + \eta_i' x_{ijt} + \varepsilon_{ijt} \tag{3.15}$$

where the individual $\beta_i$ is decomposed into a part, *b*, that is common for all individuals (the mean of the distribution of individual $\beta_i$'s) and an individual part, $\eta_i$, that differs between individuals and has mean zero in order to separate the effect of *b* from the effect of $\eta_i$.

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

$$\xi_{ijt} = \eta_i x_{ijt} + \varepsilon_{ijt} \tag{3.16}$$

that are correlated over choice (*j*) and time (*t*) for individual *i* because of the common influence of $\eta_i$. Train uses the fact that the errors are correlated over choices to explain that IIA is eliminated in the MMNL model (Train 1998). Intuitively this explanation makes sense. Using eggs as an example, the differences in taste (preferences) make the probability of choosing each egg type correlated for individual *i*. This means that the individuals that have preferences different from the mean of the population ($\eta_i \neq 0$) will not distribute their consumption according to the average distribution and will therefore not substitute according to this average distribution.

The individual is assumed to know not only his own $\beta_i$ but also the value of $\varepsilon_{ijt}$ for all *j*'s at time *t*. The individual is therefore not subject to any randomness when deciding, for example, which type of egg to choose. Given the $\varepsilon_{ijt}$'s, one type of eggs will always be preferred to all others and this type of egg will be chosen. Tomorrow the $\varepsilon_{ijt+1}$'s may well be different from the $\varepsilon_{ijt}$'s today leading the consumer to choose another type of egg with certainty given the values of the $\varepsilon_{ijt+1}$'s. If one type of egg is removed, another type of egg shifts to the top of the ranking with certainty, so the individual does not follow any kind of 'individual IIA'.

The assumption that the individual knows his own utility function perfectly is the same under conventional multinomial logit and under mixed multinomial logit. The difference lies in the assumptions about the error terms not observed by the researcher. Under the conventional multinomial logit the utility function is assumed to be

$$U_{ijt} = \beta' x_{ijt} + \varepsilon_{ijt} \tag{3.17}$$

with identical $\beta$'s for all individuals and i.i.d. extreme value error terms $\varepsilon_{ijt}$. The fact that the error terms are independent over individuals *i*, egg types *j* and time *t* creates IIA. If an individual has a high $\varepsilon_{ijt}$ for *j* it does not give any information about the value of $\varepsilon_{ikt}$ for egg type *k*. All the researcher knows is that the distribution of $\varepsilon_{ijt}$'s in the entire society creates the observed probabilities of choosing the different types of eggs. This leads to IIA.

The difference between conventional and mixed multinomial logit lies, therefore, *not* in the assumptions about the individual information level (which is assumed to be perfect information in both cases) but in the assumptions made on the error terms not observed by the econometrician. In conventional multinomial logit the error terms are assumed to be independent of each other, in the mixed multinomial logit they are allowed to be correlated.

Since the econometrician does not know the individual $\beta_i$, but only the distribution from which it is drawn, he integrates the conventional multinomial logit model over all possible values of $\beta_i$ to ascribe the best possible probabilities to the different choices for each individual. The individual likelihood function therefore becomes the conventional likelihood function of the conditional multinomial logit integrated over $\beta_i$ following the pre-defined distribution. This will be elaborated further in the following.

### *3.4.2 A general definition of the mixed multinomial logit*

In McFadden and Train (2000) MMNL is defined very generally, without mentioning the panel issue al all. The model is defined as a multinomial logit model with random coefficients, $\beta$, drawn from a cumulative distribution function $G(\beta,\theta)$:

$$\begin{aligned} P_C\left(j \middle| x,\theta\right) &= \int L_C^{cl}\left(j;x,\beta\right)\cdot G\left(d\beta;\theta\right) \\ &= \int \frac{e^{\beta' x_j}}{\sum_{l\in C} e^{\beta' x_l}}\cdot G\left(d\beta;\theta\right) \end{aligned} \tag{3.18}$$

where *C* is the choice set, *C* = {1,...,*J*}; $x_j$ is a 1×*M* vector of functions of observed attributes of alternative *j* and observed characteristics of the individuals *i*, with *x* = ($x_1$, $x_2$,..., $x_J$); $\beta$ is a *M*×1 vector of random parameters; $L_C^{cl}\left(j;x,\beta\right)$ is the likelihood value of the conventional multinomial logit model[4] for the choice set *C*; and *θ* is a vector of deep parameters in the mixing distribution *G*. The number of possible choices is therefore *J*, and the number of exogenous variables is *M*.

[4] It is important to note that the likelihood function, $L_C^{cl}\left(j;x,\beta\right)$, is only *part* of the likelihood function in the MMNL model. The superscript '*cl*' is therefore added to the notation, as an abbreviation of ***c***onditional ***l***ogit.

The present study uses panel data and the model must therefore be adjusted to take account of individuals with repeated choices. Under the assumption that the choice probabilities are independent over time for a given individual, the individual likelihood function of a conventional multinomial logit is the product of the probabilities of all the choices made by this individual:[5]

$$L_{iC}^{cl}\left(y_i, x_i, \beta\right) = P\left(y_i \middle| x_i, \beta\right) = \prod_{t=1}^{T_i} P\left(Y_{it} = y_{it} \middle| x_{ijt}, \beta\right) \tag{3.19}$$

This means that the individual likelihood function in the MMNL model with repeated choices is:

$$\begin{aligned} L_i\left(y_i, x_i, \theta\right) &= \int L_{iC}^{cl}\left(y_i, x_i, \beta\right) \cdot G\left(d\beta; \theta\right) \\ &= \int \left[ \prod_{t=1}^{T_i} P\left(Y_{it} = j' \middle| x_{ijt}, \beta\right) \right] \cdot G\left(d\beta; \theta\right) \\ &= \int \left[ \prod_{t=1}^{T_i} \frac{e^{\beta' x_{ij't}}}{\sum_{l \in C} e^{\beta' x_{ilt}}} \right] \cdot G\left(d\beta; \theta\right) \end{aligned} \tag{3.20}$$

where *j'* is the choice actually made by individual *i* at time *t*. Note that the $\beta$'s are assumed to be constant over repeated choices for each individual, since the integral is taken over the product of probabilities of the realised choices. The individual preferences are therefore assumed to be constant over time.

In a model with more than one individual, the likelihood function is the product of the individual likelihood functions, (if the likelihood values of different individuals are assumed to be independent) so the likelihood function that should be maximised in the MMNL model is:

$$L\left(y, x, \theta\right) = \prod_{i=1}^{N} L_i\left(y_{i.}, x_{i..}, \theta\right) = \prod_{i=1}^{N} \int \left[ \prod_{t=1}^{T_i} \frac{e^{\beta' x_{ij't}}}{\sum_{l \in C} e^{\beta' x_{ilt}}} \right] \cdot G\left(d\beta; \theta\right) \tag{3.21}$$

As usual it is easier to look at the log of the likelihood function, which in this case is

$$\begin{aligned} \ln L\left(y, x, \theta\right) &= \ln\left( \prod_{i=1}^{N} L_i\left(y_{i.}, x_{i..}, \theta\right) \right) \\ &= \sum_{i=1}^{N} \ln\left( L_i\left(y_{i.}, x_{i..}, \theta\right) \right) \\ &= \sum_{i=1}^{N} \ln\left( \int \left[ \prod_{t=1}^{T_i} \frac{e^{\beta' x_{ij't}}}{\sum_{l \in C} e^{\beta' x_{ilt}}} \right] \cdot G\left(d\beta; \theta\right) \right) \end{aligned} \tag{3.22}$$

[5] In general, the notation $x_{it} = x_{ijt} \forall j \in 1, \ldots, J$ will be used to describe the part of the matrix *x* where *i* and *t* are fixed for all values of *j*.

Given this definition of MMNL it is now possible to investigate IIA in the mixed multinomial logit.

### *3.4.3 IIA in the mixed multinomial logit*

If all $\beta$'s are fixed[6] (follow a one-point distribution, $\beta = \theta$), the model collapses to a conventional multinomial logit with all the advantages and disadvantages of this model (including IIA). But what happens to IIA if $\beta$ consists of a mixture of fixed and random coefficients?

Let $\beta' = \left[\beta'^{fixed}, \beta'^{random}\right]$, where $\beta^{fixed}$ follows a one-point distribution ($\eta_i^{fixed} = 0$ in equation (3.15)) and $\beta^{random}$ does not ($\eta_i^{random} \neq 0$), and let $x_{ijt} = \left[x_{ijt}^{fixed}, x_{ijt}^{random}\right]$. Since the $\beta$'s are independent of time, the probability that $y_{it} = j$ is the same for all $t$:

$$\begin{aligned} prob\left(Y_i = j \middle| x_i, \theta\right) &= \int \left[\prod_{t=1}^{T_i} \frac{e^{\beta' x_{ijt}}}{\sum_{l \in C} e^{\beta' x_{ilt}}}\right] \cdot G\left(d\beta; \theta\right) \\ &= \int \left[\prod_{t=1}^{T_i} \frac{\exp\left(\beta'^{fixed} x_{ijt}^{fixed} + \beta'^{random} x_{ijt}^{random}\right)}{\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed} + \beta'^{random} x_{ilt}^{random}\right)}\right] \cdot G\left(d\beta^{random}; \theta\right) \\ &= \int \left[\prod_{t=1}^{T_i} \frac{\exp\left(\beta'^{fixed} x_{ijt}^{fixed}\right) \exp\left(\beta'^{random} x_{ijt}^{random}\right)}{\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right)}\right] \cdot G\left(d\beta^{random}; \theta\right) \end{aligned} \tag{3.23}$$

This means that the ratio of two probabilities (the odds ratio) is now:

$$\begin{aligned} \frac{prob(y_{it} = j)}{prob(y_{it} = k)} &= \frac{\int \left[\prod_{t=1}^{T_i} \frac{\exp\left(\beta'^{fixed} x_{ijt}^{fixed}\right) \exp\left(\beta'^{random} x_{ijt}^{random}\right)}{\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right)}\right] \cdot G\left(d\beta^{random}; \theta\right)}{\int \left[\prod_{t=1}^{T_i} \frac{\exp\left(\beta'^{fixed} x_{ikt}^{fixed}\right) \exp\left(\beta'^{random} x_{ikt}^{random}\right)}{\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right)}\right] \cdot G\left(d\beta^{random}; \theta\right)} \\ &= \frac{\int \left[\frac{\prod_{i=1}^{T_i} \left(\exp\left(\beta'^{fixed} x_{ijt}^{fixed}\right) \exp\left(\beta'^{random} x_{ijt}^{random}\right)\right)}{\prod_{i=1}^{T_i} \left(\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right)\right)}\right] \cdot G\left(d\beta^{random}; \theta\right)}{\int \left[\frac{\prod_{i=1}^{T_i} \left(\exp\left(\beta'^{fixed} x_{ikt}^{fixed}\right) \exp\left(\beta'^{random} x_{ikt}^{random}\right)\right)}{\prod_{i=1}^{T_i} \left(\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right)\right)}\right] \cdot G\left(d\beta^{random}; \theta\right)} \end{aligned} \tag{3.24}$$

The products in the two denominators of the last expression are both:

$$\prod_{i=1}^{T_i} \left(\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right)\right) \tag{3.25}$$

[6] The term ’fixed effect’ must not be confused with a ’usual’ fixed effect in, e.g., a linear panel model where a fixed effect means that each individual has his own base level ($\alpha$ depends on $i$ in the equation $y_{it} = \alpha_i + x_{it}\beta + \varepsilon_{it}$). The term ‘fixed effect’ is used here because that is what Train calls the parameters that follows a one-point distribution as opposed to the random parameters (Train, 1999c; Train et al. 1999).

but since they are part of two separate integrals they do not cancel out. Therefore the odds ratios are *not* independent of the other potential choices, so even the presence of a *single* random component of $\beta$ ensures that IIA does not hold in the MMNL model.

## 3.5 Train's version of mixed multinomial logit

The following will briefly describe how Train et al. estimate the MMNL model in the GAUSS program (Train et al. 1999).[7] This is done in order to illustrate how MMNL can be estimated and to provide better intuition about the model.

The program maximises the log of the likelihood function, as defined in (3.22). In section 3.5.1 the practical use of the program is presented briefly, and the notation to be used in the following is defined. Section 3.5.2 presents the likelihood function and describes how it is simulated.

### *3.5.1 Notation and options in the GAUSS program*

In the model there are:

- $N$ individuals labelled $i$
- $J$ different commodities labelled $j$
- $M$ different exogenous variables labelled $x^m$
- For each individual there are $T_i$ observations labelled $y_{it}$

In the present study the individuals are households, the commodities are different types of eggs and $t$ describes a specific shopping trip rather than a specific time. In different chapters of this study the words *person*, *household*, *family* and *individual* are all used to describe the $N$ individuals. They all mean the same. The words *commodities*, *choices*, *goods* and *egg types* all describe the $J$ commodities. The words *time,*[8] *period*, *purchase* and *occasion* all describe $t$.

The total number of observations is $NOBS = \sum_{i=1}^{N} T_i$

[7] As a preparation for the estimations in this study the entire program has been commented thoroughly. The commented program can be requested at LMA@akf.dk. Readers with special interest in the estimation details may find useful extra information in these 50 pages. The appendix was not included in this working paper out of consideration for the readers with minor interest in estimation details.

[8] Note that in this version of the model the actual time has no influence on the choices, since the observations are simply treated as repeated choices.

In the GAUSS program the data is organised in the following way:

The matrix X of exogenous variables is a $NOBS \times M \cdot J$ matrix, with elements $x_{ijt}^{m}$, $m = 1,\ldots,M$, $i = 1,\ldots,N$, $j = 1,\ldots,J$ and $t = 1,\ldots,T_i$, where $x_{ijt}^{m}$ is the value of the exogenous variable $m$, for individual $i$, if he chooses $j$ at time $t$, and $x_{ikt}^{m}$ is the value he chooses $k$ instead.

The matrix Y of endogenous variables is a $NOBS \times 1$ column with elements $y_{it}$, $i = 1,\ldots,N$ and $t = 1,\ldots,T_i$, where the value of $y_{it}$ is $1,\ldots,J$, depending on which choice was actually made by individual $i$ at time $t$.[9]

In the general definition of the model, the $\beta$'s can follow any distribution. To make the model operational it is necessary to choose a functional form for the distribution of each $\beta$.

In the program each of the parameters $\beta^1,\ldots,\beta^M$ that describe the effect of the exogenous variables $x^1,\ldots,x^M$ is allowed to follow one of five distributions. The following is a quotation from the homepage

http://elsa.berkeley.edu/Software/abstracts/train0296.html:

*(Quote:)* Each coefficient can take any of the following five distributions:

(1) *Fixed* coefficient: the coefficient is the same for all agents (i.e., a degenerate distribution).

(2) *Normally* distributed coefficient, with the mean and standard deviation being estimated.

(3) *Uniformly* distributed coefficients, with the mean and 'spread' being estimated. A uniform distribution with mean $b$ and spread $s$ has a uniform density between $b$-$s$ and $b$+$s$.

(4) *Triangularly* distributed coefficients, with the mean and 'spread' being estimated. A triangular distribution with mean $b$ and spread $s$ has zero density below $b$-$s$, rises linearly from $b$-$s$ to $b$, decreases linearly from $b$ to $b$+$s$, and then is zero again above $b$+$s$.

*(5)* *Log-normally* distributed coefficient; the coefficient is calculated as $\exp(c + s \cdot u)$ where $u$ is a standard normal deviate and $c$ and $s$ are parameters. The program estimates $c$ and $s$. The log-normal distribution with parameters $c$ and $s$ has median $\exp(c)$, mean $m=\exp[c+((s^2)/2)]$, and standard deviation $m \cdot \sqrt{\exp\left(s^2\right)-1}$. *(Unquote).*

[9] Y is a vector of numbers between 1 and $J$. In the simulation it is more useful to have a matrix of zeros and ones, indicating which commodity was chosen. This *YPERM* matrix is $NOBS \times J$ and can be seen either as a selection of dummies for the choice, or as a selection of observed probabilities. If good $j'$ was purchased and all other goods $j \neq j'$ were not, then the observed probability of choosing good $j$ is 1 for $j=j'$ and 0 for all $j \neq j'$. In principle it should be possible to use this to rewrite the program to accept individuals choosing more than one good. This could be interesting if the budget shares were not always zero or one. One could then define an *YPERM* matrix of budget shares for each type of egg at each occasion. It is not needed in the present study, but it could be an interesting extension of the model.

In the following the parameter *c* in the log-normal distribution is labelled '*E*'($\beta$) and the parameter *s* is labelled '*std*'($\beta$), but keep in mind that the estimates are *not* direct estimates of mean and standard deviation. To keep the notation simple the spread estimated for the uniform and the triangular distribution is also sometimes referred to as 'std' in the following, but again it is important to keep in mind that they are not real standard deviations of the distributions.

Figure 3.1 to Figure 3.4 illustrate the density of the last four distributions. If the parameter is fixed, the one point estimated will have probability one. Note that all distributions are symmetric, except for the log-normal one.

**Figure 3.1 Normal distribution**

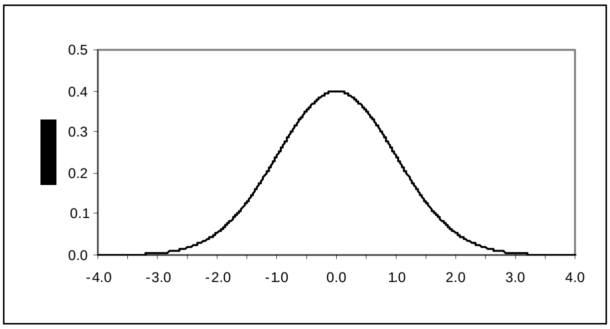



**Figure 3.2 Uniform distribution**

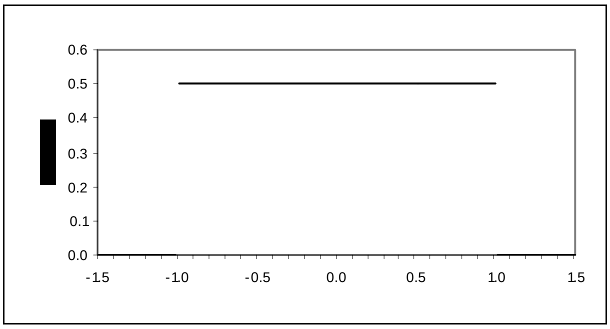



**Figure 3.3 Triangular distribution**

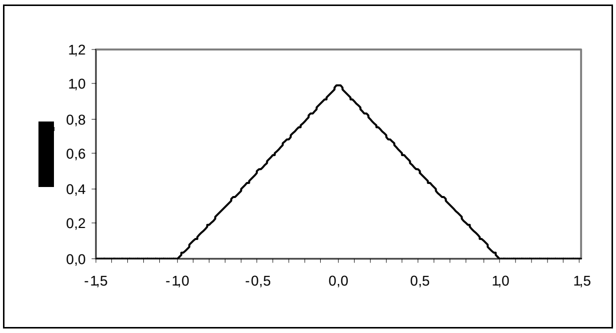
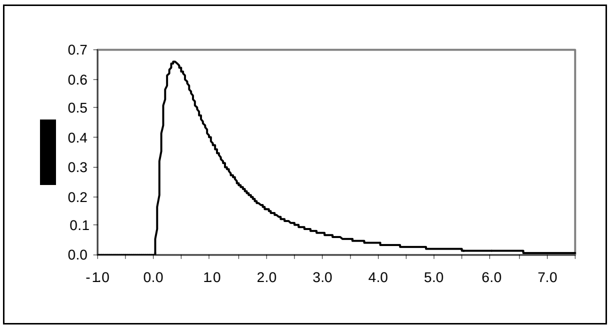



**Figure 3.4 Log-normal distribution**




In the program it is often necessary to treat fixed coefficients in one way, normal, uniform and triangular in another way and log-normal coefficients in a third way. It is therefore useful to define:

$$
\begin{aligned}
x^{fixed} &= \left\{ x^m \middle| \beta^m \text{ is fixed} \right\} \\
x^{norm} &= \left\{ x^m \middle| \beta^m \text{ follows a normal distribution} \right\} \\
x^{uni} &= \left\{ x^m \middle| \beta^m \text{ follows a uniform distribution} \right\} \\
x^{trian} &= \left\{ x^m \middle| \beta^m \text{ follows a triangular distribution} \right\} \\
x^{\lg n} &= \left\{ x^m \middle| \beta^m \text{ follows a log-normal distribution} \right\}
\end{aligned}
\tag{3.26}
$$

$$
x^{nut} = \left\{ x^{norm}, x^{uni}, x^{trian} \right\}
$$

The number of parameters in $\beta$ that follow each distribution can equivalently be defined as:

$$
M \equiv M^{fixed} + M^{norm} + M^{uni} + M^{trian} + M^{\lg n} \equiv M^{fixed} + M^{nut} + M^{\lg n} \tag{3.27}
$$

If e.g. $M^{norm} \geq 1$, then

$$
x^{norm} = \left[ x^{norm^1}, \ldots, x^{norm^m}, \ldots, x^{norm^{M^{norm}}} \right] \tag{3.28}
$$

is a $NOBS \times \left( M^{norm} \cdot J \right)$ matrix, where of course

$$
x^{norm^m} = \left[ x_1^{norm^m}, \ldots, x_j^{norm^m}, \ldots, x_J^{norm^m} \right] \tag{3.29}
$$

is a $NOBS \times J$ matrix.

The vector of parameters $\beta$ can, in general, be described as:

$$
\beta = \begin{bmatrix}
E\left(\beta^{fixed^m}\right)\} & \text{repeated } M^{fixed} \text{ times} \\
\left.\begin{matrix} E\left(\beta^{norm^m}\right) \\ std\left(\beta^{norm^m}\right) \end{matrix}\right\} & \text{repeated } M^{norm} \text{ times} \\
\left.\begin{matrix} E\left(\beta^{uni^m}\right) \\ spread\left(\beta^{uni^m}\right) \end{matrix}\right\} & \text{repeated } M^{uni} \text{ times} \\
\left.\begin{matrix} E\left(\beta^{trian^m}\right) \\ spread\left(\beta^{trian^m}\right) \end{matrix}\right\} & \text{repeated } M^{trian} \text{ times} \\
\left.\begin{matrix} "E"\left(\beta^{\lg n^m}\right) \\ "std"\left(\beta^{\lg n^m}\right) \end{matrix}\right\} & \text{repeated } M^{\lg n} \text{ times}
\end{bmatrix} \tag{3.30}
$$

where the *'E'* $\left(\beta^{\lg n^m}\right)$ and the *'std'* $\left(\beta^{\lg n^m}\right)$ are **not** true means and standard deviations, see the definition of the log-normal distribution.

The program is given initial values for all parameters in $\beta$. It is possible to restrict parameters in the following way

- To their initial value (defined by the user)
- To have the same mean as one or more of the other parameters
- To have the same variance as one or more of the other parameters

Apart from restrictions on the values of $\beta$, the program allows rationing (occurs if not all choices are available at each choice occasion), Halton[10] or random draws, Robust or usual standard errors and weighting of the individual likelihood functions with a variable defined by the user.[11] The precision of the simulated likelihood function increases with the number of iterations used in the simulation. This number is defined by the user.

### *3.5.2 The simulated likelihood function*

The integral in the individual likelihood function is simulated by adding different random errors to the parameters a number of times (iterations), and then taking the mean of the resulting likelihood functions. The errors can be created as Halton draws or random draws, and are altered to fit the distribution of $\beta$.

Using a very general notation, the simulated likelihood function for individual *i* can be written as

$$L_i^{simulated}\left(y_i, x_i, \theta\right) = \frac{\sum_{r=1}^{NREP}\left(\prod_{t=1}^{T_i}\frac{\exp\left(\left(E\left(\beta'|\theta\right)+std\left(\beta'|\theta\right)\cdot err_{ir}\right)x_{ij't}\right)}{\sum_{k\in C}\exp\left(\left(E\left(\beta'|\theta\right)+std\left(\beta'|\theta\right)\cdot err_{ir}\right)x_{ikt}\right)}\right)}{NREP} \quad (3.31)$$

where *NREP* is the number of iterations (or 'repetitions') used in the simulation. It is important to note that there is an error term for each combination of exogenous random variables, $x^m$, individual, *i*, and simulation-iteration, *r*, but NOT for each time/purchase (choice-repetition), *t*, and NOT for each possible choice, *j*. This corresponds perfectly

[10] 'Halton draws' are discussed in detail in Train (1999a). Hensher and Greene (2001), page 3, emphasise Halton draws for creating 'greatly improved accuracy with far fewer draws and computational time'. In some cases this reduced the computation time to about 10% of the time required by conventional methods. Since both Train (1999a) and Hensher and Greene (2001) recommend Halton draws, they will be used in this application.

[11] Weighting the individual likelihood functions might be reasonable in this study if focus was on the substitution effects in order to describe and forecast the market. Some families use a lot of eggs while other families use very few eggs. One might therefore weight the individual likelihood functions by the average number of eggs consumed each week by the household. This would make the estimations reflect the actual market a bit better, though the panel is not completely representative. The purpose of this study however, is not to estimate substitution effects but rather to estimate individual marginal willingness to pay, and weights are therefore not used in this application.

with the utility function presented in Train (1998): $U_{ijt} = \beta_i' x_{ijt} + \varepsilon_{ijt} = b' x_{ijt} + \eta_i x_{ijt} + \varepsilon_{ijt}$, where the error term, $\xi_{ijt} = \eta_i x_{ijt} + \varepsilon_{ijt}$, is correlated over time and over choices. The fact that the error terms for each individual are independent of time means that the individual $\beta$ is constant over repeated choices each time a $\beta$ is drawn from the distribution. The fact that the error terms are independent of the choice means that the $\beta$'s are independent of choice, and therefore that the multinomial logit is a conditional multinomial logit as opposed to a generalised multinomial logit.

The fact that the error terms depend on the exogenous variables means that there is no correlation between the effects of different exogenous variables. The program estimates covariances for the means and standard deviations of, e.g., $\beta^m$ and $\beta^k$, but the actual value of $\beta^m$ drawn from the distribution of $\beta^m$ does not depend on the value of $\beta^k$ drawn from the distribution of $\beta^k$. That means that the probability of having a given value of $\beta^m$ is independent of the value of $\beta^k$. In the same way the fact that the error terms differ from individual to individual means that the individual $\beta$'s are uncorrelated.

In each iteration (or 'repetition', *r*), the error terms are added in the following way:

The contribution from the $M^{fixed}$ fixed $\beta$'s is:

$$\left(x_{ijt}^{fixed}\beta^{fixed}\right)_r = \sum_{m=1}^{M^{fixed}} \beta^{fixed^m} x_{ijt}^{fixed^m} \tag{3.32}$$

The contribution from the normal, uniform or triangular distribution, $\beta^{nut}$ is:

$$\left(x_{ijt}^{nut}\beta^{nut}\right)_r = \sum_{k=1}^{M^{nut}} \left(E\left(\beta^{nut^k}\right) + "std"\left(\beta^{nut^k}\right)\cdot err_{ir}^{nut^k}\right)\cdot x_{ijt}^{nut^k} \tag{3.33}$$

where '*std*' is the spread in the uniform and the triangular distribution.

For the log-normally distributed $\beta$'s the contribution is:

$$\left(x_{ijt}^{\lg n}\beta^{\lg n}\right)_r = \sum_{l=1}^{M^{\lg n}} \exp\left("E"\left(\beta^{\lg n^l}\right) + "std"\left(\beta^{\lg n^l}\right)\cdot err_{ir}^{\lg n^l}\right)\cdot x_{ijt}^{\lg n^l} \tag{3.34}$$

Since the value of $\beta^{fixed}$ is not influenced by the error terms $\beta'^{fixed} x_{ijt}^{fixed}$ is constant in each iteration *r*. The value of $\left(\beta' x_{ijt}\right)_r$ is therefore

$$\begin{aligned}\left(\beta' x_{ijt}\right)_r = &\sum_{m=1}^{M^{fixed}} \left(\beta^{fixed^m} x_{ijt}^{fixed^m}\right) + \\ &\sum_{k=1}^{M^{nut}} \left(\left(E\left(\beta^{nut^k}\right) + "std"\left(\beta^{nut^k}\right)\cdot err_{ir}^{nut^k}\right)\cdot x_{ijt}^{nut^k}\right) + \\ &\sum_{l=1}^{M^{\lg n}} \exp\left("E"\left(\beta^{\lg n^l}\right) + "std"\left(\beta^{\lg n^l}\right)\cdot err_{ir}^{\lg n^l}\right)\cdot x_{ijt}^{\lg n^l}\end{aligned} \tag{3.35}$$

This is used to calculate the value of the simulated likelihood function for individual *i*, in iteration *r* (note that the exponentials are taken on sums, not matrices):

$$L_{ir}^{simulated}(y_i, x_i, \theta) = \prod_{t=1}^{T_i} \frac{\exp(\beta'_r x_{ij't})}{\sum_{j=1}^{J} \exp(\beta'_r x_{ijt})} =$$

$$\prod_{t=1}^{T_i} \left[ \frac{\exp\left( \begin{array}{l} \sum_{m=1}^{M^{fixed}} \left(\beta'^{fixed^m} x_{ij't}^{fixed^m}\right) + \sum_{k=1}^{M^{nut}} \left(E\left(\beta^{nut^k}\right) + std\left(\beta^{nut^k}\right) \cdot err_{ir}^{nut^k}\right)' \cdot x_{ij't}^{nut^k} + \\ \sum_{l=1}^{M^{\lg n}} \exp\left(E\left(\beta^{\lg n^l}\right) + std\left(\beta^{\lg n^l}\right) \cdot err_{ir}^{\lg n^l}\right)' \cdot x_{ij't}^{\lg n^l} \end{array} \right)}{\sum_{j=1}^{J} \exp\left( \begin{array}{l} \sum_{m=1}^{M^{fixed}} \left(\beta'^{fixed^m} x_{ijt}^{fixed^m}\right) + \sum_{k=1}^{M^{nut}} \left(E\left(\beta^{nut^k}\right) + std\left(\beta^{nut^k}\right) \cdot err_{ir}^{nut^k}\right)' \cdot x_{ijt}^{nut^k} + \\ \sum_{l=1}^{M^{\lg n}} \exp\left(E\left(\beta^{\lg n^l}\right) + std\left(\beta^{\lg n^l}\right) \cdot err_{ir}^{\lg n^l}\right)' \cdot x_{ijt}^{\lg n^l} \end{array} \right)} \right] \quad (3.36)$$

Where *j'* in the numerator is the type of egg that was actually chosen by individual *i* at time $t_i$ and 'std' is actually 'spread' of the uniform and the triangular distribution. Note that the only difference between two iterations of the likelihood function is the error terms. The observed values of $x_{ijt}$ and the values of the different $\beta$'s remain the same.

The final value of the simulated likelihood function for individual *i* is then the mean of the *NREP* simulated values

$$L_i^{simulated}(y_i, x_i, \theta) = \frac{\sum_{r=1}^{NREP} L_{ir}^{simulated}(y_i, x_i, \theta)}{NREP} =$$

$$\frac{1}{NREP} \sum_{r=1}^{NREP} \prod_{t=1}^{T_i} \left[ \frac{\exp(\beta'_r x_{ij't})}{\sum_{j=1}^{J} \exp(\beta'_r x_{ijt})} \right] \quad (3.37)$$

This mean simulates the integral in (3.20).

The program maximises the sum of the log of all the individual likelihood functions

$$\ln L^{simulated}(y_i, x_i, \theta) = \sum_{i=1}^{N} \ln\left(L_i^{simulated}(y_i, x_i, \theta)\right) = \ln\left(\prod_{i=1}^{N} L_i^{simulated}(y_i, x_i, \theta)\right) \quad (3.38)$$

over all the parameters of $\beta$ as defined in (3.30).

## 3.6 How to interpret the results of mixed multinomial logit models

The mean and standard deviation entering each individual likelihood function is the same for all individuals. In principle it is possible to construct a model where each individual has his own mean and standard deviation, but it would be extremely difficult to interpret (and most probably also to simulate). The integral in the total MMNL likelihood function is placed in the *individual* likelihood function, so the likelihood function for the entire panel is not an integration made over individuals maximising a conventional multinomial logit likelihood function, but rather a product of individual likelihood functions for individuals each maximising a MMNL likelihood function with the same mean and standard deviation of the $\beta$'s for all individuals.[12]

Following Train (1998) this can be interpreted as if all individuals actually know their own $\beta_i$ as well as their $\varepsilon_{it}$, but the econometrician does not, and has no way of recovering it. For each individual the econometrician therefore integrates over all possible values of $\beta_i$, and then uses the sum of the individual likelihood functions to estimate the distribution from which the different $\beta_i$'s are taken. As mentioned before (in section 3.4.3) it is the assumption about the distribution of the information that is not available to the econometrician (the error terms) that eliminates IIA, not assumptions about the individual behaviour.

The common distribution from which all $\beta_i$'s are drawn is a very interesting feature of the MMNL model. If $\beta^m$ has a small standard deviation, the population responds almost identically to changes in the exogenous variable $X_{ijt}^m$. If the standard deviation is large, it means that the individual responses to $X_{ijt}^m$ can be very heterogeneous. An example of this could be the convenience level of a food product. In some families lack of time means that convenience has a high priority, therefore it is a positive attribute that, e.g., the vegetables are pre-cut when purchased. In other families (with more time or generally more focus on food) it is important to use products that are as close to the original as possible. These families would respond negatively to a high convenience level. This interpretation of the standard deviations in the distributions of the random parameters can be very useful in this application. Estimations may show that some individuals respond negatively to a given egg type, while other individuals respond positively. A negative reaction can be explained, e.g., by mistrust of the labels describing the egg type.

The fact that the estimated distribution of the parameters is a measure of heterogeneity in the population makes it possible to say something about the distribution of, e.g., marginal willingness to pay for different egg types, in the panel. Using MMNL it is possible to tell (assuming that the estimated functional form is close to the true distribution) what percentage of the population has positive or negative marginal willingness to pay and even how many have marginal willingness to pay higher than a certain amount. These estimations depend crucially on the chosen functional form, so they must be interpreted with great care.

---

[12] In McFadden and Train (2000) the model is discussed without panel structure. On page 447, they describe MMNL as a model where: 'The random parameters, $\beta$, may be interpreted as arising from taste heterogeneity in a population of MNL decision makers' (MNL is the conditional logit model). McFadden and Train (2000) therefore interpret the estimated standard deviation as a measure of the degree of heterogeneity in the population. Revelt and Train (1998) and Train (1998) use exactly the same model is used as an individual MMNL likelihood function, and the MMNL likelihood function is now the product of the individual MMNL likelihood functions. Again the estimated standard deviation is taken as a measure of the heterogeneity between individuals.

## 3.7 Comparing the multinomial logit with the probit and the conventional model with the mixed model

As mentioned in section 3.1 the multinomial logit is the result of a random utility model with Weibull distributed error terms. Assuming instead that the error terms are normally distributed leads to a multinomial *probit* model.

In Brownstone and Train (1999) the multinomial logit and the probit are compared through the assumptions made about the utility function both models seek to estimate.[13]

$$U_i = \beta' x_i + \eta_i + \varepsilon_i \tag{3.39}$$

Here the individual utility, $U_i$, is decomposed into

1. A non-stochastic linear in parameters part that depends on observed data ($\beta' x_i$)
2. A stochastic part that is perhaps correlated over alternatives and heteroscedastic over people and alternatives ($\eta_i$)
3. Another stochastic part that is i.i.d. over alternatives and people ($\varepsilon_j$)

In short terms the difference between multinomial logit, probit, mixing and conventional can be described by the assumptions about $\eta_i$ and $\varepsilon_j$:

**Table 3.1 Comparing the conventional and mixed logit and probit**

| | $\eta_i$ | $\varepsilon_j$ |
|---|---|---|
| **Conventional multinomial logit** | One point zero distribution | i.i.d. extreme value[14] |
| **Mixed multinomial logit** | General distribution | i.i.d. extreme value |
| **Mixed multinomial probit** | General distribution | Conventional normal distribution |
| **Conventional multinomial probit** | One point zero distribution | Conventional normal distribution |

Adapted from Brownstone and Train (1999).

The choice between the different models therefore depends on the expectations of the distribution of $\eta_i$ and $\varepsilon_j$. Whether the $\eta_i$ is zero or not can be tested once the mixed model is estimated. Therefore, the choice is primarily between the multinomial logit and the probit model. Brownstone and Train (1999) conducted an experiment to compare the relative accuracy of the mixed multinomial logit with the mixed probit simulation. The result was that given the same number of *iterations*, the mixed probit produced less variance than the mixed multinomial logit, but because simulation is slow in the mixed probit, the mixed multinomial logit produced considerably lower variance given the same amount of computer *time*. Brownstone and Train (1999) therefore recommends the mixed multinomial logit instead of the mixed probit.

[13] Note that this utility function is slightly different from the one used in Train (1998): $U_{ijt} = \beta_i' x_{ijt} + \varepsilon_{ijt} = b' x_{ijt} + \eta_i x_{ijt} + \varepsilon_{ijt}$

[14] See footnote 1 on page 11 for a definition of the extreme value distribution. It is the same distribution that McFadden (1973) calls a Weibull distribution.

## 3.8 Difficulties in presenting the results

The model is not linear, which means that the effect of changes in X on the probability of Y=*j* is not given by $\beta$ alone. In the conventional multinomial logit model the problem is 'solved' by the IIA that makes it easy to use the parameters to calculate how changes in the exogenous variables will effect the relationship between the probability of choosing *j* and the probability of choosing *k*. See (3.11), (3.12) and (3.13). Getting rid of IIA makes the model more realistic, but it also makes it far more difficult to interpret the parameters.

In the previous it has been shown that under MMNL:

$$prob\left(Y_i = j \middle| x_i, \theta\right) = \int \left[ \prod_{t=1}^{T_i} \frac{\exp\left(\beta'^{fixed} x_{ijt}^{fixed}\right) \exp\left(\beta'^{random} x_{ijt}^{random}\right)}{\sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right)} \right] \cdot G\left(d\beta^{random}; \theta\right) \quad (3.40)$$

and therefore

$$\frac{prob(y_{it} = j)}{prob(y_{it} = k)} = \frac{\int \left[ \frac{\prod_{i=1}^{T_i} \left( \exp\left(\beta'^{fixed} x_{ijt}^{fixed}\right) \exp\left(\beta'^{random} x_{ijt}^{random}\right) \right)}{\prod_{i=1}^{T_i} \left( \sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right) \right)} \right] \cdot G\left(d\beta^{random}; \theta\right)}{\int \left[ \frac{\prod_{i=1}^{T_i} \left( \exp\left(\beta'^{fixed} x_{ikt}^{fixed}\right) \exp\left(\beta'^{random} x_{ikt}^{random}\right) \right)}{\prod_{i=1}^{T_i} \left( \sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right) \right)} \right] \cdot G\left(d\beta^{random}; \theta\right)} \quad (3.41)$$

The effect of a change in $x_{ijt}$ will therefore not only depend on the parameters and the relationship between the estimated probabilities. But also on the attributes of the other alternatives:

$$\frac{\partial \left( \frac{prob(y_{it} = j)}{prob(y_{it} = k)} \right)}{\partial x_{ijt}} = \frac{\partial \left( \frac{\int \left[ \frac{\prod_{i=1}^{T_i} \left( \exp\left(\beta'^{fixed} x_{ijt}^{fixed}\right) \exp\left(\beta'^{random} x_{ijt}^{random}\right) \right)}{\prod_{i=1}^{T_i} \left( \sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right) \right)} \right] \cdot G\left(d\beta^{random}; \theta\right)}{\int \left[ \frac{\prod_{i=1}^{T_i} \left( \exp\left(\beta'^{fixed} x_{ikt}^{fixed}\right) \exp\left(\beta'^{random} x_{ikt}^{random}\right) \right)}{\prod_{i=1}^{T_i} \left( \sum_{l \in C} \exp\left(\beta'^{fixed} x_{ilt}^{fixed}\right) \exp\left(\beta'^{random} x_{ilt}^{random}\right) \right)} \right] \cdot G\left(d\beta^{random}; \theta\right)} \right)}{\partial x_{ijt}} \quad (3.42)$$

which does not have a simple general solution.

This effect of the missing IIA is both an advantage and a disadvantage when trying to describe substitution patterns (how do people react to changes in attributes). In the present study the object is 'merely' to estimate the marginal willingness to pay for different types of eggs,[15] which reduces the problem, since marginal willingness to pay in MMNL is still the parameter for the attribute divided by the negative of the parameter for price. The virtue of MMNL in this connection is that it is possible to estimate a distribution of marginal willingness to pay and not only a fixed number. In most cases it is far more reasonable to assume that people have different values of marginal willingness to pay than that they all have exactly the same marginal willingness to pay. In this application the mixed multinomial logit model is therefore used in the estimation of marginal willingness to pay.

## 3.9 Conclusion

The conventional multinomial logit models should be well known to anyone who has dealt with discrete statistical models, but the mixed multinomial logit model is reasonably new and may not be as familiar. This chapter has, therefore, not only introduced the general (and theoretical) version of the model, but also briefly gone through the simulation program developed by Kenneth Train, David Revelt, and Paul Ruud at University of California, Berkeley.

One of the main advantages of the mixed multinomial logit model is that it eliminates the assumption of Independence of Irrelevant Alternatives (IIA) that is both the virtue and the Achilles' heel of the conventional multinomial logit. This property of the mixed model was, therefore, investigated further and it was found that IIA is eliminated by the fact that the econometrician is unable to observe both the individual $\beta_i$'s and the individual error component, $\varepsilon_{it}$, and therefore assumes that the $\beta_i$'s are drawn from a distribution that is common for all individuals. This means that the remaining error terms are correlated for each individual, and it is this correlation that eliminates IIA.

Finally it is noted that in this study the most important difference between a conventional multinomial logit and a mixed multinomial logit is the fact that estimated marginal willingness to pay is not assumed to be the same for all individuals when using the mixed model, but rather to follow a distribution defined by the econometrician. Since this allows a wide range of interesting conclusions about the distribution of marginal willingness to pay in the sample, it is concluded that the virtues of the mixed model exceed the extra difficulties in estimation, and the mixed multinomial logit model will therefore be used in the estimations on data in chapter 6 and 7.

[15] See chapter 2 for a definition of 'marginal willingness to pay'.

# 4 GfK data: Households, shops and products

This chapter describes the data that forms the basis for the empirical part of this study. Section 4.1 describes the entire data set in general terms and sections 4.2 and 4.3 present the subsample of data to be used in this particular study. Section 4.2.1 compares the panel with the entire Danish population, and concludes that the panel on the whole is representative of the entire population. Section 4.2.2 presents the background data recorded about the households in general, and compares customers in different stores, concluding that socio-demographics, habits and attitudes vary between customers in different stores. Section 4.3 presents the purchase data that will be used to estimate the value of the different types of eggs, and concludes that the price level varies between stores, making it important in further analysis, to include the information about the aggregated stores in which the purchases are made. Section 4.3.1 presents the four different egg types used in this study, section 4.3.2 investigates the distribution of production on egg types and section 4.3.3 describes the prices of different egg types in different store aggregates. Section 4.3.4 explores the consumption patterns observed in the purchase data. Section 4.4 concludes chapter 4 by stating that the GfK data provides an opportunity to investigate consumption of non-durables using Danish data including unique information about the individual consumer.

## 4.1 Description of the data source

The data for this study is collected by GfK ConsumerScan Denmark (GfK). The original data set consists of purchases of more than 50 different groups of food-commodities, and covers a four year period from 1997 to 2000. On an average week, almost 1,600 families report to GfK.

This section describes the data in general terms by introducing GfK (section 4.1.1), discussing the number of families in the panel (section 4.1.2), describing how and what they report (section 4.1.3) and how GfK structures this information (section 4.1.4). In section 4.1.5 the data regarding the household in general is presented. Section 4.1.6 describes how GfK recruits households, ensures that they are representative and keeps them in the panel. Section 4.1.7 concludes the general presentation of the data and introduces the subset of the data that will be used in this study.

### *4.1.1 GfK*

'GfK – ConsumerScan (Dansk HusstandsPanel)' is a product offered by GfK Danmark A/S. GfK Danmark A/S is part of the international GfK-Group, one of the largest opinion-research groups in the world. More information about GfK can be found at www.gfk.dk or www.gfk.com.

GfK has, among its many other activities, a consumer panel. Each week households[1] in the panel report their actual purchases to the GfK in a 'diary'. The diaries cover about 80 percent of each family's budget for 'everyday necessities' (GfK, 2001). Among other attributes, the households state whether the goods are organic or conventional. Once a year they answer a questionnaire about household attributes such as family income, municipality, number of children, education, profession, etc. The questionnaire also contains questions about club membership, television viewing preferences, and for some years, attitudes to convenience products (ready made meals and other processed food), special offers, cooking, etc.

All data is self-reported by the households. GfK recommends that the diaries are filled in immediately after each shopping trip to avoid problems with forgotten purchases. This means that even though it is not real scanner data it comes very close. The information provided by the families is so detailed that it must be based on actual and not 'self-estimated' behaviour.

The panel has the usual advantage of panel data: Having data that varies over both time and individuals allows estimation of more than data following either an average of consumer consumption over time (time series), or a cross-section of consumers at a given time.

### *4.1.2 Number of families reporting to GfK*

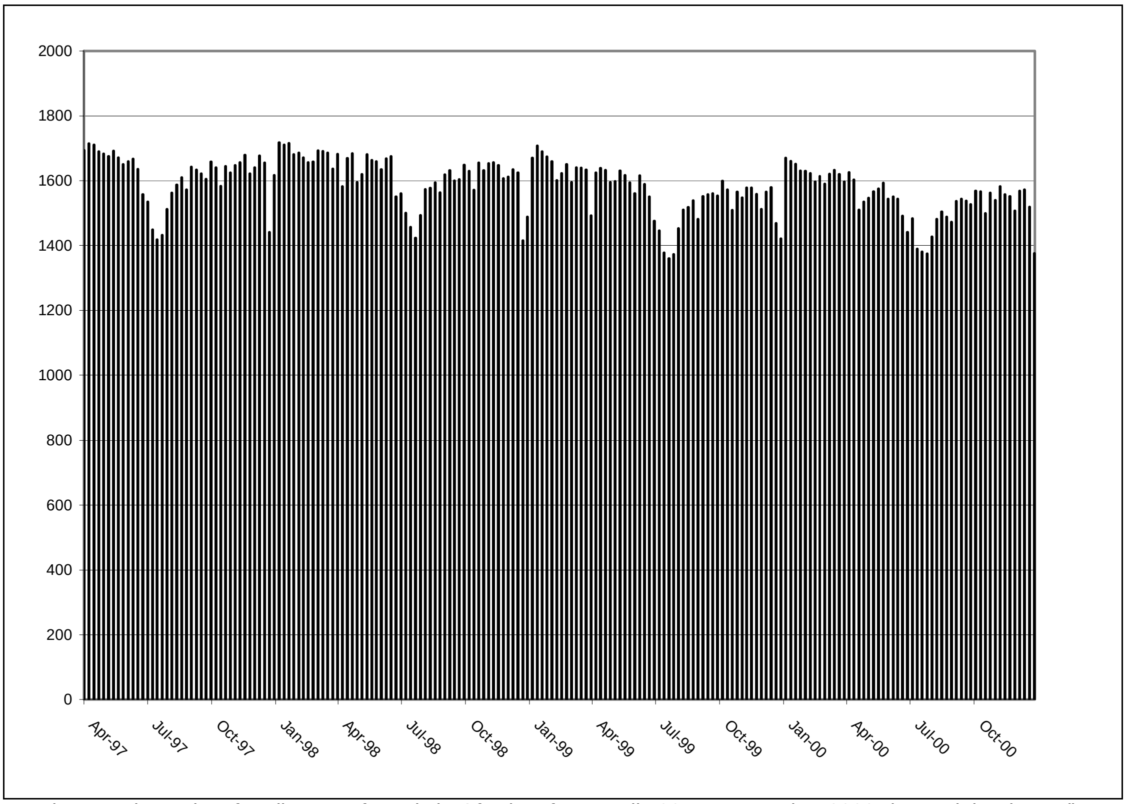
The number of families in the panel does not have an unambiguous measure. The families send in reports of purchases ('diaries') every week, but not all families send reports each week. The number of diaries in a given week can be seen as the number of active families in this particular week, but the number of families within a given month will be the number of families that sent in at least one diary during that month etc. The number of families therefore increases with the length of the period, as can be seen in Figure 4.1 and Table 4.1.

The actual number of diaries varies from week to week. As expected, the reporting rate goes down during the summer season and around Christmas, but not to a level that causes any concern. Figure 4.1 and Table 4.1 show the number of families reporting within different time spans in the period from April 1997[2] to December 2000.

[1] The households are also called families in this study. The definition of the household/family is not given by GfK, so most probably purchases made by tenants and other 'non-family' members of the household will not be included in the purchase data, and these non-family members will probably not occur in the background data about the family either.

[2] During the first quarter of 1997 the reports on organic versus conventional were not fully operational for all groups of goods. The original data set is adapted for use in a project that focuses on organic foods. The first quarter is therefore eliminated from the data set.

**Figure 4.1 Number of families reporting to GfK within a given week**




Based on purchase data for all types of goods in GfK data from April 1997 to December 2000. (Un-weighted panel)

The number of families reporting to GfK within a given month, quarter or year is described in Table 4.1 below.

**Table 4.1 Number of families reporting to GfK per year, quarter, month and week**

| | Year | Quarter | Month | Week |
|---|---|---|---|---|
| **Max** | 2,169 | 1,903 | 1,800 | 1,717 |
| **Min** | 2,034 | 1,719 | 1,658 | 1,360 |
| **Mean** | 2,108 | 1,807 | 1,731 | 1,585 |
| **Median** | 2,114 | 1,816 | 1,746 | 1,597 |

Based on purchase data for all types of goods in GfK data from April 1997 to December 2000. (Un-weighted panel)

Table 4.1 shows that during the period from April 1997 to December 2000, the maximum number of diaries within a week was 1,717, the minimum was 1,360, the mean was 1,585 and the median was 1,597. Thus on average, GfK receives approximately 1,600 diaries each week. During a year more than 2,000 families contribute to the panel, and during the entire period, 3,278 families were part of the panel for some period of time (not shown in Table 4.1).

It is important to note that it is not the same 1,600 families that report each week. The families stay in the panel for various lengths of time, and the reporting rate also varies between families. It is difficult to give a measure of the individual reporting rate since some families may be completely inactive for long periods and then return to being active members, while other families send in diaries infrequently. Both report-patterns may yield the same report rate, but the implications on estimation may be very different. The total number of diaries sent in by each family is therefore presented instead.

**Table 4.2 Quantiles of number of diaries per family in the GfK data**

| Number of diaries | Percentage of families that have this or a lover total number of diaries |
|---|---|
| 196 | Highest number of observations |
| 196 | 99% |
| 193 | 95% |
| 190 | 90% |
| 171 | 75% |
| 83 | 50% |
| 31 | 25% |
| 11 | 10% |
| 5 | 5% |
| 3 | 1% |
| 2 | Lowest number of observations |

Based on purchase data for all types of goods in GfK data from April 1997 to December 2000.

Table 4.2 shows that half of the 3,278 families send in at least 83 diaries during the data period and that 25 percent (approximately 820 families) send in more than 170 diaries. This gives an enormous amount of information about the purchasing behaviour of many different families.

The households report purchases once a week (see section 4.1.3 for more on this), and once a year they also provide more general information about the background data of the household (see section 4.1.5 for more information about the background data).

### *4.1.3 Information from the households to GfK*

The members of the panel are households. Each household chooses a 'diary keeper', a person responsible for completing the diary and mailing it to GfK. Often this person will also be responsible for most of the shopping. Each quarter the household receives a diary from GfK to be filled out by the household. The diary covers 13 weeks, (GfK splits the three months into 5, 4 and 4 weeks). Each week the diary keeper fills in information about everything purchased and sends this weekly diary to GfK.

The data recorded depends on the type of good. For pasta the data recorded includes whether it is fresh or dry, for bread it includes whether it is sliced or not. The information can also differ over time. Information about egg type is e.g. only collected during a one year period from July 1999 to June 2000 and not in the rest of the four year period.

If the diary keeper has, e.g., bought eggs, the following information is recorded:

- Scanner data (EAN-code (‘stregkode’) see www.ean.dk for more information)
- Type: Battery eggs, barn eggs, free-range eggs or organic eggs[3] (this specification is of course only relevant for eggs, for other goods the ‘type’ captures other features of the good)
- Organic/conventional (this is recorded for all goods that can possibly be organic, and therefore also for eggs)
- Eggs per unit (per egg tray)
- Number of units (number of egg trays)
- Price per unit (per egg tray)
- Price of the entire purchase (the cost of all eggs of the same type that have been purchased, not any other goods)
- Special offer or not
- Name of store

Apart from food the households report purchases of potted plants, flowers, vitamins, magazines, personal products (shampoo, deodorant, sanitary towels etc), cleaning agents, paper towels and toilet paper each week. This study only has access to an aggregated level of EAN-codes, see later.

Each shopping trip is assigned a number, and for each trip the diary keeper reports:

- The day of the week
- The time
- The name of the store
- Who participated in the trip (this information stopped in the last quarter of 2000)
- The total value of the goods purchased

Once the diaries reach GfK, they are checked to see if they are filled in correctly, before they are keyed in. Each purchase is automatically checked for consistency of brand, packaging, price, store etc. After this, a list of all purchases in a week is printed out and checked manually.

GfK then structures the data obtained from the households, thereby creating the actual data set.

### *4.1.4 How GfK structures the information*

The goods are grouped in categories such as ‘all kinds of milk’, ‘eggs’, ‘yellow cheese, sliced and un-sliced’, ‘instant coffee and other instant drinks’, ‘spaghetti, pasta and noodles’ etc The number of groups varies between 71 and 79 in different years. See appendix A for a list of commodity groups (in Danish).

[3] In Danish battery eggs are ‘bur æg’, barn eggs are ‘skrabe æg’, free-range eggs are ‘fritgående æg’ and organic eggs are ‘økologiske æg’.

Each good is identified by a 'c-code'. A c-code covers a group of goods identified by their EAN-code. For most practical purposes the EAN-codes are too detailed. GfK therefore constructs a more aggregated 'c-code', for each group of goods, based on the EAN-codes. The c-codes usually provide information about producer and brand, but the level of detail in c-codes vary from group to group and from year to year within a group. EAN-codes are not available in the data used in this study, only the c-codes. A c-code may, e.g., cover all coloured pasta products from a specific manufacturer, or even all pasta products from a given manufacturer regardless of other characteristics. Therefore a c-code may cover both organic and conventional products, and it may also cover different types of eggs. But since the organic attribute and the egg type are recorded separately by the diary keeper, these attributes are covered by a different code, and it is therefore still possible to distinguish between organic and conventional products, and between different types of eggs. In 2000 purchases of 99 different egg c-codes were reported to GfK, and since there are also other codes indicating type of egg, tray size etc it is possible to make the data even more detailed.

The level of detail varies from group to group. If different groups are to be aggregated it is important to note that the information about a common characteristic (such as whether the good is organic or conventional) is not always stored in variables with the same name, by GfK, and more general variables must therefore be constructed from the original data.

#### *4.1.5 Background information about the households*

When a household is recruited it fills out a background questionnaire. These questionnaires are updated once a year in October. The questions vary a little from year to year, but some questions are the same throughout the observation period. See appendix C for details.

For the entire period (1997-2000) information exists about the age of all members of the household (including children), education and occupation of the 'mother' and 'father',[4] household income, gender of the person mainly responsible for shopping and of main income provider. Apart from conventional socio-demographics, GfK gathers information about club memberships and membership of (non-political) organisations, equipment in households (including e.g. freezers and computers), general choice of store and store type, use and availability of different newspapers and magazines, use and availability of different TV-channels, use of weekly catalogues/flyers ('tilbudsaviser') from stores and finally, price sensitivity in general terms.

For parts of the period, information is gathered about household income left after regular outlays, opinions about 25 different chains of stores, more details about TV habits, more details about use of catalogues/flyers, access to and use of internet, cars in

[4] The 'mother' and the 'father' are the adult female and male members of the household regardless of their parental status.

the household, attitudes to convenience products,[5] general attitudes and habits related to cooking, and frequency of cinema visits.

A general overview of the questions in the background questionnaire is given in appendix B, and a more detailed description of the question and answer categories is given in appendix C. Combining data from all years lead to a data set with almost 800 variables.

#### *4.1.6 Recruitment, representativeness and payment*

GfK Denmark also conducts opinion polls by telephone interviews. These interviews are the main recruitment source. At the end of the interview each household is asked whether they would like to participate in the panel.

To make sure these households represent the entire Danish population of private households, GfK weights the observations by counting some households as two instead of one. The size of the weighted sample was 2,527 in 2000, where the size of the un-weighted sample was approximately 2,080. This working paper uses only the *un-weighted* sample. And it is the representativeness of this sample that will be investigated in section 4.2.1.

When recruiting and weighting, the primary attributes are (GfK, 2001):

- Region (urban/rural)
- Household size
- Age of main shopper
- Membership of FDB (now COOP)

The panel is continuously balanced to fit the best available public statistics on the following attributes:

- Father’s age
- Father’s profession
- Mother’s profession
- Family type
- Primary shopping place

Approximately 20 percent of the sample is replaced each year. The replacement rate varies from year to year.

[5] The convenience products mentioned in the questionnaire are: Frozen lasagne/pizza, instant sauces, instant soup, frozen pasta/rice meals, instant bouillon (‘Bouillonterninger’), spice mixes like ‘Knorr Mexican meal spice mix’.

The panel covers ‘private Danish households’ consumption of daily necessities’. According to GfK (2001), it does *not* cover

- Individual consumption outside the household (chewing gum, chocolate etc)
- Consumption in institutions (hospitals, kindergartens etc)
- Consumption in canteens, offices etc
- Commercial or industrial consumption
- Tourists shopping in Denmark

In this connection it is important to note that it does not cover durables such as cars, refrigerators, housing etc. Total consumption is therefore not covered by the data, only daily consumption is registered.

The diary might not capture all consumption even if the goods are supposed to be listed in the diaries. First, some diaries are never sent to GfK at all, and second, some diaries may be incomplete. The total consumption of food recorded in the data is therefore not necessarily equal to the actual total consumption and the difference between observed and actual consumption may vary from household to household.

The panel households are rewarded with a number of points for each questionnaire they complete. A long or difficult questionnaire gives more points than a short and easy one. These points can be used to buy goods in the ’GfK-store’, which exists only for the panel households. The members of the panel are therefore not directly paid for their work, but receive goods instead.

### *4.1.7 Why use GfK data, and for what purpose*

GfK provides a unique source of information at a very detailed level. As can be seen in appendices A, B and C, this can be a double-edged sword. First, the level of detail means that working with data and establishing usable data sets can be quite troublesome, and in this study it has taken up quite a lot of resources. The total number of observations in the purchase data exceeds five million and although the number of variables provided by GfK is less than 20 (c-code, organic/conventional, package size, price, etc), the variables cover different information for different groups of goods, and are therefore difficult to use in a more general setting. As mentioned, the background data are also very detailed and this leads to a background data set with 10,386 observations and 799 variables. The many different variables make it particularly difficult to work with background data, simply because finding the relevant information requires a good grasp of all 800 variables.

When exploiting the data it is necessary to choose between using information about many goods at an aggregate level (which requires aggregation of goods, quantities and values) or a few goods at a more detailed level. As usual, the more one knows about the true state of the world, the more complicated it gets to describe it.

It is partly for this reason that focus of this study is on only one group of goods (eggs) and only a limited part of the background information is used. This study uses a relatively new econometric method, and the complexity of the data must therefore be restricted. Information about the type of eggs that were purchased is crucial to the estimations, and this kind of information has only been collected during a one year period. This reduces the data set to approximately 24,000 observations based on purchases of approximately 2,000 families, but still with 800 different pieces of information about each of these families.

## 4.2 Background data on the households

As mentioned above, the panel consists of a number of households, who are weighted by GfK to create a more representative panel. The un-weighted panel is used here partly because no description of the weighting procedure is available, and partly because observations that occur more than once in a small subset of the data might cause problems in estimations.

In this working paper only data from July 1999 to June 2000 will be used. In section 4.2.1 the families in the entire panel will be compared with the entire Danish population in 1999, in order to find out whether the panel is representative of this entire Danish population.

The primary source of background data is data collected in October 1999, but as some families had left the panel before October 1999, data collected in October 1998 must be used for these families. When the families enter the panel they fill in the background questionnaire almost immediately, and the data is added to the data set collected in the previous October. Data for families that enter the panel during 2000 therefore appear in the October 1999 data. For the entire panel (in the relevant year) only 5 percent do not have background data obtained in 1999 and for families purchasing eggs, the number is 3 percent. Data on the entire Danish population is collected by Statistics Denmark ('Danmarks Statistik') and is available via their homepage www.dst.dk or directly at www.statistikbanken.dk.

Section 4.2.2 compares the background data of customers in different chains of stores to see if behaviour in different chains of stores can be expected to be the same and section 4.2.3 sums up the presentation of background data.

### *4.2.1 Representativeness of the panel*

It is important to know whether the panel can be used as a representation of the Danish population in general. Data on the entire Danish population is collected by Statistics Denmark, but only on conventional socio-demographics[6], and the representativeness of the panel will therefore only be investigated based on this type of information.

[6] Some data on habits and attitudes are also collected by Statistics Denmark, but these are usually based on panels like the one used by GfK. It will therefore be difficult to know which set of panel results represents the entire Danish population best.

#### *4.2.1.1 Geography*

First of all, it is important to know whether the panel covers all of Denmark. The geographical distribution of the panel can be compared with the entire Danish population by looking at counties. Figure D.1.1 in appendix D shows the distribution on counties for the panel members (July 1999 to June 2000) and the entire Danish population in 1999. The panel is perhaps a bit underrepresented in the Copenhagen area[7] (23.4 percent of the panel lives here, but 25.1 percent of the entire Danish population), but it is a marginal difference and otherwise the panel is representatively distributed over counties.

#### *4.2.1.2 Age, marital status, number of persons in household and number of children*

The households consist of one or more adults and possibly some children. It is therefore natural to investigate the representativeness of these individuals and the way they are combined. The category age shows that the most significant difference between the panel and the population in general, is that young people (below 30 years) are less likely to be part of the panel than the population in general since only 9.2 percent of the male panel members are 20-29 years old compared to 18.8 percent of the entire population. Women between 20 and 29 years constitute 13.2 percent of the female panel members compared to 17.4 percent of the population in general. Persons between 40 and 60 years are slightly overrepresented for both genders, and men above 60 are also overrepresented, but the differences are not as significant as for young people.[8]

The lack of young men in particular, also influences the 'marital' status of the panel members.[9] The panel households are more likely to include couples than the households in the general population (60.7 percent of the households in the panel but only 53.0 percent of all Danish households). Since single women are also overrepresented (33.7 percent of the panel, 27.8 percent of all Danish households), this can be accredited to the fact that single men are much underrepresented in the panel (5.6 percent in the panel, 19.2 percent of all Danish households).

The underrepresentation of single men also means that the number of households consisting of only one person is underrepresented (33.2 percent in the panel, 36.6 in the population), and the number of households consisting of exactly two persons are overrepresented (37.2 percent in the panel, 33.1 in the population).[10] It is therefore natural to look at the number of children in the households, and it turns out that the fraction of households with children is higher in the panel than in the population in general (30.9 percent in the panel, 22.4 in the population).[11]

Summing up, the underrepresentation of single young men means that the panel is generally a bit older than the population in general, and therefore also more often consists of couples with children.

---

[7] Copenhagen Municipality, Frederiksberg Municipality and Copenhagen County.
[8] Figure D.1.2 and D.1.3 in appendix D.
[9] Figure D.1.4 in appendix D.
[10] Figure D.1.5 in appendix D.
[11] Figure D.1.6 in appendix D.

#### *4.2.1.3 Occupation and income*

Participating in the panel is quite time consuming which raises concerns that mainly persons outside the labour market participate. This would decrease the representativeness of the panel. In order to compare the working status of panel members with the entire Danish population, the panel (and the population) have been divided into the categories 'working', 'assisting spouse', 'unemployed' and 'pensioner, student, on leave etc'.

As mentioned above, men 60 years old or above are overrepresented in the panel. This means that male 'pensioners, students, on leave etc'[12] are overrepresented in the panel too (31.0 percent in the panel 27.2 in the population)[13] and that working men are underrepresented (64.2 percent in the panel 69.8 in the population). Elderly women are not overrepresented in the panel, which means that 40 percent are 'pensioners, students, on leave etc' in both the panel and the population. Unemployed women are slightly overrepresented in the panel (3.9 percent in the panel and 3.3 in the population), which means that working women are slightly underrepresented (55.4 percent in the panel, 56.2 in the population), but the differences are marginal.[14]

Data about the type of work (blue collar/white collar)[15] is also collected by GfK. However, it is very difficult to compare this information with information about the entire population, since 'blue collar/white collar' is not defined by GfK, but by the persons answering the background questionnaires. People are simply asked to state whether the male and the female member of the household is blue collar or white collar. Representativeness regarding type of work is therefore not explored here.

An examination of household income shows that households with very low incomes and relatively high incomes are underrepresented in the panel (20.7 percent of the panel has a household income lower than 150,000 DKK, compared to 24.7 percent of the general population. 28.7 percent of the panel has an income higher than 400,000 DKK, compared to 35.2 of the general population), while households with incomes between 200,000 and 400,000 DKK are overrepresented (38.4 percent of the panel, 30.0 of the population). The overrepresentation of couples may explain why low household incomes are underrepresented.[16]

Summing up, the distribution of persons inside and outside the labour market is not much different from the population in general, but household income is generally less extreme in the panel than in the population in general.

#### *4.2.1.4 Conclusion on representativeness*

The un-weighted panel is quite representative on most of the demographics investigated here. In general, the tendency is that people are more 'average' than the population in

---

[12] Male students are probably underrepresented since young men are underrepresented, and very few men are on leave, thus, the overrepresentation of this group is induced by the overrepresentation of elderly men.

[13] Figure D.1.7 in appendix D.

[14] Figure D.1.8 in appendix D.

[15] Blue collar is 'arbejder' in Danish, white collar is 'functionær'.

[16] Figure D.1.9 in appendix D.

general, since young people, single men, households without children and households with very low or very high income are all underrepresented.

### *4.2.2 Comparing segments of the data*

One of the things that are registered about a purchase is the store in which the purchase was made. The store can be a single store located at a particular place in Denmark, or it can be a chain consisting of sometimes more, sometimes less homogeneous stores. Corner stores ('kiosk') are an example of a group of very heterogeneous stores, while 'SuperBrugsen' is an example of a chain of rather homogeneous stores.

When estimating marginal willingness to pay[17] for different types of eggs in chapter 6 it turns out that the results vary when data from different stores is used in the estimations. It is therefore interesting to investigate possible differences between customers in different stores.

It is important to notice that the purpose of this section is *not* to explore whether the customers in different stores are representative of the entire panel. On the contrary, the purpose is to reveal differences between the customers in different stores, which might explain the differences in behaviour in these stores.

There are more than 130 different definitions of stores in the original data, and in order to make it feasible to use this information, it is necessary to aggregate these stores into groups. Table E.1.2 in appendix E shows which stores are included in each store aggregate. An aggregated store may be either a chain of stores (e.g. SuperBrugsen), or a group of similar stores (e.g. 'greengrocers etc').

Using these aggregated stores it is possible to make a nesting of the data that makes the store aggregates in the data more and more homogeneous. Corner stores and petrol stations are very heterogeneous and can be expected to have extreme prices (compared to the rest of the market) and limited variety, and they are therefore excluded from the data along with non-food stores and other stores with very few observations. The remaining data are called 'subsample A'. In 'subsample B', 'directly from farms' (including sales directly from vans and markets) are removed, mainly because the variety can be expected to vary within this type of outlet. Further removing aggregates that are very heterogeneous (in both price and variety) lead to 'subsample C'. 'Irma' is also deleted in subsample C because battery eggs are purchased only once leading to a very poor price estimate. Table 2.3 presents an overview of the subsambles.

[17] Marginal willingness to pay was defined in chapter 2.

**Table 4.3 Aggregated stores in the different subsamples used in the study (defined by the author)**

| Aggregated stores in the entire sample | Subsample A | Subsample B | Subsample C |
|---|---|---|---|
| SuperBrugsen | √ | √ | √ |
| DagligBrugsen | √ | √ | √ |
| Kvickly and OBS | √ | √ | √ |
| Irma | √ | √ | |
| Fakta (Discount) | √ | √ | √ |
| Føtex | √ | √ | √ |
| Netto (Discount) | √ | √ | √ |
| Aldi (Discount) | √ | √ | √ |
| Prima | √ | √ | √ |
| Favør | √ | √ | √ |
| Various grocers | √ | √ | |
| Various discount stores | √ | √ | |
| 'greengrocers etc' | √ | √ | |
| 'Directly from farms' | √ | | |
| Corner store/petrol station | | | |
| Non-food | | | |
| Only convenience | | | |
| Other stores | | | |
| Bilka | √ | √ | √ |
| A-Z (DS) | | | |
| *Number of observations* | | | |
| 24,147[18] | 23,720[19] | 21,050 | 15,816 |
| *Number of families* | | | |
| 1,943[20] | 1,941 | 1,846 | 1,693 |

Note: '√' means that this store aggregate is included in the subsample. The aggregated stores are defined in Table E.1.2 in appendix E.

Each household can of course make purchases in more than one store, and typically do so, though not usually in all of the aggregated stores. The purchases made in different stores can therefore be made by households that have different characteristics, and in some cases it can even be expected that e.g. the geographical characteristics differ. Some stores are restricted to a specific area, and mainly families in that area will be using the stores. More complex and interesting differences occur when the characteristics are attitudes and habits. Eighty-five percent of the families in subsample A are also represented in subsample C (and 95 percent in subsample B) and the characteristics of the families in the three subsamples therefore only differ marginally.

In section 4.2.2.1 the geographical distribution in stores is investigated, in section 4.2.2.2 age, marital status and number of children of customers in different stores is

[18] Includes 144 purchases of free-range eggs that are possibly also organic. These purchases are excluded in the subsamples.

[19] Originally 23,818 purchases, but purchases of the same type of eggs (from different producers or in different tray sizes) on the same shopping trip are aggregated to one purchase, leaving 23,720 observations.

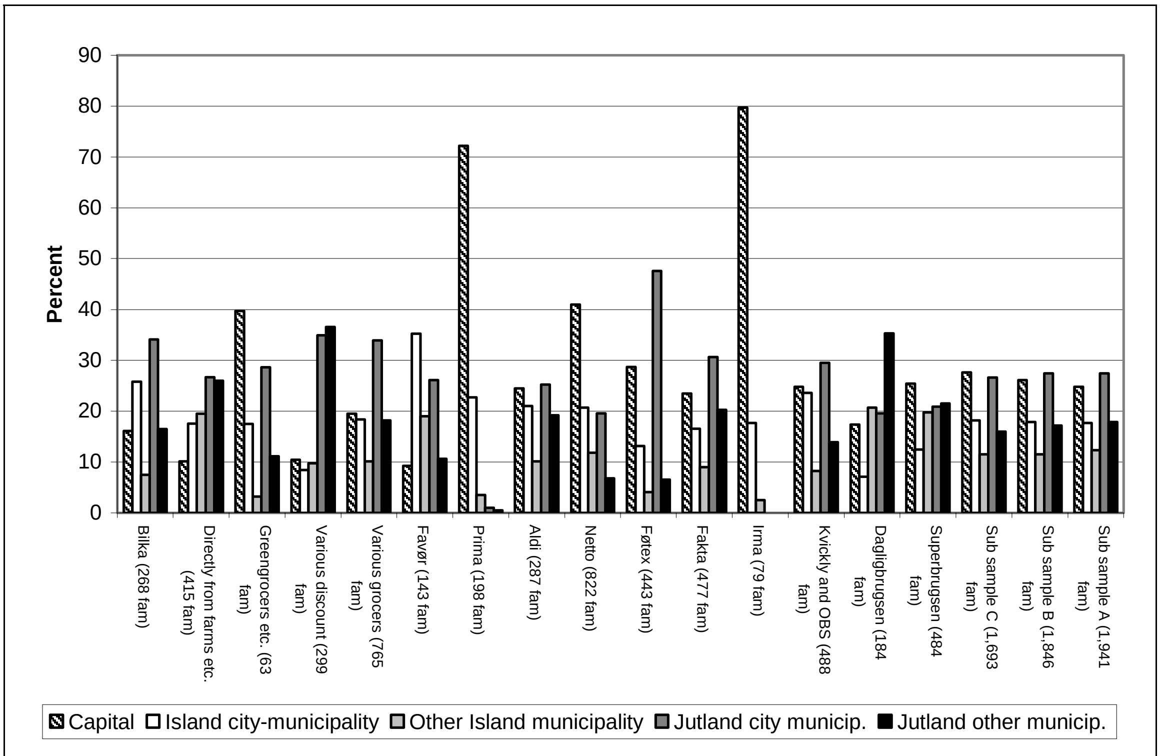
[20] The entire panel consists of 2,196 families during the period from 26 June 1999 to 30 June 2000. Only 1,943 families purchase eggs within this period, which means that 255 families do not.

presented. Section 4.2.2.3 looks at occupation and income and section 4.2.2.4 investigates habits and attitudes of the customers.

#### *4.2.2.1 Geography*

First, the geographical distribution of families in the different subsamples is investigated to find out whether all stores are found in all of Denmark, and if their market shares are the same throughout Denmark. In section 4.2.1.1, the geographical distribution of the panel was compared with the general population by examining the distribution on counties. The results of comparing the distribution of customers on 14 counties for each of the 18 subsamples would be incomprehensible; therefore, another GfK-defined measure of geography is used.[21] The Danish municipalities are divided into three zones, Capital[22], Jutland and 'Islands', and Jutland and 'Islands' are further divided into 'city' municipalities and 'other' municipalities:

**Figure 4.2 Distribution of families in different subsamples of the GfK data by Capital, Jutland and Other**




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

Prima and Irma appear to be operating primarily on Zealand, and not in Jutland, which also fits the actual geographical distribution of the stores. It is, however, generally important to note that people may very well make purchases in other areas than the one they are living in and this will disturb the picture of the actual geographical location of the stores.

[21] See exact definition of geographical categories in appendix F.
[22] Copenhagen Municipality, Frederiksberg Municipality and Copenhagen County.

Small stores such as greengrocers and cheese shops seem to be used more in the cities, especially in the Capital area. Relatively many Føtex customers come from city municipalities in Jutland, but Føtex is found throughout the entire country. The group 'various discount stores' covers one or more stores that are strongly represented in Jutlandic non-city municipalities and these municipalities are also heavily represented among the customers in 'DagligBrugsen'.

The geographical differences lead to differences in the distribution of customers living in one-family houses, apartments and on farms. In subsample A (all families purchasing eggs) 61.3 percent lived in one-family houses, 32.0 percent in apartments, 3.6 percent on farms and 3.1 percent in two family houses or rented rooms. However, the distribution among customers living in one family houses and apartments varies a great deal among the stores. In particular, in the two non-Jutlandic chains, Prima and Irma, house owners are underrepresented compared to the entire panel (29.1 percent in Irma and 47 percent in Prima). People who live on a farm more often buy directly from other farms (5.8 percent of these purchases are made by people who live on a farm) and farmers are especially overrepresented in DagligBrugsen (10.3 percent against 3.1 percent of the entire sample). The latter result is not unexpected since DagligBrugsen is overrepresented in rural areas.[23]

The conclusion is that not all stores are found in all of Denmark and that even if they are, the distribution of customers from urban areas and rural areas vary from store to store.

#### *4.2.2.2 Age, marital status and number of children*

When looking at age distribution for men[24] and women[25] in the different subsamples it turns out that customers in Bilka are younger than the panel members in general, while customers in Irma, 'directly from farms' or speciality stores such as greengrocers are generally older.

Investigating marital status[26] for the customers in different stores shows that singles are overrepresented in Netto and Irma, and couples are overrepresented in Bilka and DagligBrugsen. These differences are also reflected in the fraction of customers without children.[27] In general, 69.5 percent of the households have no children, but in Bilka the fraction is only 59.6 percent and in DagligBrugsen only 62.5. In Irma, on the other hand, 87.3 percent of the customers have no children.

As expected, there is a connection between marital status and number of children and to some extent also age. The distribution of these parameters varies from store to store.

#### *4.2.2.3 Occupation and income*

In section 4.2.1.3 the panel was compared with the Danish labour market statistics for the whole population. In this section the *type* of work is investigated. The differences

[23] Figure D.2.1 in appendix D.
[24] Figure D.2.2 in appendix D.
[25] Figure D.2.3 in appendix D.
[26] Figure D.2.4 in appendix D.
[27] Figure D.2.5 in appendix D.

are most pronounced for the father's profession.[28] In DagligBrugsen more men are self employed in the primary sector (4.9 percent in DagligBrugsen compared to 2.0 percent in the entire panel) and more men are blue collar workers[29] than in the panel in general (35.7 percent in DagligBrugsen, 27.9 percent in the panel). Discount stores such as Aldi and Fakta also attract more blue collar workers than stores in general (29.5 percent in Aldi and 30.5 percent in Fakta), whereas Netto (which is also a discount store) attracts relatively many white collar workers (40.2 percent in Netto, 35.6 percent in the panel).

The differences are not as significant for the mother's profession,[30] but it is interesting to note that while the fraction of white collar workers is almost the same for men and women (35.6 percent for men and 37.9 percent for women in the entire panel) the fraction of blue collar workers differs dramatically, since 27.9 percent of the men are blue collar workers, but only 14.9 percent of the women are blue collar workers. The fraction of self employed women is also lower than for men, meaning that (compared to men) women are placed outside the labour market instead of being blue collar workers. (45.4 percent of the women are outside the labour market, as compared to 30.9 percent of the men).

An examination of income[31] shows no major differences between the different subsamples, except perhaps a slight tendency towards people with low incomes using discount stores and people with very high incomes being a bit overrepresented in Prima and Irma. Irma generally has a higher price level as it tends to stock more luxury and speciality goods.

The most interesting thing about the GfK data is that it not only provides information about usual socio-demographic characteristics, but also about attitudes and habits. As mentioned in section 4.1.5 (and described in detail in appendices A, B and C) the information covers numerous areas, and varies in the level of detail. In the following section, newspaper habits and attitudes to prices will be presented.

#### *4.2.2.4 Habits and attitudes*

Habits and attitudes are expected to be correlated with preferences for different types of eggs, and might even be more strongly correlated than conventional socio-demographics. It is therefore interesting to investigate whether habits and attitudes varies between customers in different store aggregates.

[28] Figure D.2.6 in appendix D.

[29] White collar workers are called 'funktionærer' in Danish, whereas blue collar workers are called 'arbejdere'. The households in the panel are simply asked to state whether the mother and the father are 'abejder' or 'funktionær', no detailed definition is given by GfK.

[30] Figure D.2.7 in appendix D.

[31] Figure D.2.8 in appendix D.

Newspaper habits can be described by subscription[32] (availability) and number of papers read[33] (use). There are quite big differences in the fraction of non-subscribers,[34] especially on weekdays. In general less than 50 percent of the households subscribe to a newspaper on weekdays, but subscription is more frequent among customers in 'directly from farms', 'greengrocers etc', 'various grocers', Irma, 'Kvickly and OBS', DagligBrugsen and to some extent also SuperBrugsen. Customers in 'greengrocers etc', Irma, 'Kvickly and OBS' and DagligBrugsen read the paper more frequently than the population in general, and customers in greengrocers and Irma often read more than one paper per day.

Attitude to prices is a very interesting type of information since it can be expected to influence purchasing habits, and it is usually unknown. GfK measures this attitude by asking the respondents to mark the statement that best fits their attitude. There are three questions, each with three possible answers

1. Choose between
    a) I prefer brand labels ('mærkevarer') to cheaper products, to be sure to get good quality
    b) No-name products are often just as good as brand labels. I buy cheaper no-names just as often as brand labels
    c) Don't know/not answered

2. Choose between
    a) I look carefully for special offers when shopping. This saves me quite a lot of money
    b) Looking for special offers is too much trouble. It is not worth the effort
    c) Don't know/not answered

3. Choose between
    a) I almost always shop in stores where I know the prices are low
    b) It is not decisive if the prices are a bit higher in the store where I shop. As long as the location and variety is good
    c) Don't know/not answered

If a person chooses b), a) and a) it means the person always prefers a low price. If the person chooses a), b) and b) it means that the person always ranks something else higher than the price. The three statements are combined in a variable that states the number of times the cheapest solution was preferred.

[32] Many people do not subscribe to any newspaper, but read these, either by buying them one at a time, or by reading them at work, or somewhere else. On average (subsample A), less than 20 percent have not read any papers within the last week. Compared with the fraction of non-subscribers (more than 50 percent), it means that 30 percent of the households read papers without subscribing (but not necessarily without paying).

[33] On weekdays, the newspapers in the GfK data are: Berlingske Tidende, BT, Børsen, Aktuelt, Ekstra Bladet, Fyens Stiftstidende, Jyllandsposten, Jydske Vestkysten, Politiken, Aalborg Stiftstidende, Århus Stiftstidende, 'The local paper in my area' (not free papers) and Information. GfK asks how many papers out of the last six possible have been read by the person filling in the questionnaire (probably often the person responsible for most of the shopping). Adding the answers for the papers mentioned yields a measure of the number of papers read each week. The distribution of answers is shown in figure D.2.10 in appendix D.

[34] Figure D.2.9 in appendix D.

Looking at the statements for customers in different stores yields the interesting results[35] presented in Figure 4.3 to Figure 4.6.

Figure 4.3 shows the percentage of customers in different stores who prefer brand labels. People who use small stores such as ‘greengrocers etc’ are particular fond of brand labels, and so are people using the COOP stores (previously FDB) Irma, ‘Kvickly and OBS’ and SuperBrugsen. Customers in Føtex also prefer brand labels more often than the panel in general. Customers in Prima, Aldi, Netto and Fakta care more about price than brands compared to the entire panel. This partitioning of the stores supports the general perception of the stores, where the first stores focus on quality rather than price, and the last stores are, to varying degrees, discount stores.

Figure 4.4 shows the percentage of customers in different stores who prefer *not* to spend time looking for special offers. This shows a somewhat different picture. Where people in DagligBrugsen did not care more about brand labels than the panel in general, they are more reluctant to spend time looking for special offers. People who shop in Prima cared less about brand labels than the average, but they also care slightly less about looking for special offers. Customers in Bilka, ‘Directly from farms’, Aldi and Fakta are more interested in looking for special offers than the panel in general.

Figure 4.5 shows the percentage of customers who find store location and assortment more important than prices when deciding which store to use. This picture looks more like the one for brand labels, but note that store loyalty is more widespread than brand loyalty since more than 50 percent rank other attributes above price when choosing stores, but only slightly more than 25 percent do so when choosing brands (Figure 4.3). Note that this does not mean that almost 75 percent look only at price when choosing brands, many other factors such as, e.g., freshness enter the decision.

In Figure 4.6 the three statements are combined. Bilka, ‘Various discount stores’, Aldi, Netto and Fakta have a higher rate of customers who always choose the cheapest product. Particularly, Aldi and to some extend Fakta have many price conscious customers. ‘Greengrocers etc’, Føtex, Irma, Kvickly and OBS, DagligBrugsen and SuperBrugsen all have relatively small fractions of very price conscious customers. Again the distinction fits the distinction between discount stores and non-discount stores.

[35] The percentage of ’unknown/not answered’ (c) varies from 0.2 to 2.2, which is very small compared to the fractions of a) and b) answers. Only one of the a) and b) answers is therefore presented in order to make it easier to read the figures.

**Figure 4.3 Distribution on statement one, brand labels and no-name products**

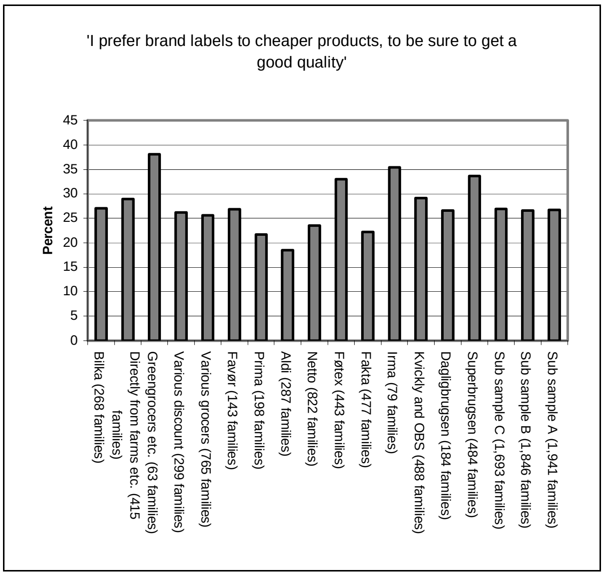



**Figure 4.4 Distribution on statement two, searching for special offers**

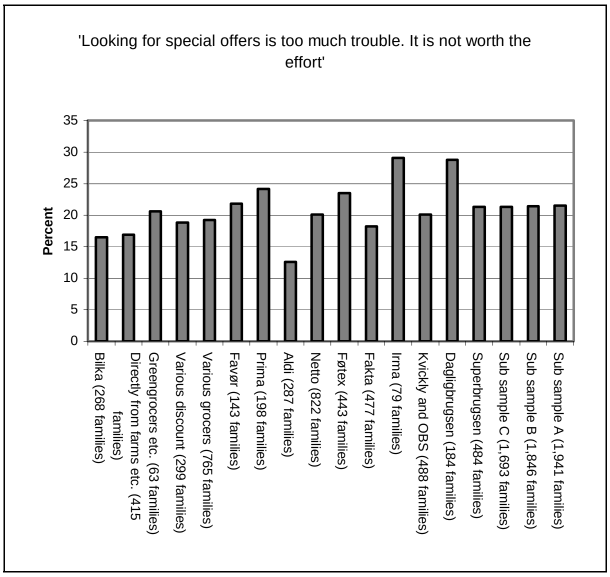



**Figure 4.5 Distribution on statement three, price level not important when choosing store**

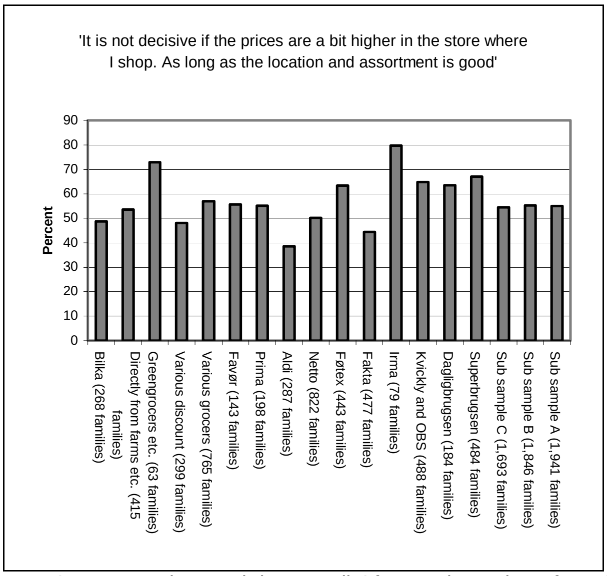
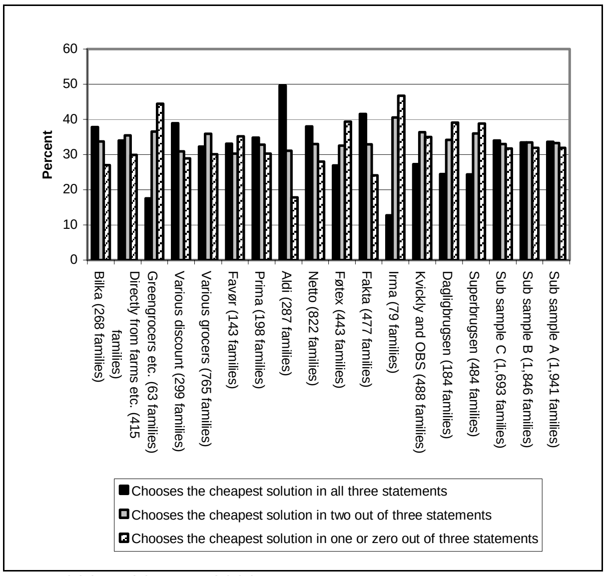



**Figure 4.6 Distribution on a combination of the three price statements**




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsamples A to C represent nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

Compared to socio-demographics and newspaper habits, the information about price consciousness seem to give a far clearer picture of the differences between stores, and it may well prove to be yield a better way revealing differences in preferences for different types of eggs.

#### *4.2.2.5 Conclusion on segments of the data*

In general, the differences between the stores exist, and may often be ascribed to differences in geographical dispersion. Prima and Irma do not cover Jutland, and DagligBrugsen is overrepresented in Jutlandic non-city municipalities. Among the discount stores, Netto is more of a city phenomenon than Aldi and Fakta, but all three chains are represented in all of the regions.

Stores primarily located in cities have a higher fraction of customers that are flat dwellers, single persons and white collar workers. Stores that are not conventional supermarkets (‘directly from farms’ and ‘greengrocers etc’) have a lower fraction of young customers.

Important differences between discount stores and non-discount stores are captured by information about habits and attitudes, information that is available in the GfK data, but will often be unavailable in other data sets.

Since customers differ between aggregated stores, it is reasonable to expect preferences for different types of eggs to vary between customers in different stores.

### *4.2.3 Conclusion on background data*

The un-weighted panel is reasonably representative of the entire Danish population, at least as far as conventional socio-demographics can tell. The panel may well differ in some respects that are not usually measured in the entire population, but this is difficult to determine since attitudes and habits are rarely recorded by Statistics Denmark, and if they are, the numbers are based on test samples, not the entire population.

Many habits and attitudes are recorded by GfK, and investigating these along with the conventional socio-demographics leads to the conclusion that customers differ from store to store, not only in socio-demographics but perhaps more importantly in habits and attitudes. This may prove to be important in the further analysis of the data, since it means that customers with different attitudes towards different types of eggs may choose to place their purchases in different stores, and thereby making the distribution of preferences among customers vary from store aggregate to store aggregate.

## 4.3 Purchase data for eggs

This study only uses data on eggs. Detailed data on eggs is available during a one year period from July 1999 to June 2000. Less detailed data is available during the entire period from 1997 to 2000 but is not used in the present study.

In section 4.3.1 the different egg types will be presented, and it will be discussed briefly how the data can be used to estimate the value that people ascribe to different types of eggs. Section 4.3.2 describes the Danish production of eggs. Section 4.3.3 introduces the sales channels in the egg market that are defined by aggregated stores and describes the prices observed in the market. Section 4.3.4 presents data on consumption of the different egg types in general and from different stores.

### *4.3.1 The different egg types*

Danish eggs fall into five categories

- Battery eggs ('bur æg')
- Aviary eggs ('voliere æg')
- Barn eggs ('skrabe æg')
- Free-range eggs ('æg fra fritgående høns')
- Organic eggs ('økologiske æg')

Aviary eggs are not recorded as a separate type of eggs in the GfK data, but investigating the Danish egg production (in section 4.3.2) shows that aviary eggs constitute an extremely low share of the total production, and the missing information is therefore not a problem.

All eggs are assumed to have the same objective nutritious qualities (contents of vitamins, proteins etc). The labels 'barn eggs' and 'free-range eggs' mainly indicates increased animal welfare compared to battery eggs, whereas the 'organic' label indicates a more environmentally friendly production as well as a higher level of animal welfare. Some households may also perceive the organic eggs as being healthier than other egg types because the hens are fed with organic feed. This might contribute to a higher marginal willingness to pay for organic eggs. The label 'organic' may also have a value of its own, since the governmentally authorised 'Ø-label' is used on many different goods and is a familiar label. The estimated marginal willingness to pay for organic eggs compared to battery eggs is therefore not only a measure of the marginal willingness to pay for animal welfare related to organic eggs compared to the animal welfare related to battery eggs. In this study it is impossible to distinguish the health effect from the effect of the 'Ø-label'.[36]

The contents of the labels are regulated by the rules under which the different eggs are produced. These rules are summarised in Table 4.4.[37]

[36] Eggs may be infected with various bacteria (salmonella etc) and some people may perceive the level of risk to vary from egg type to egg type. Unfortunately no information about the general perception of risk related to different egg types is available in this study. Some people may perceive the risk related to organic eggs as being lower than the general risk, and other might perceive it as being higher. It is therefore impossible to make any hypothesis about the sign of the marginal willingness to pay for security in different egg types and thus this issue is ignored.

[37] The information in Table 4.4 is taken partly from http://www.poultry.dk/ where a similar table is listed. The information from this table has been supplemented with information from European Commission (1991), Ministry of Agriculture (1992) and Danish Plant Directorate (2000).

**Table 4.4 Rules for production of different egg types** (see source in footnote 37 on page 46)

| | **Battery eggs** | **Barn eggs** | **Free-range eggs** | **Organic eggs** |
|---|---|---|---|---|
| **Average flock size** | 4 | 3,000-10,000 | 3,000-10,000 | 1,000-15,000<br>3,000* (4,500 in transition period) |
| **Access to henhouse** | No | Yes | Yes | Yes |
| **Space requirements (hens per $m^2$)** | 16 | Max 7 | Max 7 | Max 6 |
| **Access to outdoor areas** | No, live in cages with 4 hens in each | No, live in stables | Yes, during daytime | Yes |
| **Outdoor space requirements** | - | - | 10 $m^2$ per hen (0.4 $m^2$ per hen under 'intensive free-range production'**) | 4 $m^2$ per hen |
| **Vegetation required outdoors** | - | - | Mainly covered with vegetation | Mainly covered with grass or other crop, there must be access to shade and shelter |
| **Windows required in henhouse** | No | No | No | Yes, the daylight must be allowed to enter to a degree that can activate the animals. |
| **Bedding ('strøelse')** | No | At least 1/3 of the indoor area is covered with bedding such as straw, shavings, sand or peat | At least 1/3 of the indoor area is covered with bedding such as straw, shavings, sand or peat | At least 1/3 of the indoor area is covered with bedding such as straw, shavings ('høvlspåner'), sand or peat ('tørv') |
| **Nests** | No | Yes | Yes | Yes, (either 12.5 single nests per 100 hens or 1.2 $m^2$ shared nest per 100 hens*) |
| **Perches ('siddepinde')** | No | Yes | Yes | Yes (at least 18 cm per hen*) |
| **Beak trimming ('næbtrimning')** | Yes | Yes | Yes | No |
| **Preventive antibiotics** | No*** | No*** | No*** | No |
| **Organic feed** | No | No | No | Yes, at least 75% |
| **Change of pen ('foldskifte')** | | | | Yes, to allow the vegetation to be re-established and to reduce the risk of transferring infections, the outdoor area must be free of poultry every second year or the top 10 cm soil or covering must be replaced. |

* Recommended by LØJ (2000) (Landsorganisationen for Økologisk Jordbrug/'The Danish Organisation for Organic Farming').
** Eggs produced under the 'intensive free-range' must be marked with 'Æg fra fritgående høns – intensivt system' (Free-range eggs – intensive system).
*** According to the LØJ rules, adding growth-promoting antibiotics and other additives is only *prohibited* in organic farming. According to The Danish Poultry Association, these are not used in any of the other production forms either.

*Battery eggs* are very different from other egg types because the hens are kept in small cages (equivalent to ½ m × ½ m) with four hens in the same cage (leaving 25 cm×25 cm for each hen). They never leave the cages and most probably only experience artificial light.

*Barn eggs* are laid by hens that are kept in big flocks of up to 10,000 hens in a big stable. There are no rules about artificial light versus daylight, and they have no outdoor areas available. There can be up to 7 hens per square meter (leaving 38 cm×38 cm for each hen). At least one third of the stable must be covered by dry bedding, allowing the hens to scratch. The hens must also have access to perches (‘siddepinde’), but there are no rules about how many or how long these perches must be.

*Free-range eggs* are laid by hens that are treated like barn hens, except for the fact that they must have access to outdoor areas, and these areas must mainly be covered by vegetation. There are two different kinds of free-range production; the normal and the intensive version. In the normal version each hen must have at least 10 $m^2$ of outdoor area, in the intensive version only 0.4 $m^2$. Free-range eggs produced under intensive conditions must be labelled ‘Free-range eggs – intensive system’. The GfK data does not distinguish between free-range eggs produced under the normal and the intensive system.

*Organic eggs* are laid by hens that are, in some respects, better off than free-range hens and in other respects perhaps a bit worse off. Inside the stable there must not be more than 6 hens per $m^2$ (41 cm × 41 cm for each hen) which is better than free-range hens, but outdoors each hen only has to have access to 4 $m^2$, where the free-range hens must have access to 10 $m^2$ under one system and 0.4 $m^2$ under another system. The organic hens are fed primarily with organic feed, but it is hard to say if this increases the level of welfare for the hen. There are also rules about replacing the top-soil or keeping the area free of poultry every second year. Again something that does perhaps not influence the welfare of the hen, but could decrease the risk of infected eggs.

Although the rules give an objective description of the life of the different types of hens, the level of animal welfare under different production methods is constantly debated. A major problem for non-battery eggs is that the high-productive breeds are not bred for the purpose of being kept in large flocks, but rather in small cages. The large flocks lead to big problems with the pecking order among the hens. In some cases hens are pecked to death or stressed so badly that they start plucking themselves in an abnormal way.

The pecking problems are, to a certain degree, avoided by trimming the beaks on all hens except the organic ones. The organic farmers argue that the hens need the tip of the beak to peck on vegetables and other coarse food, and that trimming the beaks therefore reduces the level of animal welfare. The debate for and against beak trimming has only reached the general public within the last year or two, so families in the panel would not be expected to be influenced by this during the relevant period.

Two other things distinguish the organic eggs from the other egg types. In organic egg production there are rules against additives in the feed (e.g. growth promoting antibiotics). Under the other production types it is claimed that additives are not used, but there is nothing in the rules that prevent it. The other difference is the fact that organic hens must experience daylight indoors. There is of course nothing that prevents a barn egg producer from putting windows in his stable, but there are no rules demanding it either.

Apart from differences in rules for production, organic eggs have the advantage of using a familiar label (the ‘Ø-label’ which identifies organically produced goods) that is used on many different goods. The consumers have a general perception of the Ø-label, and do not have to spend time and energy investigating a new label such as barn eggs or free-range eggs.

Even if the level of animal welfare related to different egg types could be measured precisely in some way, the animal welfare perceived by the consumers could easily diverge from this objective measure. It is very difficult to calculate a ‘net-level’ of animal welfare, since all production forms have pros and cons. Some people even claim that battery hens have better animal welfare because they are less exposed to stress and natural deceases. It is therefore not possible to estimate the value of one ‘unit’ of animal welfare, only the value of different egg types.

It can be expected that people have different levels of information about the egg types. This will contribute to differences in utility and thereby the value from person to person, but can never (unless information about knowledge is gathered for specific individuals) be separated from differences in the utility gained from the perceived contents of the labels. These restrictions on measuring utility of the different egg types must be kept in mind when estimating and interpreting the results.

### *4.3.2 Danish production of the different egg types*

Before investigating the purchase data for eggs it is important to look at the production of the different egg types, primarily to investigate whether the lack of information about aviary eggs causes problems. Table 4.5 presents details of the 1997 Danish egg production.

**Table 4.5 Total 1997 production of eggs**

| 1997 production | Producers delivering to authorised egg packing departments | | Hens producing eggs to authorised egg packing departments | | Average number of hens per producer | Distribution of egg production as percentage of weight of total production |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| Battery eggs | 132 | 33 | 2,525,875 | 65 | 19,135 | 68.2 |
| Aviary eggs | 2 | 1 | 30,060 | 1 | 15,030 | 0.5 |
| Barn eggs | 127 | 31 | 719,362 | 19 | 5,664 | 16.5 |
| Free-range eggs | 66 | 16 | 352,828 | 9 | 5,346 | 8.4 |
| Organic eggs | 78 | 19 | 242,000 | 6 | 3,103 | 6.5 |
| *Total* | *405* | *100* | *3,807,125* | *100* | *9,400* | *100.0* |

Source: The Danish Poultry Association ('Det Danske Fjerkræraad') [38]

From Table 4.5 it can be seen that in 1997 battery eggs were generally produced by large producers (almost 20,000 hens on average per producer) and that the number of hens per producer was far lower for barn eggs and free-range eggs (less than 6,000). The average number of hens per producer was even lower for organic eggs with slightly more than 3,000 hens per producer. Since the percentage of hens differs from the percentage of produced eggs, some hens must be more productive than others. Apparently battery hens and organic hens are slightly more productive than the rest. Aviary eggs were only produced by two producers, and therefore have a negligible share of the market.

Data from Statistics Denmark illustrates the development of the product shares since 1997.

**Table 4.6 Danish production of eggs in different years**

| Percentage of total production | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|
| Battery eggs | 68 | 67 | 64 | 62 | 60 |
| Aviary eggs | 0 | 0 | 0 | 0 | 0 |
| Barn eggs | 17 | 14 | 16 | 17 | 17 |
| Free-range eggs | 9 | 9 | 8 | 9 | 9 |
| Organic eggs | 7 | 10 | 12 | 13 | 13 |
| *Total* | *100* | *100* | *100* | *100* | *100* |

Source: Statistics Denmark ('Danmarks Statistik') (ANI8)

The two sources in Table 4.5 and Table 4.6 agree on the distribution of production in 1997. The production of organic eggs has changed dramatically since 1997, stealing market shares primarily from the battery eggs. The number of hens per producer presented in Table 4.5 may therefore not be valid for the years following 1997.

[38] **Det Danske Fjerkræraad.** 'The Danish Poultry Association' is a trade association for all Danish producers of poultry.

In GfK data aviary eggs are not recorded as a separate type of egg, but will most probably fall into the category ‘unknown type of egg’. As can be seen in Table 4.5 and Table 4.6, this is not a very big problem. In Table 4.6 it clear that there has been a shift from production of battery eggs to production of organic eggs. Aviary eggs have apparently not benefited from this development since the production remains close to zero percent of total production. It is therefore not crucial to the analysis of choice of egg type to include information about aviary eggs, and this study estimates the marginal willingness to pay for barn eggs, free-range eggs and organic eggs, all compared to the marginal willingness to pay for battery eggs.

The following section gives a brief overview of the actual purchase data for eggs.

### *4.3.3 Sales channels and prices*

The data contains information about ordinary unprocessed eggs (‘shell eggs’) as well as pasteurised eggs. The two products are very difficult to compare. Pasteurised eggs can be purchased as yolks or whites only, which make it difficult to compare the prices. If one only needs yolk, pasteurised yolks are cheaper than buying ordinary eggs, but if one needs both yolk and whites, the pasteurised eggs are more expensive than ordinary eggs. If pasteurised eggs were to be included in the model, it would require a nested structure, where people first choose the level of security by choosing between pasteurised or non-pasteurised eggs, and then later choose either which type of pasteurised egg to buy or which kind of shell egg to buy. This lies beyond the scope of this study and purchases of pasteurised eggs are therefore disregarded.

The GfK definition of the egg types means that some eggs end up in a category where it is impossible to tell if they are free-range or organic. This happens in 0.6 percent of all purchases (see Table 4.8) and the observations are deleted. In other cases stores must be coded wrong. There are, e.g., six purchases of eggs in a DIY (do-it-yourself) centre. These purchases are also deleted along with two purchases made in a Canteen, a single one in the store ‘A-Z’ and three purchases where price or number of eggs seemed to be unrealistic. The total number of observations deleted due to errors in the coding is only twenty.

The eggs are sold from stores that are aggregated into 17 different store aggregates[39] that could possibly sell eggs (non-food stores are excluded). The distribution of *purchases* of eggs and *number* of eggs purchased in these aggregated stores are presented in Table 4.7.

[39] See appendix E for a definition of the aggregated stores.

**Table 4.7 Distribution of purchases and number of eggs purchased in aggregated stores**

| Aggregated store | Number of purchases in this store | Number of eggs purchased in this store | Percentage of all eggs purchased by the panel |
|---|---|---|---|
| SuperBrugsen | 2,379 | 24,905 | 8.63 |
| DagligBrugsen | 636 | 7,593 | 2.63 |
| Kvickly and OBS | 1,787 | 17,838 | 6.18 |
| Irma | 223 | 1,734 | 0.60 |
| Fakta (Discount) | 1,875 | 19,125 | 6.63 |
| Føtex | 1,591 | 16,510 | 5.72 |
| Netto (Discount) | 4,636 | 43,863 | 15.20 |
| Aldi (Discount) | 957 | 12,138 | 4.21 |
| Prima | 646 | 6,811 | 2.36 |
| Favør | 515 | 6,352 | 2.20 |
| Various grocers | 3,632 | 40,245 | 13.94 |
| Various discount stores | 1,204 | 16,084 | 5.57 |
| 'greengrocers etc' | 225 | 2,600 | 0.90 |
| 'Directly from farms' | 2,798 | 60,980 | 21.13 |
| Corner store/petrol station | 87 | 678 | 0.23 |
| Other stores | 99 | 1,174 | 0.41 |
| Bilka | 857 | 10,005 | 3.47 |
| *Total* | *24,147* | *288,653* | *100.00* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, and also possibly organic too. A purchase may include more than one tray of eggs, but only one type of eggs.

The most important thing to note is that 'Directly from farms' constitute 20 percent of the market for eggs, and therefore is an important factor that must be taken account of in the analysis. Some stores are very small even when aggregated. The aggregates Irma, 'Greengrocers etc', Corner stores and 'Other stores' are particularly small, and it must be considered further if they are to remain in the data. This will be taken up again later.

After having introduced the producers and the store aggregates it is therefore natural to proceed by investigating their influence on the price. Table 4.8 presents the purchases of different egg types by the entire panel.

**Table 4.8 Purchases of different egg types**

| Egg type | Number of eggs purchased | Percentage of all eggs purchased by the panel | Number of purchases | Percentage of all purchases made by the panel |
|---|---|---|---|---|
| Battery eggs | 128,425 | 44.5 | 9,842 | 40.8 |
| Barn eggs | 44,392 | 15.4 | 3,917 | 16.2 |
| Free-range eggs | 57,097 | 19.8 | 3,875 | 16.0 |
| Organic eggs | 55,793 | 19.3 | 6,369 | 26.4 |
| Free-range eggs, possibly also organic | 2,928 | 1.0 | 144 | 0.6 |
| *Total* | *288,635* | *100* | *24,147* | *100* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range and possibly also organic. A purchase may include more than one tray of eggs, but only one type of eggs.

As mentioned, there are a few purchases for which it can not be decided if the eggs are free-range or organic. These eggs are included in the table to give an idea of the size of the problem. It is obviously not a very big problem, and the purchases are just deleted from the data.

The prices of eggs ought to depend on the level of animal welfare related to the eggs, but as can be seen the influence from store and producer is also important. Table 4.9 gives a general overview of the price of the different egg types.

**Table 4.9 Price in DKK per egg for different egg types**

| Egg type | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price |
|---|---|---|---|---|---|---|
| Battery eggs | 9,842 | 128,425 | 0.25 | 2.99 | 1.25 | 0.279 |
| Barn eggs | 3,917 | 44,392 | 0.28 | 3.33 | **1.56** | 0.414 |
| Free-range eggs | 3,875 | 57,097 | 0.27 | 3.16 | **1.50** | 0.545 |
| Organic eggs | 6,369 | 55,793 | 0.33 | 4.33 | 1.96 | 0.460 |
| *Total* | *24,003* | *285,707* | *0.25* | *4.33* | *1.53* | *0.497* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. *Not* including data on eggs that are free-range, possibly organic too. A purchase may include more than one tray of eggs, but only one type of eggs. Special offers are included in the price calculations.

Note that the mean price of free-range eggs is *lower* than the mean price of barn eggs, whereas the opposite would be expected. The prices vary a great deal, especially the price of free-range eggs.

If one simply looks at the mean of the prices of different types of eggs in different stores (Table 4.10),[40] one finds that the mean price of barn eggs is only higher than the mean price of free-range eggs in the small stores such as 'greengrocers etc' and in Corner stores and 'Other stores'. This is consistent with the fact that all of these aggregates can be expected to be very heterogeneous in both prices and variety. As seen in Table 4.7, the market shares for these stores and for Irma are very small. Irma and 'greengrocers etc' can be expected to attract a special kind of customers as this was also the result when looking at the background data in section 4.2. They are therefore kept in the data set although they can be expected to cause some problems. Corner stores and 'Other stores', however, merely represent two small, and very heterogeneous aggregates, and are therefore deleted from the data.

[40] Details about the prices in Table 4.10 can be found in appendix E.

**Table 4.10 Mean price in DKK per egg for each aggregated store**

| **Aggregated store** | **Battery eggs** | **Barn eggs** | **Free-range eggs** | **Organic eggs** |
|---|---|---|---|---|
| SuperBrugsen | 1.36 | 1.62 | 1.97 | 2.24 |
| DagligBrugsen | 1.25 | 1.84 | 2.26 | 2.52 |
| Kvickly and OBS | 1.38 | 1.68 | 2.07 | 2.25 |
| Irma | 1.06 | 2.04 | 2.32 | 2.39 |
| Fakta (Discount) | 1.17 | 1.31 | 1.79 | 1.79 |
| Føtex | 1.26 | 1.80 | 1.91 | 2.17 |
| Netto (Discount) | 1.29 | 1.42 | 1.78 | 1.80 |
| Aldi (Discount) | 1.02 | 1.34 | 1.45 | 1.87 |
| Prima | 1.29 | 1.77 | 2.19 | 2.31 |
| Favør | 1.06 | 1.65 | 1.75 | 2.27 |
| Various grocers | 1.30 | 1.56 | 1.77 | 2.19 |
| Various discount stores | 1.19 | 1.37 | 1.77 | 1.97 |
| 'greengrocers etc' | 1.45 | **1.78** | **1.48** | 2.15 |
| 'Directly from farms' | 1.06 | 1.01 | 1.02 | 1.10 |
| Corner store/petrol station | 1.68 | **2.13** | **1.65** | 2.76 |
| Other stores | 1.10 | **1.81** | **1.21** | 2.30 |
| Bilka | 1.21 | 1.72 | 1.86 | 1.88 |
| *Total* | *1.25* | *1.56* | *1.50* | *1.96* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. *Not* including data on eggs that are free-range, possibly organic too. Special offers are included in the price calculations.

The price level obviously varies from store to store and, in general, the prices are lower in discount stores and higher in non-discount stores. To eliminate the effect of the price level one may look at relative mean prices (which is not necessarily the same as mean relative prices) by dividing all mean prices with the mean price of battery eggs. Table 4.11 presents the relative mean price of eggs.

**Table 4.11 Relative mean prices per egg for each egg type in each aggregated store**

| Aggregated store | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|
| SuperBrugsen | 1.00 | 1.19 | 1.45 | 1.65 |
| DagligBrugsen | 1.00 | 1.47 | 1.81 | 2.02 |
| Kvickly and OBS | 1.00 | 1.22 | 1.50 | 1.63 |
| Irma | 1.00 | 1.92 | 2.19 | 2.25 |
| Fakta (Discount) | 1.00 | 1.12 | 1.53 | 1.53 |
| Føtex | 1.00 | 1.43 | 1.52 | 1.72 |
| Netto (Discount) | 1.00 | 1.10 | 1.38 | 1.40 |
| Aldi (Discount) | 1.00 | 1.31 | 1.42 | 1.83 |
| Prima | 1.00 | 1.37 | 1.70 | 1.79 |
| Favør | 1.00 | 1.56 | 1.65 | 2.14 |
| Various grocers | 1.00 | 1.20 | 1.36 | 1.68 |
| Various discount stores | 1.00 | 1.15 | 1.49 | 1.66 |
| 'greengrocers etc' | 1.00 | 1.23 | 1.02 | 1.48 |
| 'Directly from farms' | 1.00 | **0.95** | **0.96** | 1.04 |
| Corner store/petrol station | 1.00 | 1.27 | **0.98** | 1.64 |
| Other stores | 1.00 | 1.65 | 1.10 | 2.09 |
| Bilka | 1.00 | 1.42 | 1.54 | 1.55 |
| *Total* | *1.00* | *1.25* | *1.20* | *1.57* |

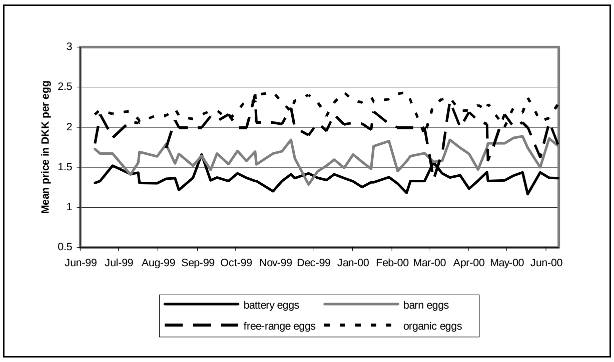
Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. *Not* including data on eggs that are free-range, possibly organic too. Special offers are included in the original price calculations.

Here it becomes clear that three of the mean prices are actually *lower* than the mean price of battery eggs, which is a problem especially in the 'directly from farms' that have 20 percent of the egg market. It is also worth noticing that stores with high absolute prices of organic eggs such as SuperBrugsen, 'Kvickly and OBS', and Føtex, do not have remarkably high relative prices. The price of battery eggs in Irma is based on a single purchase, so the relative prices are probably not very precise.

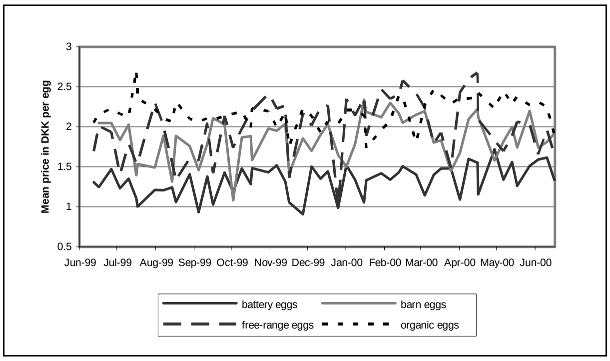
Even though the mean prices generally are consistent with the level of animal welfare related to the different egg types, the observed prices may be less consistent in a given week. The mean price for each type of egg in each store each week has therefore been calculated and investigated graphically. A few of them will be presented here.

**Figure 4.7 Examples of week by week prices in the GfK data**

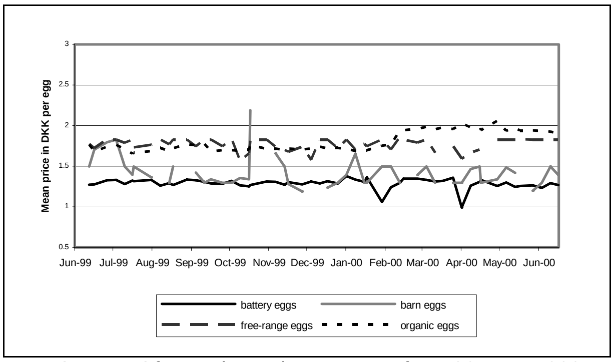
**SuperBrugsen**



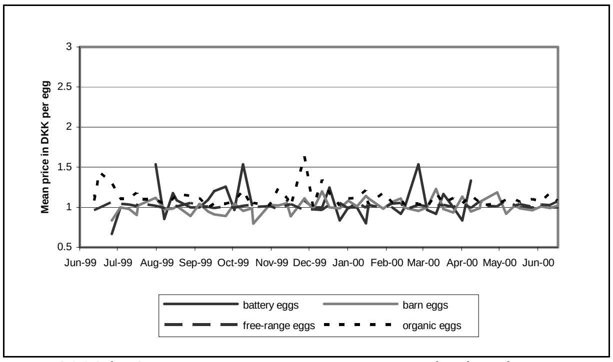

**Føtex**




**Netto (discount)**




**'Directly from farms'**




Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000 in SuperBrugsen, Føtex, Netto and 'Directly from farms'. Special offers are included in the price calculations.

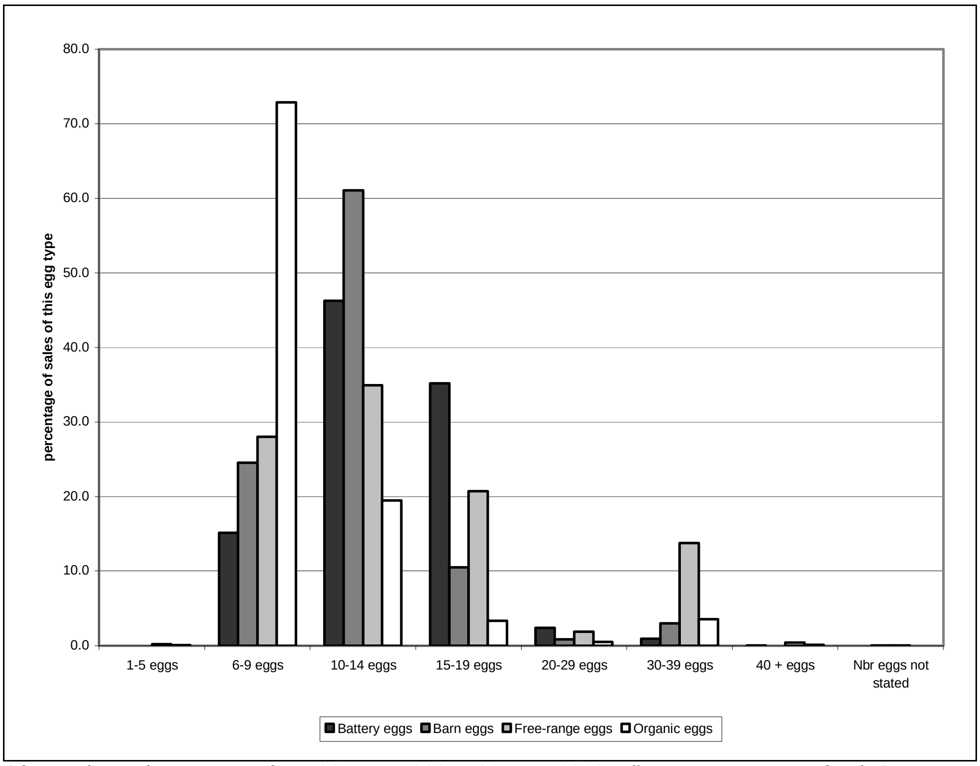
The development in prices clearly varies from store to store. SuperBrugsen has relatively stable prices, whereas Føtex has large jumps from week to week. Netto has even more stable prices for battery eggs and organic eggs, but the price of barn eggs is volatile and has holes in weeks with no observations. The price of an egg purchased directly from a farmer collapses around one DKK per egg, independent of egg type. Using the information about which store the purchase was made in therefore seems crucial.

At least three other things can be expected to influence prices. One is the *size* of the eggs in the tray. The tray is marked with the egg size which can be small, medium, large or extra large. This will obviously influence the price, since the quantity of eggs in the tray (but not the number) depends on this size. Unfortunately, this information was not recorded by GfK during the period investigated in this study, and it is therefore necessary to assume that the distribution of egg size is independent of egg type, and remember that it causes unobserved heterogeneity in prices.

According to The Danish Poultry Association, the average weight of a battery egg is 62 g and the average weight of a non-battery egg is 63 g, so differences can be expected, but no information about the distribution by classes of egg types for the different types of eggs is available in this study. Even though the means are almost identical, the distributions may differ, and it is the distributions that are of interest in this connection.

The other thing that might influence the price is the *number* of eggs in the tray. The eggs are typically (and in supermarkets, always) sold in trays with 6, 10, 12, 15 or occasionally 30 eggs per tray. In general, it would be expected that the price per egg would decrease as the tray size increases, and this might be interesting when comparing prices of different types of eggs. In Figure 4.8 it is shown how the purchases are distributed by different tray sizes. It is obvious that organic eggs are sold in six-packs far more frequently than other egg types, and that battery eggs are generally sold in bigger trays. Note the large fraction of free-range eggs that are sold in 30 egg trays. These eggs are purchased directly from a farmer, and constitute a rather large share of all free-range eggs purchased by the panel. This is probably the main reason for the low mean price of free-range eggs in Table 4.9 where the store was not accounted for.

**Figure 4.8 Distribution of sales by tray size for different eggs in the GfK data**




Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. *Not* including data on eggs that are free-range, and possibly also organic.

The huge differences in the distribution of tray sizes means that it can not be assumed that a person first decides which size of egg tray he wants, and then which type of egg. The tray size is, to some extent, a part of the egg type, not something that is chosen a priori.

The third thing that may influence prices is the egg packing company that supplies the eggs to the store. Based on the EAN codes provided by the households, GfK creates 'c-codes' for the eggs. These c-codes can be used to identify the producer[41] of the eggs.

[41] In this context the producer is actually the egg packing companies that buy eggs from various producers and pack them in trays labelled with the name of the egg packing company. Note that it is *not* the same as the producers in Table 4.5.

There are 13 professional egg producers in the data plus ‘unknown producer’ and ‘farm gate selling’, which is ‘directly from farms’ sales. Not all producers supply eggs to all stores and most stores mainly purchase eggs from one or two producers.[42]

As can be seen in Table 4.12, the small producers, in particular, concentrate on a few types of eggs, which mean that the choice of egg type and producer will not always be independent of each other, just as tray size depends on egg type.

**Table 4.12 Distribution of purchases by egg type for different producers**

| | **Distribution of purchases as *percentage*** | | | | | **Total *number* of purchases** |
|---|---|---|---|---|---|---|
| | **Battery eggs** | **Barn eggs** | **Free-range eggs** | **Organic eggs** | ***Total*** | |
| Danæg | 55 | 16 | 9 | 20 | *100* | 6,683 |
| Hedegaard/Farmæg | 62 | 4 | 8 | 26 | *100* | 5,445 |
| FDB incl. e.g. Danæg, Natura | 23 | 30 | 12 | 36 | *100* | 4,353 |
| Farmer/farm gate selling | 2 | 11 | 63 | 24 | *100* | 2,584 |
| Unknown type/brand | 22 | 36 | 25 | 17 | *100* | 1,467 |
| Brd. Honum | 66 | 9 | 6 | 19 | *100* | 1,005 |
| Æg Fra Friske Burhøns | 50 | 26 | 22 | 1 | *100* | 940 |
| Dueholm | - | - | - | 100 | *100* | 679 |
| Møllebjerggård | 14 | 14 | 7 | 65 | *100* | 277 |
| Nemli | 47 | 52 | 1 | - | *100* | 249 |
| Heslegård | 68 | 8 | 18 | 6 | *100* | 154 |
| Økologisk balance æg | - | - | 1 | 99 | *100* | 116 |
| Three producers with less than 50 purchases | 43 | 49 | 4 | 4 | *100* | 51 |
| *Total* | *41* | *16* | *16* | *27* | 100 | *24,003* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. *Not* including data on eggs that are free-range, and possibly also organic. A purchase may include more than one tray of eggs, but only one type of eggs. ‘-‘ means that purchases of this type of egg was not observed for this producer.

### *4.3.4 Consumption*

As described in section 4.3.1, the fact that the labels for non-battery eggs include different levels of e.g. animal welfare related to the different types of eggs may imply that the marginal willingness to pay varies over egg types. Section 4.3.3 focused on the price of these different types (supply), this section focuses on purchases (demand).

As shown in Table 4.2 on page 36 the number of diaries reported per family varies a great deal. Even if a family sends in diaries each week, it may well be the case that they do not purchase eggs every week. Since the observation period for eggs is only one year compared to the four years in Table 4.2, the number of reported egg purchases per family will automatically be much lower. The number of purchases per family is important because the strength of the panel structure depends on the number of families with repeated choices. In Table 4.3 on page 45 the egg data was nested in different subsamples covering a decreasing number of store aggregates. As the number of stores decreases the number of purchases per family naturally decreases too. The differences

[42] See appendix E for more on the relationship between producers, prices and stores.

between the subsamples are shown in Table 4.13. SuperBrugsen is included as an example of a very small subsample.

**Table 4.13 Number of egg purchases per family in different subsamples of the GfK data**

| Number of purchases per family | Subsample A | | Subsample B | | Subsample C | | SuperBrugsen | |
|---|---|---|---|---|---|---|---|---|
| | No. of families | Percent | No. of families | Percent | No. of families | Percent | No. of families | Percent |
| 1 | 146 | 7.5 | 173 | 9.4 | 224 | 13.2 | 167 | 34.5 |
| 2 | 119 | 6.1 | 129 | 7.0 | 162 | 9.6 | 81 | 16.7 |
| 3 | 111 | 5.7 | 126 | 6.8 | 143 | 8.4 | 44 | 9.1 |
| 4 | 110 | 5.7 | 113 | 6.1 | 118 | 7.0 | 38 | 7.9 |
| 5 | 101 | 5.2 | 92 | 5.0 | 101 | 6.0 | 32 | 6.6 |
| 6-10 | 432 | 22.3 | 410 | 22.2 | 375 | 22.2 | 61 | 12.6 |
| 11-20 | 570 | 29.4 | 505 | 27.4 | 385 | 22.7 | 48 | 9.9 |
| 21-30 | 236 | 12.2 | 203 | 11.0 | 133 | 7.9 | 6 | 1.2 |
| 31-40 | 82 | 4.2 | 65 | 3.5 | 34 | 2.0 | 3 | 0.6 |
| 41-50 | 27 | 1.4 | 25 | 1.4 | 16 | 0.9 | 2 | 0.4 |
| 51-59[43] | 7 | 0.4 | 5 | 0.3 | 2 | 0.1 | 2 | 0.4 |
| *Total* | *1,941* | *100.0* | *1,846* | *100.0* | *1,693* | *100.0* | *484* | *100.0* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. *Not* including data on eggs that are free-range, and possibly also organic. See definition of the subsamples in Table 4.3 on page 45. A purchase may include more than one tray of eggs, but only one type of eggs.

In the big heterogeneous subsamples the panel structure is very strong since more than 85 percent of the families purchases eggs more than once. In the small subsample, SuperBrugsen, only two out of three families purchase eggs more than once and only 40 percent of the families purchase eggs at least four times. In the bigger subsamples almost 70 percent of the families purchase eggs at least four times. It is obvious that the panel structure decreases with the size of the sample, but also that it does not disappear even in very small samples.

The distribution of sales/purchases by egg types differs from store to store and from household to household. The purchase share is defined as the share of *purchases* of different types of eggs, not number of eggs purchased or value of money spent on different types. Note that Corner stores and ‘Other stores’ are excluded from the data, and therefore not included in the table.

The purchase shares differ substantially from store to store. This can be caused by a combination of differences in price level as seen in section 4.3.3 and differences in socio-demographic characteristics, habits and attitudes as seen in section 4.2.2. It can therefore be expected that the marginal willingness to pay for different types of eggs may vary when using different stores in the estimations, something that will be investigated further in chapter 6.

[43] Data covers 53 weeks, so families with more than 53 weeks purchase eggs more than once a week (on average).

**Table 4.14 Purchase shares in for different egg types in different aggregated stores**

| | Total number of purchases | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|---|
| SuperBrugsen | 2,367 | 20.3 | 32.2 | 11.7 | 35.9 |
| DagligBrugsen | 634 | 44.3 | 21.9 | 14.4 | 19.4 |
| Kvickly and OBS | 1,775 | 20.3 | 32.1 | 11.4 | 36.2 |
| Irma | 221 | 0.5 | 24.4 | 22.6 | 52.5 |
| Fakta (Discount) | 1,874 | 44.9 | 17.4 | 10.5 | 27.3 |
| Føtex | 1,583 | 30.8 | 20.1 | 18.3 | 30.8 |
| Netto (Discount) | 4,625 | 63.1 | 2.2 | 5.9 | 28.7 |
| Aldi (Discount) | 953 | 70.8 | 0.6 | 1.3 | 27.3 |
| Prima | 644 | 53.7 | 15.5 | 9.0 | 21.7 |
| Favør | 512 | 60.7 | 17.8 | 7.2 | 14.3 |
| Various grocers | 3,600 | 52.9 | 20.4 | 9.9 | 16.8 |
| Various discount stores | 1,200 | 58.2 | 18.1 | 9.2 | 14.6 |
| 'greengrocers etc' | 213 | 7.0 | 13.6 | 56.8 | 22.5 |
| 'Directly from farms'. | 2,670 | 2.7 | 11.6 | 61.7 | 24.1 |
| Bilka | 849 | 41.3 | 14.3 | 12.8 | 31.6 |
| *Total* | *23,720* | *41.1* | *16.3* | *16.2* | *26.4* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. *Not* including data on eggs that are free-range, and possibly also organic. *Total* does not include non-food, Canteens, A-Z, Corner stores and 'Other stores'. (Subsample A defined in Table 4.3). A purchase may include more than one tray of eggs, but only one type of eggs.

Each family distributes its purchases over the four different types of eggs. It is therefore possible to calculate a purchase share for each type of egg for each family. In Table 4.15 the resulting purchase shares are summed up for families in subsample C (the smallest and most homogenous of the big subsamples) and for SuperBrugsen. Families with less than four purchases are excluded in the table to avoid 'false' observations of 100 percent purchase shares. If a household only reports one egg purchase to GfK the purchase share of the type chosen will be 100, but the information will not be relevant.

Table 4.15 shows that customers in SuperBrugsen clearly tend to have higher purchase shares of non-battery eggs than the panel in general (represented by subsample C). More than 60 percent of the customers in SuperBrugsen never buy battery eggs, and almost 20 percent always buy organic eggs. In subsample C only 22 percent never buy battery eggs and only 7 percent always buy organic eggs. Brand loyalty is weak for barn eggs and free-range eggs compared to battery eggs and organic eggs, since the share of customers who persistently purchases barn eggs and free-range eggs are low. This could be caused by rationing,[44] but is also likely to be a sign of low 'brand value'. Organic eggs and battery eggs are easily distinguished from other types of eggs, but the 'image' linked to barn eggs and free-range eggs is probably less clear.

[44] If a type of egg is not available in a given choice situation, it is said to be 'rationed'.

**Table 4.15 Percentage of families in GfK data with a given purchase share of the four different egg types**

| Individual purchase shares | Battery eggs | | Barn eggs | | Free-range eggs | | Organic eggs | |
|---|---|---|---|---|---|---|---|---|
| | Sub-sample C | Super-Brugsen | Sub-sample C | Super-Brugsen | Sub-sample C | Super-Brugsen | Sub-sample C | Super-Brugsen |
| **0** | 22 | 61 | 42 | 32 | 54 | 63 | 44 | 42 |
| **]0,10]** | 6 | 3 | 12 | 5 | 13 | 7 | 8 | 4 |
| **]10,25]** | 13 | 7 | 22 | 20 | 22 | 17 | 14 | 15 |
| **]25,50]** | 14 | 11 | 16 | 16 | 8 | 6 | 10 | 6 |
| **]50,75]** | 15 | 8 | 5 | 13 | 2 | 4 | 8 | 6 |
| **]75,90]** | 10 | 3 | 1 | 6 | 1 | 1 | 6 | 7 |
| **]90,100[** | 6 | 3 | 1 | 3 | 0 | 1 | 4 | 2 |
| **100** | 14 | 5 | 1 | 6 | 0 | 1 | 7 | 18 |
| ***Total*** | *100* | *100* | *100* | *100* | *100* | *100* | *100* | *100* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000 in subsample C as defined in Table 4.3 on page 45, and SuperBrugsen. *Not* including data on eggs that are free-range, and possibly also organic. A purchase may include more than one tray of eggs, but only one type of eggs.
Note: Includes families with at least four purchases; 1,164 families in subsample C and 192 families in SuperBrugsen.

An important difference between Table 4.14 and Table 4.15 is that Table 4.14 only illustrates differences between stores, whereas Table 4.15 also shows differences between individuals. For both subsamples it is clear that the individual households have different purchasing patterns. Some families prefer battery eggs others prefer organic eggs. It is therefore reasonable to assume that people derive different levels of utility from a given type of egg, and it is important to incorporate this heterogeneity in the model when estimating marginal willingness to pay.

### *4.3.5 Conclusion on purchase data*

The data on purchases of eggs yields information about consumption and prices of four different types of eggs with identical nutritional value, but different labels. Among other things, the contents of these labels include animal welfare, which cannot be defined objectively and the perception of animal welfare may therefore differ from individual to individual. It is also likely that individuals would ascribe different values to the labels, even if they could agree on a common perception of the contents of these labels. The utility gained by purchasing a given type of egg may therefore vary between individuals for two reasons:

1. different perceptions of the contents of the labels (e.g. the ‘number of units’ of animal welfare)
2. different utility of one ‘unit’ of e.g. animal welfare

These two effects can only be untangled if further information can somehow be gained about individual values of 1 or 2. This kind of information is not available in the present study.

Investigation of prices revealed a great deal of variation between the different stores. Prices can also be expected to be influenced by the tray size, the producer and the size of the egg itself.

The typical tray size varies between types of egg which means that the choice of tray size is not independent of the choice of egg type. The typical egg type varies between producers, which also mean that the choice of producer is not independent of the choice of egg type. The egg size is unfortunately not recorded in the data, but the mean weight of a battery egg is very close to the mean weight of the other eggs.

Individual, as well as aggregate purchase patterns, differ from subsample to subsample when the subsamples are defined by the stores included in the subsample, and purchase patterns clearly differ between individuals, indicating heterogeneity in the utility of eggs.

## 4.4 Conclusion

The GfK data set contains enough information for years of study. The combination of observed purchase behaviour over long periods of time, combined with information about the socio-demographics as well as habits and attitudes of the households, provides a unique opportunity to investigate consumption of non-durables using Danish data. This chapter documents not only the purchase data, but also the background information about the households in the panel. Hopefully, this may help researchers using the GfK data in future studies.

The enormous amount of information can, however, also make it difficult to find clear patterns in the data, and it certainly makes working with the data sets a very demanding job. Months of work has gone into structuring the data and trying to get an overview of the contents and possibilities of the data. In this chapter a brief overview of the contents of the entire data set has been given (section 4.1). This study only uses purchases of eggs during a one year period and it has been established that during this period the panel is representative even when ignoring the weighting that GfK uses to increase representativeness (section 4.2.1). It has also been established that the data indicates that customers in different stores differ not only on conventional socio-demographics, but perhaps more importantly on habits and attitudes (section 4.2.2). In section 4.3 it was found that prices also vary from store to store, and that the price variation from week to week within a chain of stores can also be substantial. Purchase patterns (distribution by egg types) also vary from store to store, both on the individual level and on aggregate level.

# 5 Adjusting theory to practice

When estimating marginal willingness to pay using market data, several practical problems occur. This chapter describes the most important decisions needed before the actual estimations can be carried out, and supplies solutions to the practical problems.

First the utility function to be used in the estimations in chapter 6 and 7 will be presented and the assumptions about this utility function will be outlined (section 5.1). In section 5.2 the choice set used in the analysis will be discussed. Section 5.3 provides a solution to the problem of defining which types of eggs are realistic alternatives in a given choice situation. Section 5.4 discusses the need for imputed prices and provides a method for creating these. Section 5.5 shows how marginal willingness to pay is derived from estimation results, and discusses the assumptions needed when interpreting the estimated marginal willingness to pay. Finally, in section 5.6 the implications of normalising the utility function to allow relative prices instead of absolute prices are examined.

In general the theory presented in chapter 2 and 3 was based on the utility function of a single individual. The data presented in chapter 4 is recorded at household level, which means that the observed purchases can be the result of a decision process involving more than one individual. Modelling the decision process within the household lies beyond the scope of this working paper, and the households are therefore treated as individuals in the following.

## 5.1 The utility function

It is assumed that the households have rational continuous preferences and therefore can be described as utility maximising consumers. The GfK data does not cover purchases of durable goods and this study only uses data on purchases of eggs. It is therefore necessary to assume that consumption of eggs can be modelled without using information about purchases of other goods. As mentioned in chapter 2, weakly separable and homothetic preferences are enough to ensure this property.

Utility maximising consumers are assumed to have preferences that can be represented by a utility function. This utility function is used to maximise utility given prices and budget constraints. The result is the *indirect* utility function that yields the maximum utility level given individual preferences, individual budget constraints and market prices. Prices may also depend on the individual but in this application it is assumed that all individuals face the same prices when entering the same store. In the estimations the price parameter is an expression of utility of money, and this allows estimation of the marginal willingness to pay, as it is derived by Hanemann (1984) and discussed in chapter 2.

It is assumed that the utility function is not perfectly observable to the econometrician and a Random Utility Model (RUM) is therefore used (see chapter 2 and 3). In the random utility model the utility is expected to be the mean of the true utility plus an unobserved error term with zero mean.

In this working paper an extremely simple version of the utility function is used:

$$U_i\left(e_i, p\right) = \sum_{t=1}^{T_i} U_{it}^{e_t}\left(e_{it}, p_{jt}\right) = \sum_{t=1}^{T_i}\left(\beta^j + \beta^p p_{jt} + \varepsilon_{ijt}\right) \tag{5.1}$$

where $T_i$ is the number of observations for individual $i$, $j$ is the type of egg that was actually chosen at time $t$ by individual $i$, $e_i$ is a $T_i \times 1$ vector of all eggs chosen within the period covered by the utility function, so that $e_{it} = j$. The $T_i \times J$ matrix of prices of all egg types at all times is represented by $p$, so that $p_{jt}$ is the price of egg type $j$ at time $t$. $\varepsilon_{ijt}$ is an error term and follows an extreme value distribution. This leads to the multinomial logit model. Note that this is a mixture of a conditional and a generalised multinomial logit (defined in chapter 3) since the constant of the utility function ( $\beta^j$ ) depends on the type of egg chosen, whereas the price parameter is independent of the choice actually made, which means that, the reaction to price is assumed to be the same for all types of eggs. In more complicated versions of the model the parameters for the egg types are allowed to depend on characteristics $c$ of the individual so that $\beta^j$ becomes $\beta_c^j$.

The utility function is additive which means that preferences for eggs are homothetic. As long as the conventional multinomial logit model is used, it is assumed that all individuals have essentially the same utility functions, only with individual error terms. When turning to the mixed multinomial logit however, it is assumed that all individuals has their own set of $\beta_i^j$'s, and that these $\beta_i^j$'s are all drawn from the same distribution. Note that in this study the price parameter remains fixed (the same for all individuals), only the reactions to the egg types are mixed. The mixed multinomial logit therefore changes the utility function on which the analysis is based to

$$U_i\left(e, p\right) = \sum_{t=1}^{T_i}\left(\beta_i^j + \beta^p p_{jt} + \varepsilon_{ijt}\right) = \sum_{t=1}^{T_i}\left(b^j + \eta_i^j + \beta^p p_{jt} + \varepsilon_{ijt}\right) \tag{5.2}$$

where $\eta_i^j$ has zero mean. The individual knows both $\beta_i^j$ and $\varepsilon_{it}$, but the researcher only knows that all $\beta_i^j$'s are drawn from the same distribution and therefore estimates parameters describing the distribution of $\beta_i^j$ which is common for all individuals. The common mean is labelled $b^j$ and the individual deviation from the mean is labelled $\eta_i^j$.

As mentioned in chapter 3, the individual likelihood function in the conventional multinomial logit is

$$L_i(\beta, e_i, p) = \prod_{t=1}^{T_i}\left(\frac{\exp(U_{it}(e_{it}))}{\sum_{j=1}^{J}\exp(U_{it}(j))}\right) \tag{5.3}$$

but since the $\beta_i^j$'s in the mixed multinomial logit are assumed to be unobservable and drawn from a distribution, the likelihood function is integrated over all possible values of $\beta_i^j$ in the mixed multinomial logit.

$$L_i(\theta, e_i, p) = \int\left[\prod_{t=1}^{T_i}\left(\frac{\exp(U_{it}(e_{it}))}{\sum_{j=1}^{J}\exp(U_{it}(j))}\right)\right]\cdot G(d\beta;\theta) \tag{5.4}$$

where $G$ is the density function of $\beta$, given the parameters $\theta$ that determines the distribution from which $\beta$ is drawn.

It is important to note that the functional form of the utility function is essentially the same in both the conventional and the mixed multinomial logit, and that apart from the fact that the $\beta$'s are allowed to differ between individuals in the mixed multinomial logit (something that can also be achieved in the conditional multinomial logit, e.g. by using dummies for individuals) the main difference between the two models is that the $\beta$'s are estimated as single points in the conventional multinomial logit, and in the mixed multinomial logit the distribution from which they are drawn can be estimated. The conventional multinomial logit can be seen as a special case of the mixed multinomial logit, where all $\beta$'s follow one-point distributions. The extra information gained in the mixed multinomial logit not only improves the information about the $\beta$'s, but also eliminates IIA (see chapter 3).

The utility functions in (5.1) and (5.2) are very simple and perhaps not very close approximations of the preferences of the individual. However, this application concentrates on the challenges of the data set and of the mixed multinomial logit, instead of searching for the optimal representation of the preferences. Experimenting with the functional form of the utility function could be an interesting topic for further study.

The likelihood functions in (5.3) and (5.4) use information not only about the purchase that was made, but also about purchases *not* made. This means that the definition of the *choice set* from which the purchase was made is essential to the analysis.

## 5.2 The choice set

The choice set is the set of possible choices. In this application it is the set of possible purchases. The preferences for eggs are assumed to be weakly separable from other goods and the preferences for different egg types are assumed to be homothetic. Weak separability of preferences ensures that the distribution on egg types can be modelled given only the total quantity of eggs purchased. This quantity of eggs is influenced by purchases of other goods, but homothetic preferences for the different egg types implies that the distribution of purchases on egg types is independent of the total quantity of eggs purchased. The choice of egg type can therefore be modelled without including purchases of any other goods. This means that only the *egg* should enter the description of the choice, not the goods that may have been purchased along with the egg.

When a consumer chooses to purchase a given type of egg, a store and a time for the purchase are also chosen. Both time and store can be seen as part of the choice and therefore as dimensions in the choice set. If there were five different stores and ten different occasions on which the purchase could be made, it would mean that each type of egg would lead to 50 choices in the choice set. The egg that is purchased has many different attributes itself and the choice set would therefore, even in relatively simple settings, be extremely large. This can be avoided by conditioning on some of the dimensions in the choice set. The most critical decision is whether to condition on purchase of eggs, time and choice of store. Note that this conditioning is not a modelling constraint, it is merely a way of delimiting the data used in the estimations.

### *5.2.1 Should purchasing 'nothing' be part of the choice set?*

If the choice set is not conditioned on purchase, all zero purchases must be included in the choice set. This causes problems with the definition of the choice occasion. In principle, the choice occasion becomes continuous instead of discrete, meaning that the individual is constantly choosing between not purchasing and purchasing various types of eggs. In this particular application the nature of the original data would allow defining the choice occasion as any shopping trip made by the consumer, which solves this part of the problem. The choice of whether to buy eggs or not can, however, hardly be assumed to be independent of consumption of other goods, which means that separability breaks down. In this study the choice set is therefore conditioned on a purchase being made, and only the choice of egg type is modelled.

### *5.2.2 Time of the purchase*

If the good in question is not perishable it can be stored for a period of time, which means that purchases today can be substitutes for purchases made in the future. If a good is on sale, the customer may buy more than needed in the immediate future, and then consume out of stock instead of purchasing new items for a while. Deciding whether to buy new today or consume out of stock is then a relevant part of the decision and individual expectations about future prices must be included in the model. Eggs, however, can only be stored for a relatively short period of time (approximately a month) and the quality of the eggs decrease with time. Old eggs do not have the same qualities as fresh eggs (taste, risk of infection etc.) so often eggs will be purchased because they are needed within a few days. Purchases made in the past or the future can therefore be disregarded in the choice set, and the choice set is conditioned on the time of purchase.

#### 5.2.2.1 Defining the time of purchase

In the data the time of the purchase is recorded as a time interval of a few hours within a day. This can be combined with data for the week of the purchase and the day of the week (Monday, Tuesday etc.) to give an exact date and a rather precise hour of purchase. Using this very precise definition means that only very few purchases are made by the panel within these few hours on a given day. This would, in most cases, make it impossible to find data on alternative purchases, and a more crude time measure is therefore needed. In this study it is assumed that prices and variety are constant during a *week*, and the choice set is therefore restricted to choices within a given week.

### 5.2.3 The store in which the purchase was made

If a person is observed to purchase a tray of eggs in a given store, it is important to decide whether eggs in other stores should be part of the choice set. The question is whether the person considered the price and other attributes of the eggs in other stores when choosing which egg to purchase, or whether he simply compared the eggs confronting him in the store he had chosen to enter. If the choice set is not conditioned on the store in which the purchase was made, it means that the consumer is expected to have full information about detailed prices of all types of eggs in all stores in Denmark. It is most certainly unreasonable to assume that eggs on Skagen (the most northern part of Denmark) are perceived as substitutes for eggs in Copenhagen, and defining a set of reasonable stores from which substitutes can be purchased is very difficult and lies beyond the scope of this study.[1] The choice set is therefore conditioned on the store in which the purchase was made.

#### 5.2.3.1 Definition of the store in which the purchase was made

Only the chain of stores is recorded in the data, not the actual store in which the purchase was made. Even if information about the actual store was available, it might not yield any additional information, since it would be very unlikely that all four different egg types were purchased by the panel in the same store in the same time interval. This is necessary to ensure information about the alternative purchases. It is, therefore, important to find a store definition that is as precise as possible without creating too many holes in the information needed to describe the choice set.

Data on the geographical location of the household is also available at municipality and thereby at county level. The panel consists of approximately 2,000 families distributed all over Denmark. Examination of background data on the GfK panel in 1999 shows that half of the panel live in municipalities where only 15 families or less participate in the panel (). It is unlikely that these families will purchase all types of eggs in all frequented stores in a municipality within a given week. There are 275 municipalities in

[1] Using information from the original dataset it is possible to find all stores frequented by the individual household. If the stores are frequented 'often enough', they could be considered to supply substitutes to the egg that was actually chosen. The number of alternatives using this method would vary systematically between individual households, and may, for some households, be a very large number. It is also very uncertain if the alternatives defined this way will capture the alternatives as they are perceived by the individual making the purchase, and it will most certainly increase the complexity of the estimation enormously.

Denmark, but only 14 counties and the number of panel families in each county is 70 or more except for Bornholm which is a relatively small Danish county. Geographical information could be used to define the store more precisely than just the aggregated store alone, but it may again lead to lots of missing information about alternative egg types and false observations if a family purchases eggs outside their own geographical area.

Table 5.1 shows the percentage of observations where no information about a given type of egg is available within the same store and the same week, given three different definitions of the store.

**Table 5.1 Missing information given different levels of precision of the store definition**

| Percentage of observations where no information about this type of egg is available within the same week and… | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|
| …aggregated store and municipality | 39.9 | 72.1 | 80.4 | 56.9 |
| …aggregated store and county | 19.8 | 47.4 | 57.1 | 32.1 |
| …aggregated store | 1.9 | 10.1 | 9.0 | 1.2 |

Source: Calculations based on GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except Kiosks, ‘other stores’, non-food stores, Canteens, A-Z and ‘directly from farms’. (Subsample B defined in Table 4.3 in chapter 4).

It is natural that the level of missing information is higher for barn eggs and free-range eggs since they have a lower market share (in subsample B the purchases are distributed as 46 percent battery eggs, 17 percent barn eggs, 10 percent free-range eggs and 27 percent organic eggs). It is, however, clear that including county or municipality in the store definition creates an additional difference between the share of missing observations for battery eggs and organic eggs.

The more observed purchases of a given type of egg there are in a given store in a given week, the better is the information about this type of egg. Table 5.2 shows the distribution of the number of observations. If there is only one observation, the information is highly vulnerable to errors, if there are many observations; the influence of potential errors is reduced.

**Table 5.2 Number of similar observations, only cases where at least one purchase is observed**

| Number of observed purchases | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|
| **In the same aggregated store, week and municipality** | | | | |
| 1 | 46.7 | 45.2 | 40.1 | 41.8 |
| 2-5 | 33.4 | 32.6 | 39.2 | 32.5 |
| More than 5 | 19.9 | 22.2 | 20.7 | 25.7 |
| *Total* | *100.0* | *100.0* | *100.0* | *100.0* |
| **In the same aggregated store, week and county** | | | | |
| 1 | 22.2 | 34.9 | 34 | 26.1 |
| 2-5 | 45.3 | 41.6 | 44.1 | 38.1 |
| More than 5 | 32.5 | 23.5 | 21.9 | 35.8 |
| *Total* | *100.0* | *100.0* | *100.0* | *100.0* |
| **In the same aggregated store and week** | | | | |
| 1 | 1.8 | 12.4 | 11.1 | 3.4 |
| 2-5 | 10.5 | 30.9 | 48 | 17.8 |
| More than 5 | 87.7 | 56.7 | 40.9 | 78.8 |
| *Total* | *100.0* | *100.0* | *100.0* | *100.0* |

Source: Calculations based on GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except Kiosks, 'other stores', non-food stores, Canteens, A-Z and 'directly from farms'. (Subsample B defined in Table 4.3, chapter 4).

Even if only the crude geographical categorisation, counties, is used, the fraction of single observations is between 22 and 35 percent, which is alarmingly high. Estimations where information from these single observations is used will be vulnerable to errors. If the geographical distribution of the stores is not included, the share of single observations drops dramatically to between 2 and 12 percent. As mentioned above, the difference between barn and free-range eggs on the one hand and battery and organic eggs on the other hand comes from the difference in purchase shares and can not be eliminated.

Based on the information in Table 5.1 and Table 5.2, the 'store of purchase' is defined as the aggregated store, and the geographical distribution is disregarded in the present study.

### *5.2.4 Attributes of the egg*

Conditioning on store and time, the egg purchased can be defined by the type of egg, the producer of the egg, the tray size, the egg size and the production date. If all of these factors are known, the egg should be almost perfectly defined (apart from cracked eggs in the tray, presentation on the shelf etc.). *Egg size* and *production date* are unfortunately not recorded in the data, so these factors can not be used in the estimation.[2] In chapter 4 it was found that the size of the *egg tray* is not independent of the egg type. First, this means that the choice set can not be conditioned on tray size, since the choice of tray size is correlated with the choice of egg type. Second, it means that it is not reasonable to assume that all types of eggs have been present in all sizes of egg trays. Allowing all combinations of tray size and egg type to enter the choice set would therefore not be a good description of the actual choice set. Like for the definition of stores, the problem is to define relevant substitutes, and again to keep the number of choices in the choice set down in order to make estimations feasible. The tray size is therefore ignored in the estimations.

If the *producer* formed part of the choice it would require knowledge of the producers present in the store and it would be difficult to reveal whether all types of eggs were offered by each producer in a given store within a given week. The only way to include producers would therefore be to condition on the chosen producer. However, this seems behaviourally unreasonable, and in practice it will be very difficult to find purchases of alternative egg types from the same producer. The producer is therefore also ignored.

The eggs in the choice set are therefore described by their type (defined as battery eggs, barn eggs, free-range eggs and organic eggs) and their price per egg. The type is easy to identify but problems with rationing can complicate the number of types entering a specific choice.

### 5.3 Shortage of supply (rationing)

When the choice set is defined as four types of eggs in a given store within a given week, it means that information about every possible choice is not always available, since not all types of eggs are purchased by the panel in each aggregated store every week, as shown in Table 5.1. This may occur as a result of shortage of supply (rationing) and if this is the case it is important to include the rationing in the model. In the estimations it is possible to ration choices so that not all choices are present in all choice situations and the problem is therefore not how to include varying choice sets in the estimations, but rather how to define the appropriate alternatives to a given choice.

Unfortunately (but quite reasonably) the households only report the purchase that was actually made, not the types of eggs that were available. In the estimations the best possible measure of rationing is to assume that a type of egg that was not purchased in a given aggregated store within a given week was not available at all within this week. According to Table 5.1 this happens for battery eggs in 1.9 percent of all purchases in

---

[2] In Denmark the eggs trays are marked with the size of the eggs defined in categories such as Small, Medium, Large and Extra Large. According to The Danish Poultry Association ('Dansk Fjerkræråd') an average battery egg weighs 62 g, and an average non-battery egg weighs 63 g. Therefore, one could expect non-battery eggs to fall into the higher weight classes a bit more often than the battery eggs.
There could potentially be a difference in the processing of different egg sizes for different types of eggs. Not all eggs are sold to consumers as shell eggs; some are used in other products such as cakes pasteurised eggs etc. It is possible that e.g., small organic eggs are used in other products more often than small battery eggs, which would shift the average size of the organic eggs sold as *shell eggs* upwards compared to battery eggs.

subsample B (defined in Table 4.3, chapter 4), 10.1 percent for barn eggs, 9.0 percent for free-range eggs and 1.2 percent for organic eggs. The percentages depend on the size of the subsample used in the estimations.

This way of revealing rationing is very crude. First, the fact that an egg type was not purchased by the panel in a given store in a given week, does not necessarily mean that the eggs were not present in the aggregated store in this particular week. The panel is too small to cover all types of eggs in all aggregated of store every week. Second, the fact that a free-range egg was purchased in 'Dagligbrugsen' in Copenhagen does not necessarily mean that free-range eggs were present in 'Dagligbrugsen' in a small Jutlandic city, or even in 'Dagligbrugsen' in the other end of Copenhagen. Therefore this way of revealing rationing will most probably often delete egg types in cases where they were actually present, and fail to delete egg types in cases where they were rationed. It is however the best possible method given the data available.

The distribution of the sales varies a great deal from aggregated store to aggregated store, as can be seen in Table 5.3.

**Table 5.3 Distribution of purchases by egg types for each aggregated store**

| **Percentage of sales in aggregated store** | **Battery eggs** | **Barn eggs** | **Free-range eggs** | **Organic eggs** | ***Total*** |
|---|---|---|---|---|---|
| Superbrugsen | 20 | 32 | 12 | 36 | *100* |
| Dagligbrugsen | 44 | 22 | 14 | 19 | *100* |
| Kvickly and OBS | 20 | 32 | 11 | 36 | *100* |
| Irma | 1 | 24 | 23 | 53 | *100* |
| Fakta (Discount) | 45 | 17 | 11 | 27 | *100* |
| Føtex | 31 | 20 | 18 | 31 | *100* |
| Netto (Discount) | 63 | 2 | 6 | 29 | *100* |
| Aldi (Discount) | 71 | 1 | 1 | 27 | *100* |
| Prima | 54 | 16 | 9 | 22 | *100* |
| Favør | 61 | 18 | 7 | 14 | *100* |
| Various grocers | 53 | 20 | 10 | 17 | *100* |
| Various discount stores | 58 | 18 | 9 | 15 | *100* |
| Greengrocers etc. | 7 | 14 | 57 | 23 | *100* |
| Bilka | 41 | 14 | 13 | 32 | *100* |
| *In all stores:* [3] | *46* | *17* | *10* | *27* | *100* |

Source: Calculations based on GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except Kiosks, 'other stores', non-food stores, Canteens, A-Z and 'directly from farms'. (Subsample B defined in Table 4.3, chapter 4).

In most of the aggregated stores it is reasonable to expect all types of eggs to be present in all weeks. But as some aggregates have very low shares of certain egg types, rationing must be expected here. For example, Aldi has a share of purchases of barn

[3] Sales directly from farms are not included in the table because the available way of revealing rationing is particularly unreasonable here. It must be expected that most farmers produce only one type of egg and therefore only sell one type of egg. The shares of purchases of different types of eggs sold directly from farms are 3% battery eggs, 12% barn eggs, 62% free-range eggs and 24% organic eggs. It is clear that the farmers who sell eggs directly to the consumers are not randomly chosen from all Danish farmers, since the distribution of egg types is far from the average distribution.

eggs and free-range eggs that is so low, that rationing can be expected to be quite frequent. Some of the aggregates consist of stores that must be expected to be very heterogeneous, and the available way of revealing rationing is likely to be even cruder here.

### *5.3.1 Implications of unobserved rationing*

If rationing occurs but is not revealed it might mean that a person is perceived as choosing not to buy a specific type of eggs even though this type might have been preferred if it was present. This will lead to a lower marginal willingness to pay for this type of egg and might contribute to an estimated negative marginal willingness to pay. 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.

Once the eggs entering the choice set are defined, the price of the eggs is needed for calculations of marginal willingness to pay. The definition of these prices turns out to be very important in the actual analysis, and will therefore be discussed in some detail in the following section.

## 5.4 Prices

For each observed purchase, the price paid is recorded by the household. The prices of alternatives are not recorded, but other families may have purchased eggs that can be used to describe the alternatives. In the present study the mean of the observed prices is used as an imputed price of alternative purchases.

### *5.4.1 Which prices should be imputed*

For the eggs that were not purchased, it is necessary to use the prices observed in purchases made by other members of the panel to create an imputed price. The question is what to do with prices of observed purchases.

As mentioned above, there are many unknown attributes of the purchased egg. Apart from the missing egg size and the imperfect information about the store in which the purchase was made, one could, for instance, mention the freshness of the eggs. 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 with mean prices for the eggs that were not purchased (by this family on this occasion) means that one compares the price of an egg with 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 (within a store aggregate) and a mixture of different degrees of freshness.[4] This will disturb the estimated effect of the prices, and thereby the estimated effect of the brands and other variables entering the model.

[4] The degree of freshness is not expected to influence the *price* of the eggs, but rather the *choice* of which eggs to buy.

For this study many different models (different subsamples, different level of detail in the price definition) were tested on data using observed prices for the purchase made, and imputed prices for alternative purchases. They all gave the result that people would pay significantly less for barn and free-range eggs compared to battery eggs and sometimes a bit less for organic eggs compared to battery eggs, sometimes almost the same. These results seemed counter-intuitive since they turned up even when looking only at families with a high share of organic purchases, or when looking only at stores with low shares of battery eggs-sales. Using imputed prices for all egg types did not solve the problem completely, but clearly reduced it. This application therefore uses a set of prices where all prices are imputed.

### 5.4.2 *The imputed prices*

It is crucial that the imputed prices are as close to the real prices as possible, but it is equally crucial that the prices of the different types of eggs are *equally* close to the true prices, and that the elimination of unobserved heterogeneity is the same. It will usually be impossible to achieve these three goals at the same time since some types of eggs are sold more frequently than others, as can be seen in Table 5.4.

**Table 5.4 Distribution of purchases by egg types**

| Egg type | Percentage of all purchases |
|---|---|
| Battery eggs | 46 |
| Barn eggs | 17 |
| Free-range eggs | 10 |
| Organic eggs | 27 |

Source: Calculations based on GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except Kiosks, ‘other stores’, non-food stores, Canteens, A-Z and ‘directly from farms’. (Subsample B defined in Table 4.3, chapter 4).

This means that a more precise definition of the eggs entering the imputed prices may increase the precision of the prices of, e.g., battery eggs, but if the precision of the prices of the free-range eggs are not increased too, it will not help the estimation. It was these considerations that lead to the choice of one week as the time unit and the aggregated store as the store definition in section 5.2.2 and 5.2.3. Using these definitions, the imputed price of egg type $j$ at time $t$, for individual $i$ who purchased egg type $j'$ in store $s$ at time $t$ is calculated as the mean price per egg of all purchases of egg type $j$ in store $s$ at time $t$:

$$\hat{p}_{ijts} = \frac{\sum_i p_{ijts}}{\sum_i 1_{ijts}} \tag{5.5}$$

where $\sum_i 1_{ijts}$ is the total number of purchases of egg type $j$, in store $s$ at time $t$.

Using the price per egg in the calculations means that the prices are not weighted in any way, neither by number of eggs purchased nor by value of eggs purchased. Each observed purchase contributes equally to the imputed price.

Whether the simple mean is a good way of imputing prices can be investigated by a regression of the observed prices on the imputed prices.

$$p_{ijt} = \pi_1 \hat{p}_{ijt} + \pi_2 + \varepsilon_{ijt} \quad (5.6)$$

This ought to lead to the parameter estimates: $\hat{\pi}_1 = 1, \hat{\pi}_2 = 0$. The $R^2$ value is a measure of the share of variance explained by the imputed prices, and the standard errors of $\hat{\pi}_1$ and $\hat{\pi}_2$ also give an indication of the differences between the imputed and the observed prices.

Data sets containing purchases of only one type of eggs were created, and for each dataset the observed price was regressed on the imputed price. To illustrate the importance of *store*, the prices were also imputed using only information about the type of egg and the week of the purchase, so that the imputed prices were means of all observed prices of this type of egg within a given week, in any store.

**Table 5.5 Results of regressions of the observed prices on imputed prices, using two different imputation systems**

| | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|
| **Imputed prices based on type of egg, aggregated store and week** | | | | |
| $R^2$ | 0.261 | 0.420 | 0.511 | 0.546 |
| standard error of $\hat{\pi}_1$ | 0.017 | 0.020 | 0.021 | 0.012 |
| standard error of $\hat{\pi}_2$ | 0.021 | 0.032 | 0.040 | 0.025 |
| **Imputed prices based on type of egg and week** | | | | |
| $R^2$ | 0.029 | 0.044 | 0.028 | 0.048 |
| standard error of $\hat{\pi}_1$ | 0.059 | 0.078 | 0.125 | 0.059 |
| standard error of $\hat{\pi}_2$ | 0.073 | 0.125 | 0.235 | 0.122 |

Source: Calculations based on GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except Kiosks, 'other stores', non-food stores, Canteens, A-Z and 'directly from farms'. (Subsample B defined in Table 4.3, chapter 4).
Note: In all regressions the parameter estimates were one and zero as expected.

Table 5.5 clearly shows that the aggregated store is tremendously important when imputing prices. The value of $R^2$ increases by a factor ten when stores are included in the imputation. The table also shows that the standard errors of the estimated parameters are generally very low so there is no doubt that they are truly one and zero as expected.

In another study it might be interesting to try to improve the precision of the prices that are now imputed using only egg type and aggregated store. One way could be to combine geography with longer time periods, which might decrease the problem of single or no observations created when introducing geography.

Another interesting way of improving the prices could be to try to account for as much unobserved heterogeneity as possible. If it is possible to establish that eggs sold on Funen are generally 0.02 DKK cheaper than eggs sold in the rest of the country, this could be used in the imputation of the prices. If an egg was purchased on Funen, all prices entering the mean and coming from other parts of the country can have 0.02

DKK subtracted to make them more ‘Funen-like’. If the eggs were actually purchased in Jutland, the prices of the eggs purchased at Funen can have 0.02 DKK added when entering the imputed price of the Jutlandic egg. This way all observed prices could enter the imputation. However, these extensions lie beyond the scope of this study, and the mean of observed prices within a given aggregated store within a given week as defined in (5.5) is therefore used as the imputed price in this study.

In the estimations the parameter for price is an expression of the utility of money. This measure can be used to calculate the marginal willingness to pay as described in chapter 2.

## 5.5 Marginal willingness to pay

The marginal willingness to pay derived in chapter 2 depends on the utility function. The median marginal willingness to pay for one egg compared to another egg is defined as the amount $(\pi_2 - \pi_1)^*$ that can be added to the price $\pi_1$ of the ‘base-egg’, leaving the person indifferent between buying one or the other egg. In chapter 2 the utility function was defined as $u(b, y; s) = v(b, y; s) + \varepsilon_b, \quad b \in \{0,1\}$ with $v(b, y; s) = \alpha_b + \beta y, \quad \beta > 0,$ $b \in \{0,1\}$, which meant that the marginal willingness to pay was expressed as the point where the probability that the utility of one egg was higher than the utility of another egg equalled 0.5:

$$\begin{aligned} &\Pr\{v(1, y-\pi_2; s) - v(0, y-\pi_1; s) > \varepsilon_1 - \varepsilon_2\} = \\ &\Pr\{\alpha_2 + \beta(y-\pi_2) - (\alpha_1 + \beta(y-\pi_1)) > \varepsilon_1 - \varepsilon_2\} = \\ &\Pr\{\alpha_2 - \alpha_1 - \beta(\pi_2 - \pi_1) > \varepsilon_1 - \varepsilon_2\} = 0.5 \end{aligned} \tag{5.7}$$

This leads to the following expression of median marginal willingness to pay:

$$\begin{gathered} \alpha_2 - \alpha_1 - \beta(\pi_2 - \pi_1)^* = 0, \quad \beta > 0 \\ \Updownarrow \\ (\pi_2 - \pi_1)^* = \frac{\alpha_2 - \alpha_1}{\beta} \end{gathered} \tag{5.8}$$

In section 5.1 the utility function was defined as

$$U_i(e, p) = \sum_{t=1}^{T_i} \left(\beta_i^j + \beta^p p_{jt} + \varepsilon_{ijt}\right) = \sum_{t=1}^{T_i} \left(b^j + \eta_i^j + \beta^p p_{jt} + \varepsilon_{ijt}\right) \tag{5.9}$$

where $\beta_i^j$ is the utility of one unit of egg type *j* for individual *i*, and $\beta^p$ is the utility of money which is assumed to be constant over individuals.

Using the utility function in (5.9) and the battery eggs as the base, (5.7) changes to

$$
\begin{aligned}
&\Pr\left\{U_{it}^{j} > U_{it}^{battery}\right\} = \\
&\Pr\left\{\beta_i^j + \beta^p p_{jt} + \varepsilon_{jit} > \beta_i^{battery} + \beta^p p_{battery,t} + \varepsilon_{battery,it}\right\} = \\
&\Pr\left\{\beta_i^j - \beta_i^{battery} + \beta^p \left(p_{jt} - p_{battery,t}\right) > \varepsilon_{battery,it} - \varepsilon_{jit}\right\}
\end{aligned} \tag{5.10}
$$

and (5.8) changes to

$$
\begin{gathered}
\beta_i^j - \beta_i^{battery} + \beta^p \left(p_{jt} - p_{battery,t}\right)^* = 0, \quad \beta^p < 0 \\
\Updownarrow \\
\left(p_{jt} - p_{battery,t}\right)^* = \frac{\beta_i^j - \beta_i^{battery}}{-\beta^p}
\end{gathered} \tag{5.11}
$$

Notice that the sign of $\beta^p$ is reversed to reflect the signs obtained by the estimations and to reflect the fact that the reaction to prices intuitively is expected to be negative. The higher the price, the lower the utility.

In the actual estimations $\beta_i^{battery}$ is normalised to zero so the resulting marginal willingness to pay for an egg of type *j* compared to a battery egg is:

$$
wtp_{battery}^{j} = \frac{\beta_i^j}{-\beta^p} \tag{5.12}
$$

If the mixed multinomial logit assumes that $\beta_i^j$ follows a normal distribution with mean $b^j$ and standard deviation $s^j$, the marginal willingness to pay will therefore also follow a normal distribution, but with mean $b^j / \left(-\beta^p\right)$ and standard deviation $s^j / \left|\beta^p\right|$ (which means that the mean can be either positive or negative, but the standard deviation must always be interpreted as positive).

The estimated parameters, $b^j$, $s^j$ and $\beta^p$, all have standard errors. These standard errors are usually ignored when discussing marginal willingness to pay in a mixed multinomial logit (Revelt and Train 1998, Train 1998, Train 1999b and Rouwendal and Meijer 2001) but Hensher and Greene (2001) proposes to incorporating the sampling variance by using the estimated covariance matrix of the parameters along with the point estimates, $b^j$, $s^j$ and $\beta^p$, in a simulation of the distribution of marginal willingness to pay. This yields a more correct distribution of marginal willingness to pay, but requires simulations which are beyond the scope of this study. Using more complicated and more correct measures of marginal willingness to pay could be an interesting topic for further study, but in this application the point estimates are taken as given and the standard errors are ignored.

Note that the functional form of the utility function means that income does not enter the derived marginal willingness to pay. This is a result of the restrictive assumption that income enters the utility function *linearly,* as was demonstrated in chapter 2. In this study income is simply omitted from the utility function, but allowing it to enter linearly would lead to the same estimate of marginal willingness to pay.

### *5.5.1 Assumptions needed when interpreting marginal willingness to pay*

As discussed in chapter 4, the animal welfare that is signalled by the labels for the non-battery eggs, is not easily defined and the perception of animal welfare in different types of eggs may vary from individual to individual. The utility of the different types of eggs depends (among other things) on the perceived level of animal welfare in each egg type, as well as the individual utility of one ‘unit’ of animal welfare.

In the simplest possible scenario a person’s utility from a given type of egg is the product of the number of units of animal welfare perceived to be connected with the egg and the individual utility of one unit of animal welfare. When interpreting the sign of the estimated marginal willingness to pay, it is therefore necessary to make an assumption about the sign of either the level of animal welfare that each individual ascribes to the egg, or the utility that the individual gains from animal welfare. Even in more complicated cases where the individual’s utility depends on e.g. the level of animal welfare through a more sophisticated functional form than the simple product, it is necessary to make assumptions about the sign of either animal welfare or utility of animal welfare.

If no assumptions were made, a positive utility could be the result of either an individual having negative utility of animal welfare combined with a perceived negative influence on animals, or an individual having positive utility of animal welfare and perceiving the animal welfare as positive in this type of egg. The two cases can only be distinguished by gaining information about either the individual perception of animal welfare in different eggs or the individual utility of animal welfare, or by making assumptions about one of them.

In the present study it is assumed that people have positive (or zero) utility of animal welfare and that the sign of the marginal willingness to pay can be seen as an indication of the expected level of animal welfare in the different egg types. As also mentioned in chapter 4 things like the effect on health, and perhaps also risk of infection, may influence the utility of the different egg types. It is therefore also assumed that an increase in health or a decrease in the risk of infections provide positive utility.

## 5.6 Relative prices

In section 5.4, it was described how the absolute prices of all available egg types can be imputed. In chapter 4 mean prices of the egg types in different stores were presented, and it was concluded that the price level varied from store to store. Eliminating the price level by using prices relative to the battery price changed the picture so that stores with high absolute prices on organic eggs did not necessarily have high relative prices on these eggs. Using absolute or relative prices can therefore be expected to change the results of the estimations. The question is, what the implications of using relative prices are on the utility function and on the estimated marginal willingness to pay.

Relative prices mean that $p^r_{jt} \equiv p_{jt} \big/ p_{battery,t}$ is used as the price of egg *j* instead of $p_{jt}$. This changes (5.2) into

$$U_i^r\left(e_i,p^r\right)=\sum_{t=1}^{T_i}\frac{U_{it}^{e_t}\left(e_{it},p_{jt}\right)}{p_{battery,t}}$$
$$=\sum_{t=1}^{T_i}\frac{\left(\beta_i^j+\beta^p p_{jt}+\varepsilon_{jit}\right)}{p_{battery,t}}$$
$$=\sum_{t=1}^{T_i}\left(\frac{\beta_i^j}{p_{battery,t}}+\beta^p\frac{p_{jt}}{p_{battery,t}}+\frac{\varepsilon_{jit}}{p_{battery,t}}\right) \tag{5.13}$$
$$\equiv\sum_{t=1}^{T_i}\left(\tilde{\beta}_i^j+\beta^p\frac{p_{jt}}{p_{battery,t}}+\frac{\varepsilon_{jit}}{p_{battery,t}}\right)$$

which is a transformation of the original utility function $U_i\left(e_i,p\right)$.

This new utility function changes the marginal willingness to pay in (5.12) to

$$\frac{\beta_i^j-\beta_i^{battery}}{p_{battery,t}}+\beta^p\left(\frac{p_{jt}-p_{battery,t}}{p_{battery,t}}\right)^*=0,\quad \beta^p<0$$
$$\Updownarrow$$
$$\left(\frac{p_{jt}-p_{battery,t}}{p_{battery,t}}\right)^*=\frac{\beta_i^j-\beta_i^{battery}}{p_{battery,t}}\Big/-\beta^p \tag{5.14}$$
$$\Updownarrow$$
$$\left(p_{jt}^r-1\right)^*=\frac{\tilde{\beta}_i^j-\tilde{\beta}_i^{battery}}{-\beta^p},\quad \tilde{\beta}_i^k=\frac{\beta_i^k}{p_{battery,t}}$$

where $\tilde{\beta}$ is the parameter achieved from the estimation. Since $\beta_i^{battery}$ is normalised to zero, $\tilde{\beta}_i^{battery}$ is also normalised to zero and the estimated median relative marginal willingness to pay is therefore:

$$w\tilde{t}p_{battery}^j=\frac{\tilde{\beta}_i^j}{-\beta^p} \tag{5.15}$$

Therefore, working with relative prices does not create difficulties in the estimations and the derivation of marginal willingness to pay; it merely changes the interpretation of the marginal willingness to pay. When absolute prices are used, the marginal willingness to pay measures the *amount* in DKK a person is willing to pay extra for, e.g., an organic egg compared to a battery egg. Using relative prices leads to a *relative* marginal willingness to pay where the extra amount is divided by the price of battery eggs. The relative marginal willingness to pay is therefore measured as a percentage divided by 100. If the relative marginal willingness to pay is 0.5 it means that the person is willing to pay 50 percent extra for an egg of type *j* compared to a battery egg.

## 5.7 Conclusion

Assuming that the individual utility of different types of eggs can be described by a simple linear (and thereby homothetic) utility function that is separable from purchases of other goods, yields a simple way of estimating marginal willingness to pay. The estimations are based on a choice set that is conditioned on the time and store of the purchase but ignores producers, tray size and egg size, and therefore only includes up to four choices characterised by the type of egg (battery, barn, free-range or organic) and the price of the egg. Prices and variety are assumed to be constant during the week, and prices in aggregated stores are assumed to be the same all over the country.

In the present study, the best feasible method for detecting rationing is to assume that egg types that were not purchased in a given store aggregate within a given week were not available within this week. This is a very crude method that will surely assume rationing in cases where there was none, and reject rationing in cases where rationing occurred, but it is the best method possible given the data available.

In the estimation imputed prices will be needed not only for the types that were not purchased, but also for the type that was actually purchased. Otherwise the results will be biased by unobserved heterogeneity in prices. The prices are imputed as the mean of all prices observed for that type of egg in that store aggregate within that week.

In this study the mean and standard deviation of the distribution of marginal willingness to pay is calculated as the parameters of the distribution of the reaction to egg type divided by the negative of the parameter of price. This ignores the standard errors of the estimated parameters. Including this variance is an interesting topic for further study.

Finally, using relative prices instead of absolute prices means that the utility function in each time period is divided by the price of battery eggs within that specific time period. It is assumed that consumption can not be substituted over time and the fact that utility varies over time is therefore irrelevant when making the actual choice.

It is now possible to proceed with the actual estimations.

# 6 SuperBrugsen

Chapter 2 and 3 presented the theoretical foundation for the estimations. Chapter 4 presented the data and chapter 5 provided solutions to practical problems arising when theory is combined with practice. This chapter proceeds with the actual estimations. The multinomial logit model used here is a mixture of a conditional (the price parameter is independent of choice) and a generalised multinomial logit (there is a parameter for each egg type).[1]

Section 6.1 presents results of estimations using the conventional multinomial logit model. These results are used as starting values when proceeding to the more flexible mixed multinomial logit. In section 6.2 estimations are conducted using the mixed multinomial logit and this increases the likelihood function dramatically in all models estimated, indicating that the mixed multinomial logit is significantly better than the conventional multinomial logit in this study. Section 6.2.1 discusses the choices that must be made when using mixed multinomial logit, and in section 6.2.2 the results of the mixed multinomial logit are compared with the ones obtained using the conventional multinomial logit in section 6.1. Section 6.2.3 examines models where reaction to egg type depends on background variables yielding information about the household in general. Geographical location of the household residence improves the model significantly, while age of the main buyer and attitude to branded goods do not. Section 6.2.4 sums up the results of the mixed multinomial logit and section 6.3 sums up this entire chapter.

Based on the description of the four different types of eggs in chapter 4 it is expected that marginal willingness to pay, compared to battery eggs, will be *positive* for all three types of non-battery eggs, and that marginal willingness to pay will be higher for free-range eggs than for barn eggs. Marginal willingness to pay for organic eggs is expected to be higher than marginal willingness to pay for the two other types because the 'Ø-label' (indicating that a good is organic and guaranteed by a governmental authority) is a familiar label that is used on a wide variety of goods, and apart from animal welfare also indicates environmentally friendly production, and to some consumers also healthier products or taste.[2] The parameter for price is expected to be *negative*, indicating that the utility of money is positive. Or put in another way: An increase in the price of the product implies a decrease in utility.

---

[1] As in the previous chapters, the four different egg types are: Battery eggs, barn eggs, free-range eggs and organic eggs.

[2] According to Statistics Denmark (2002) three quarters of consumers purchasing organic goods in general state that animal welfare is important or very important for their choice of organic goods over conventional goods. Environmental benefits are also important or very important to three quarters of the households purchasing organic goods. Sixty percent of these consumers also find personal health important or very important. The taste of organic products compared to conventional products is only important or very important to forty percent of the families, whereas it is of no importance at all for another forty percent. The value of the 'Ø-label' is therefore clearly composed of several different effects.

According to Statistics Denmark (2002) approximately 70 percent of all Danish households buy organic goods at least once in a while. Approximately thirty percent of these families state that they are unwilling to pay extra for organic goods. Combining these figures leads to the conclusion that approximately half of all Danish families have positive willingness to pay for organic goods in general. If this is taken as an indication of the distribution of willingness to pay for organic eggs or other non-battery eggs it means that the median of the distribution is expected to be close to zero. Investigating background characteristics (in chapter 4) revealed differences in both socio-demographics and attitudes for the costumers in different chains of stores. It is expected that price level, variety and general image varies between chains of stores. Some stores focus on quality and ethics (e.g. SuperBrugsen) other on low prices (e.g. Bilka). These differences may mean that the chains attract different costumers, and this could be the reason for the observed differences among costumers. If marginal willingness to pay for different types of eggs is correlated with one or more of the background characteristics it means that the marginal willingness to pay will vary between different chains of stores. Given that median marginal willingness to pay for e.g. organic eggs is expected to be close to zero, this median will be positive in some chains of stores, and negative in others. This chapter focuses on SuperBrugsen as an example of a chain of stores where marginal willingness to pay is expected to be higher than in the general population, and where shortage of supply is expected to be rare. Chapter 7 briefly presents results from other stores.

## 6.1 The conventional multinomial logit for panels

When estimating mixed multinomial logit models (MMNL), it is important to always start out with a quite conventional multinomial logit model. First, it is important to get an indication of the unrestricted signs of the parameters, especially if the log-normal distribution is to be assumed, since the functional form at the log-normal distribution restricts the parameter to be positive. Second, Train (1999c) recommends that the results of the conventional multinomial logit are used as starting values for the mixed model.

The estimation program in Train et al. (1999) has been used to estimate the conventional multinomial logit model.[3] As described in chapter 5, the prices are imputed by taking the mean of all observed prices of a particular type of eggs in a particular store within a particular week as described in chapter 5. This means that the customer is expected to compare prices only of the eggs presented in the store where the purchase was made. See chapter 5.2.3 for more on the definition of the choice set and the imputed prices. Battery eggs are used as the base, so that parameters for barn eggs,

[3] Let all parameters be fixed to achieve a conventional logit model, and set the number of repetitions to 1 to save time in the estimation. For more information about the program see chapter three. To confirm the accuracy of the GAUSS estimations, they have been compared with results using Limdep. Limdep did not allow for robust standard errors but when 'usual' standard errors were used in the GAUSS program, they turned out to be half the size of the ones obtained using Limdep. The estimated parameters were the same when measured to two decimals. The estimations are conducted using the program in Train et al. (1999) because it uses the panel structure which Limdep ignores. Results of the Limdep estimations are presented in Table G.1 and Table G.2 in appendix G.

free-range eggs and organic eggs could be estimated, allowing calculation of the marginal willingness to pay for these egg types compared to the marginal willingness to pay for battery eggs.[4]

The conventional multinomial logit model is estimated using information about purchases only, not information about the households making the purchases. The purchase information available in the data is price, egg type chosen, store of purchase and whether the purchase was made on sale or at the regular price. The ‘on sale’ parameter seems to be too closely correlated with the price variable and it is therefore disregarded in all estimations.[5] Estimations on data including purchases in many different chains of stores indicate that using relative prices work better than using absolute prices. Estimations using the measurement of rationing developed in chapter 5 improve the results of estimations using data from stores with frequent rationing.

As can be seen in Table 6.2 relative prices do not change the essence of the results in SuperBrugsen. Rationing is not very frequent in SuperBrugsen (at least not when measured in this way) and as expected, introducing it in the model hardly affects the results at all.[6]

---

[4] Estimations could also be conducted using other types of eggs as the base. In general changing the base may lead to different estimates and the effect of different base egg types is therefore investigated. In appendix G, Table G.3, the results of four separate estimations, using different egg types as the base are presented. The marginal willingness to pay for non-battery eggs, compared to battery eggs, differs slightly on the third decimal when the base egg type changes, but for all practical purposes it does not matter which egg type is used as the base. The value of the likelihood function is also the same. Estimations using data from subsample C or data from Bilka yields the same result.

[5] Including the ‘on sale’ variable in estimations using relative prices, as defined in chapter 5, lead to positive price parameters in both subsample B and C (defined in chapter 4, Table 4.3), regardless of whether rationing is included in the model or not.

[6] Using relative prices, in stead of absolute, do not change the optimal value of the log-likelihood function, but introducing rationing increases the value by 3, from -3,075 to -3,072. This is not unexpected since deleting an option automatically increases the probability of the chosen alternative, and thereby the value of the likelihood function.

**Table 6.1 Estimated parameters using different specifications of the conventional multinomial logit**[7]

| ***Parameter* for:** | **Price** | | **Barn eggs** | | **Free-range eggs** | | **Organic eggs** | |
|---|---|---|---|---|---|---|---|---|
| a) Absolute prices | -1.07 | ** | 0.75 | ** | 0.13 | | 1.52 | ** |
| *(Standard error)* | *(0.225)* | | *(0.168)* | | *(0.236)* | | *(0.257)* | |
| b) Relative prices | -1.29 | ** | 0.73 | ** | 0.06 | | 1.42 | ** |
| *(Standard error)* | *(0.281)* | | *(0.170)* | | *(0.238)* | | *(0.251)* | |
| c) Relative prices, rationing | -1.28 | ** | 0.73 | ** | 0.08 | | 1.41 | ** |
| *(Standard error)* | *(0.280)* | | *(0.170)* | | *(0.238)* | | *(0.250)* | |

Source: Estimations using GfK purchase data on eggs in SuperBrugsen from 26 June 1999 to 30 June 2000.
'**' means that the parameter is significantly different from zero at the one percent level.
Notes for the different approaches:
a: The mean price of a battery egg in SuperBrugsen is 1.36 DKK.
b: Using relative prices means that the price of a battery egg is always one.
c: Rationing is measured as described in Chapter 5.

The estimated parameters lead to the different values of marginal willingness to pay presented in Table 6.2.

**Table 6.2 Marginal willingness to pay using different specifications of the conventional multinomial logit**

| **Approach:** | **Barn eggs** | **Free-range eggs** | **Organic eggs** |
|---|---|---|---|
| a) Absolute prices in DKK | 0.70 | 0.12 | 1.42 |
| b) Relative prices | 0.57 | 0.05 | 1.10 |
| c) Relative prices, rationing | 0.57 | 0.06 | 1.11 |

Source: Estimations using GfK purchase data on eggs in SuperBrugsen from 26 June 1999 to 30 June 2000.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated marginal willingness to pay therefore has no standard error.

Using relative prices and rationing (approach c) means that people are willing to pay 57 percent *extra* for a barn egg, 6 percent extra for a free-range egg and 111 percent extra for an organic egg. The parameter for free-range eggs is 0.08 (see Table 6.1), but the standard error of 0.238 means that it is not significantly different from zero and the positive willingness to pay should therefore be interpreted with care. In chapter 4, the mean of the observed absolute prices and the relative values of these means were presented. Using imputed prices, as described in chapter 5, changes the means slightly, and taking the mean of the relative imputed prices leads to a slightly different result than calculating the relative mean of the observed prices. Table 6.3 show the means of the observed absolute prices (as presented in chapter 4) along with the means of the means of the imputed absolute prices and the means of the relative imputed prices.

[7] This table presents standard errors for the first time and in this connection it is important to note that McFadden and Train (2000) emphasise (page 455) that for a finite number of repetitions/draws the usual standard error 'may substantially underestimate the covariance of the estimator', and they therefore recommend using robust standard errors. Comparing estimated 'usual' standard errors with estimated robust standard errors on these data show that the robust standard errors are between four and seven times bigger than the non-robust standard errors, and robust standard errors are therefore used throughout this whole study. However the standard errors may still be underestimated since the imputed prices are taken as given in the estimations, thereby ignoring the variance of these prices. Including this variance in estimations is an interesting topic for further research but is ignored in this study.

**Table 6.3 Mean prices of different egg types in SuperBrugsen**

| Mean price | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|
| Observed absolute prices | 1.36 | 1.62 | 1.97 | 2.24 |
| Absolute imputed prices | 1.36 | 1.63 | 2.01 | 2.25 |
| Relative imputed prices | 1.00 | 1.20 | 1.48 | 1.66 |

Source: Calculations using GfK purchase data on eggs in SuperBrugsen from 26 June 1999 to 30 June 2000.

The mean relative imputed price of a barn egg is 1.20 in SuperBrugsen which means that on average a barn egg costs 20 percent extra compared to a battery egg. The estimated marginal willingness to pay for this type of eggs compared to battery eggs is 57 percent and thus higher than the observed. The mean imputed relative price of an organic egg is 66 percent higher than the price of a barn egg, so the estimated marginal willingness to pay of 111 percent is almost twice as high as the observed. For free-range eggs the opposite is the case. The mean relative imputed price is 48 percent higher than the price of a battery eggs, but the estimated marginal willingness to pay compared to battery eggs is almost zero.

A similar result is found in Baltzer (2002).[8] Here the marginal willingness to pay for free-range eggs compared to battery eggs is found to be almost the same as for barn eggs (i.e. significantly different from zero) but lower than the observed price of free-range eggs. The marginal willingness to pay for barn eggs and organic eggs in Baltzer (2002) are higher than the observed prices, just as it is found in Table 6.2 above.

It is reasonable to expect the value of different labels to vary between households. Animal welfare may be very important to some households, but have little or no value in other households. These differences are perceived as 'heterogeneity of preferences' in the econometric model. The labels 'barn eggs' and 'free-range eggs' mainly indicates increased animal welfare, whereas the 'organic' label indicates a more environmentally friendly production as well as a higher level of animal welfare. Some households may also perceive the organic eggs as being healthier than other egg types because the hens are fed with organic feed. The heterogeneity of marginal willingness to pay for organic eggs can therefore be induced by differences in the perception and evaluation of at least three different attributes, whereas the heterogeneity of marginal willingness to pay for barn and free-range eggs is expected to arise only from differences in perception and evaluation of animal welfare.

The mixed multinomial logit estimates a distribution of the mixed parameters. The standard deviation can be used as a measure of the degree of heterogeneity, in the sample, related to this particular parameter. The estimated parameter distribution means that a standard deviation of willingness to pay can be calculated too.

Train (1999c) recommends that the results from the conventional multinomial logit are used as starting values for the mean of the parameters when estimating mixed multinomial logit, and that 0.1 is used as the starting value for the standard deviation of

[8] Baltzer (2002) uses data for purchases of eggs in 75 different Kvickly stores from the third quarter of 2000 to the third quarter of 2001 (both included). The data is observed on store level not on costumer level. This means (among other things) that observed prices are available for all types of eggs present in a store within a given week, and imputed prices are therefore not needed.

the parameters. The results obtained in Table 6.2 will therefore be used in the following estimations.[9]

## 6.2 Mixed multinomial logit

In chapter 4 investigations of individual purchasing patterns showed a great deal of heterogeneity between the households in the panel. This is not captured by the conventional multinomial logit model and the estimations therefore proceed using the more flexible mixed multinomial logit model (described in chapter 3) that allows heterogeneity in individual preferences by assuming that one or more of the parameters of the utility function are drawn from a common distribution. This means that individual parameters are not assumed to be identical, and that the heterogeneity is assumed to be captured by the common distribution estimated. Estimating the parameters of the *distribution* of the parameters of the utility functions yields not only a measure of the marginal willingness to pay for different types of eggs, but also a measure of the degree of heterogeneity in the population.

In section 6.2.1, the decisions needed before estimation of mixed multinomial logit models are discussed, and the choices are made. Section 6.2.2 re-estimates the model from section 6.1 in a mixed multinomial logit setting. Section 6.2.3 expands the model by including information about the households and section 6.2.4 sums up the results of the mixed multinomial logit.

### *6.2.1 Decisions needed before the actual estimation of the mixed multinomial logit model*

When mixing a multinomial logit there are, among other things, three important decisions to make:

- Which parameters should be assumed to be heterogeneously distributed?
- Which distribution(s) should be used to capture the heterogeneity?
- How many repetitions should be used in the simulation of the distribution(s)?

These questions are addressed in the following sections:

#### *6.2.1.1 Which parameters should be assumed to be heterogeneously distributed*

If a parameters is assumed to be heterogeneously distributed it is said to be ‘mixed’, hence the name ‘mixed multinomial logit’. In this study the reaction to price is never mixed, because this makes the interpretation of the marginal willingness to pay much easier. If the parameter for a given egg type is mixed with a given distribution, it is relatively easy to figure out which distribution the marginal willingness to pay follows, as long as the price parameter is fixed in a one point distribution. Allowing the reaction to both price and egg type to vary between individuals makes it difficult to calculate the

[9] Most of the models in the following have *also* been tested using 1 as the starting value for the mean of the egg parameters, and -1 as the starting value for the price. The results differed slightly but not in essence, which indicates that the estimations are not highly sensitive to changes in starting values.

distribution of marginal willingness to pay. One could also keep the reaction to egg type fixed and mix the reaction to price, but calculating the distribution of an inverted variable is more difficult than in the non-inverted case.[10] More importantly the purpose of this study is to examine people's marginal willingness to pay for different egg types. If only the price is mixed it means that the relative reactions to the egg types are the same for all people, only the *level* of the marginal willingness to pay can change. This type of results is not very interesting since one of the purposes of this study is to investigate the heterogeneity, within the population, of marginal willingness to pay for different types of eggs. Therefore, the parameter for price is always kept fixed in the following, and the parameters for the different egg types are always mixed.

#### 6.2.1.2 The distribution used in the mixing

The most important decision when choosing the mixing distribution is whether the distribution should be symmetric or not. The estimation program in Train et al. (1999) allows the parameters to follow a normal, a uniform, a triangular or a log-normal distribution (see chapter 3). The first three distributions are symmetric, meaning that the number of individuals with marginal willingness to pay lower than the mean should be equal to the number of individuals with marginal willingness to pay higher than the mean. The log-normal distribution is not symmetric, which among other things means that the mean differs from the median. In the log-normal distribution, the number of people with marginal willingness to pay lower than the *median* equals the number of people with marginal willingness to pay higher than the median. But in the log-normal distribution the distribution of marginal willingness to pay among the people with marginal willingness to pay lower than the median is not the same as among the people with marginal willingness to pay higher than the median. The shape of a log-normal distribution means that people with lower marginal willingness to pay generally will have marginal willingness to pay closer to the median, while the people with higher marginal willingness to pay can have infinitely large marginal willingness to pay. (The tail of the log-normal distribution is heavier than the tails in the normal distribution, which emphasises the problem).

The log-normal distribution has the property that it restricts the sign of the parameter to be positive. If a negative sign is needed the sign of the variable entering the model must be reversed. However, in this case restricting the reaction to an egg type to be either negative or positive seems to ruin the estimation.[11] This is supported by the fact that only 50 percent of the population state that they have positive willingness to pay for e.g. organic goods in general (Statistics Denmark 2002). The preferences for these households can not be described by a log-normal distribution at the same time as the

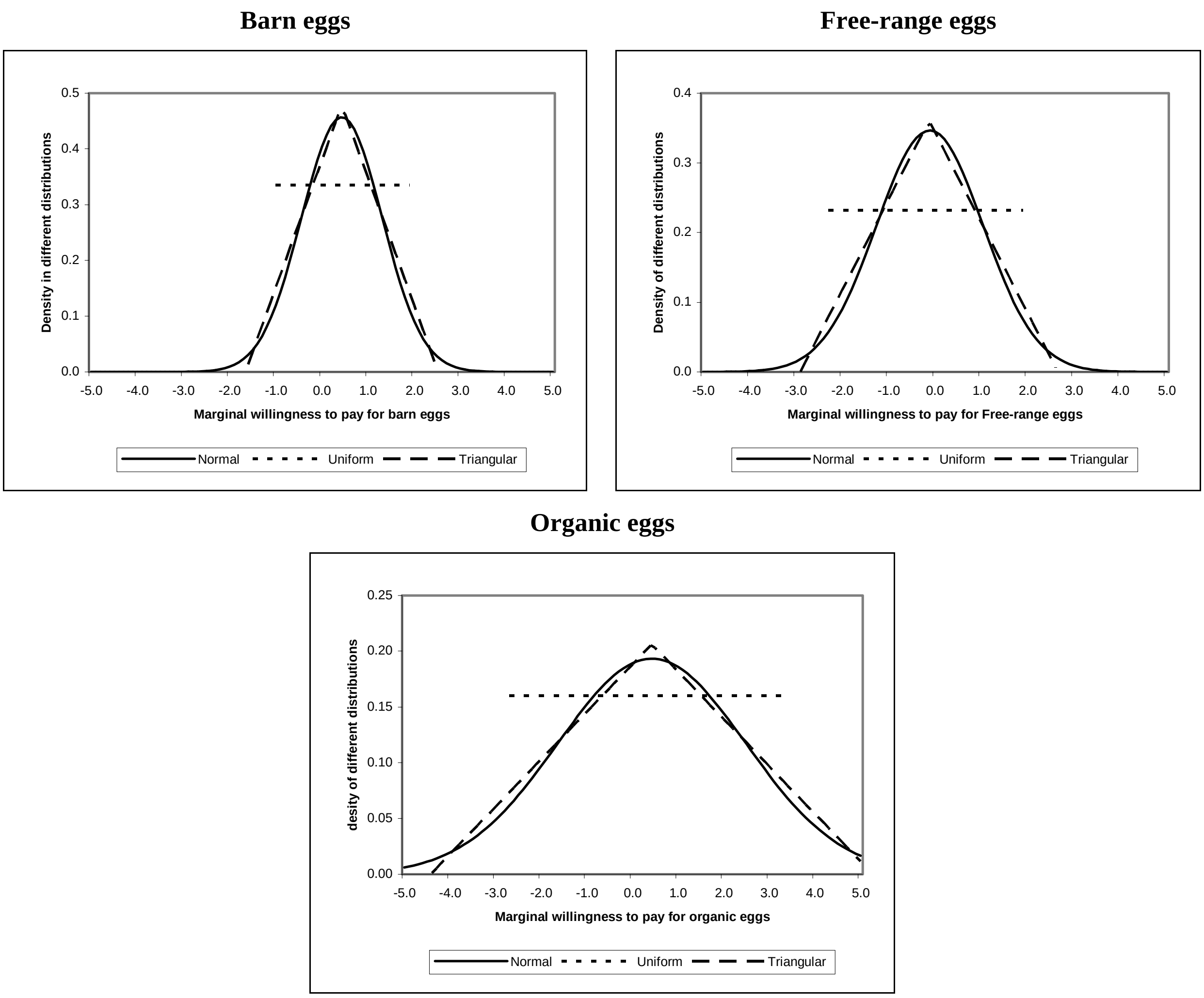
[10] The normal distribution is always defined on zero. The uniform and the triangular distribution are potentially defined on zero and will often be so. This leads to serious problems when defining the inverted distribution since this will go to infinity in zero. This problem must be considered carefully when deciding whether, and how, prices should be mixed.

[11] Estimations with log-normal parameters for the different egg types using data from SuperBrugsen with relative prices and rationing have been attempted for this study, but the Hessian turned singular during estimation, and the likelihood function could therefore not be maximised.

preferences for the 50 percent of the population that do not have positive willingness to pay. This study therefore disregards the log-normal distribution.

In order to compare the different functional forms of the distributions available in Train et al. (1999), the results of mixing the reaction to egg type with three different distributions are presented graphically in Figure 6.1. The mean of the marginal willingness to pay for free-range eggs is not significantly different from zero in any of the distributions, and the mean of the marginal willingness to pay for organic eggs is not significantly different from zero in the uniform distribution. All other parameters are significantly different from zero. The uniform distribution is naturally very different from the others, but the triangular and the normal distribution are very much alike in this case.

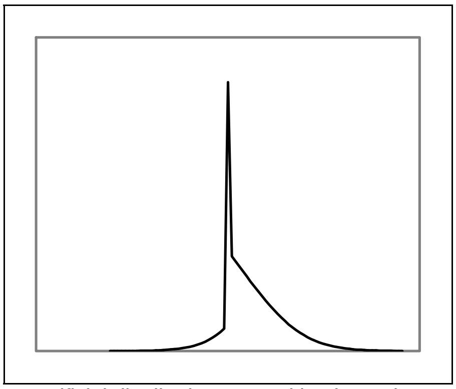
**Figure 6.1 Normal, uniform and triangular distribution of marginal relative willingness to pay for barn, free-range and organic eggs**




Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Only data from SuperBrugsen. Prices are relative which means that the relative marginal willingness to pay is measured as a percentage of the battery egg price divided by one hundred. Rationing is allowed. Number of repetitions:500, starting values: -1 for price, 1 for the mean of the reaction to egg types and 0.1 for the ‘standard deviation’ of the reaction to egg types.

In another study it could be interesting to find a distribution that recognises that some people have a positive marginal willingness to pay while others have a negative marginal willingness to pay, and that the distribution is not symmetric. Perhaps the distribution should even be a combination of a discrete and a continuous model allowing the point zero to have positive probability indicating the share of individuals that are indifferent between two egg types. Figure 6.2 illustrates a more desirable distribution. Unfortunately this type of distribution is not included in existing software for mixed multinomial logit modelling (e.g. in Train et al. 1999), so the normal distribution will be used instead.

**Figure 6.2 More desirable distribution of marginal willingness to pay**



Artificial distribution created by the author

*6.2.1.3 Number of repetitions used in the simulation*

As mentioned in chapter 5 the individual likelihood function has the form:

$$L_i(\theta, e_i, p) = \int \left[ \prod_{t=1}^{T_i} \left( \frac{\exp\left(U_{it}^{e_t}\right)}{\sum_{e_t=1}^{J} \exp\left(U_{it}^{e_t}\right)} \right) \right] \cdot G(d\beta; \theta) \tag{6.1}$$

The integral must be optimised by simulation, and the last decision regards the number of draws (or repetitions) used to simulate this integral. The precision of the estimation increases as the number of repetitions increases, but so does the time required for the simulations. Choosing the number of repetitions is therefore a matter of choosing between accuracy and feasibility.

In McFadden and Train (2000) results from Brownstone and Train (1999) are presented, using 250 repetitions. In Revelt and Train (1998) the number of repetitions is 500. A recent paper by Layton and Brown (2000) uses 1000 repetitions. Estimations using the *triangular* distribution and subsample C with relative prices and rationing using 250, 500, 1,000, 1,500 and 2,000 repetitions showed that the most significant difference lies between using 250 and 500 repetitions.[12] The present study therefore uses 500 repetitions as in Revelt and Train (1998).

[12] The results are presented in Table G.4 in appendix G.

Now that these three important decisions have been made, the estimation may proceed using the mixed multinomial logit with fixed price parameter, normally distributed parameters for the three different egg types that are compared to the base type battery eggs, and 500 repetitions. Prices are relative and rationing is allowed as described in chapter 5.

### *6.2.2 Results of mixing using only purchase data*

The conventional multinomial logit can be seen as a mixed multinomial logit in which the standard deviations of all parameters are zero. This study introduces three standard deviations in the mixed multinomial logit (one for each egg type), and it is therefore possible to conduct a likelihood ratio test with three degrees of freedom, to see if the conventional multinomial logit is almost as good as the mixed multinomial logit.

Mixing the parameters for the three egg types with the normal distribution increases the optimal value of the likelihood function from -3,072 to -1,972. The likelihood ratio test is asymptotically $\chi^2$ distributed with three degrees of freedom, and the test-probability is the probability that the conventional multinomial logit is just as good as the mixed multinomial logit. This probability is lower than 0.1 percent, indicating that the mixed multinomial logit yields a better description of data than the conventional multinomial logit.

Table 6.4 presents the estimation results. Note the difference between the standard *errors* and the standard *deviations*. The standard errors reflect the precision of the estimated parameters, whereas the standard deviations reflect the degree of heterogeneity in the population. All three standard deviations are significantly different from zero, indicating that mixing is necessary for all three parameters.

**Table 6.4 Results of mixing the reaction to egg type with the normal distribution**

| **Mean and standard deviation of the parameters for:** | **Mean** | | **Standard deviation** | |
|---|---|---|---|---|
| Price | -2.38 | ** | - | |
| *(Standard error)* | *(0.534)* | | - | |
| Barn eggs | 1.03 | ** | 2.12 | ** |
| *(Standard error)* | *(0.185)* | | *(0.174)* | |
| Free-range eggs | -0.30 | | 2.70 | ** |
| *(Standard error)* | *(0.366)* | | *(0.266)* | |
| Organic eggs | 1.18 | ** | 4.82 | ** |
| *(Standard error)* | *(0.452)* | | *(0.410)* | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Only data from SuperBrugsen.
'**' means that the parameter is significantly different from zero at the one percent level.
Prices are relative which means that the relative marginal willingness to pay is measured as a percentage of the battery egg price divided by one hundred. Rationing is allowed.
Number of repetitions: 500. Starting values for the price and the means of the mixed parameters are taken from the conventional logit estimation in section 6.1, and 0.1 is used as starting value for the standard deviation of the reaction to egg types.

The estimated distributions of the parameters for the three different egg types combined with the estimated price parameter leads to a distribution of marginal relative willingness to pay for each egg type. These distributions are described in Table 6.5.

**Table 6.5 Estimated distribution of marginal relative willingness to pay**

| Marginal relative willingness to pay for: | Mean | Standard deviation |
|---|---|---|
| Barn eggs | 0.43 | 0.891 |
| Free-range eggs | -0.13 | 1.135 |
| Organic eggs | 0.50 | 2.024 |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Only data from SuperBrugsen. Prices are relative which means that the relative marginal willingness to pay is measured as a percentage of the battery egg price divided by one hundred. Rationing is allowed.
Number of repetitions: 500. Starting values for the price and the means of the mixed parameters are taken from the conventional logit estimation in section 6.1, and 0.1 is used as starting value for the standard deviation of the reaction to egg types.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated marginal willingness to pay therefore has no standard error.

As in the conventional logit in section 6.1 the mean of the parameter for free-range eggs is not significantly different from zero, and the negative mean should therefore be interpreted with care.

Table 6.6 shows the relative imputed prices along with the marginal relative willingness to pay estimated using the conventional and the mixed logit. In the mixed logit the mean marginal relative willingness to pay for barn eggs is still higher than the observed price difference, but the interpretation of this is now that more than 60 percent have a marginal relative willingness to pay for barn egg that is higher than the mean relative imputed price of 1.20. The mean marginal relative willingness to pay for organic eggs is lower than the observed price difference, but this just means that less than 50 percent of the households have a marginal relative willingness to pay for barn egg that is higher than the mean relative imputed price of 1.66. The percentage is 47, so the difference is quite small.

**Table 6.6 Comparing marginal relative willingness to pay (wtp) in the conventional and the mixed logit with the imputed prices**

| Mean price | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|
| Imputed relative prices | 1.20 | 1.48 | 1.66 |
| Wtp using conventional logit | 0.57 | 0.06 | 1.11 |
| Mean wtp using mixed logit | 0.43 | -0.13 | 0.50 |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Only data from SuperBrugsen. Prices are relative which means that the relative marginal willingness to pay is measured as a percentage of the battery egg price divided by one hundred. Rationing is allowed.

It is not possible to say anything in general about the relationship between the parameters of the conventional logit and the mean of the parameters in the mixed logit,[13] but in this particular case the mean is generally lower than the parameter in the conventional logit.

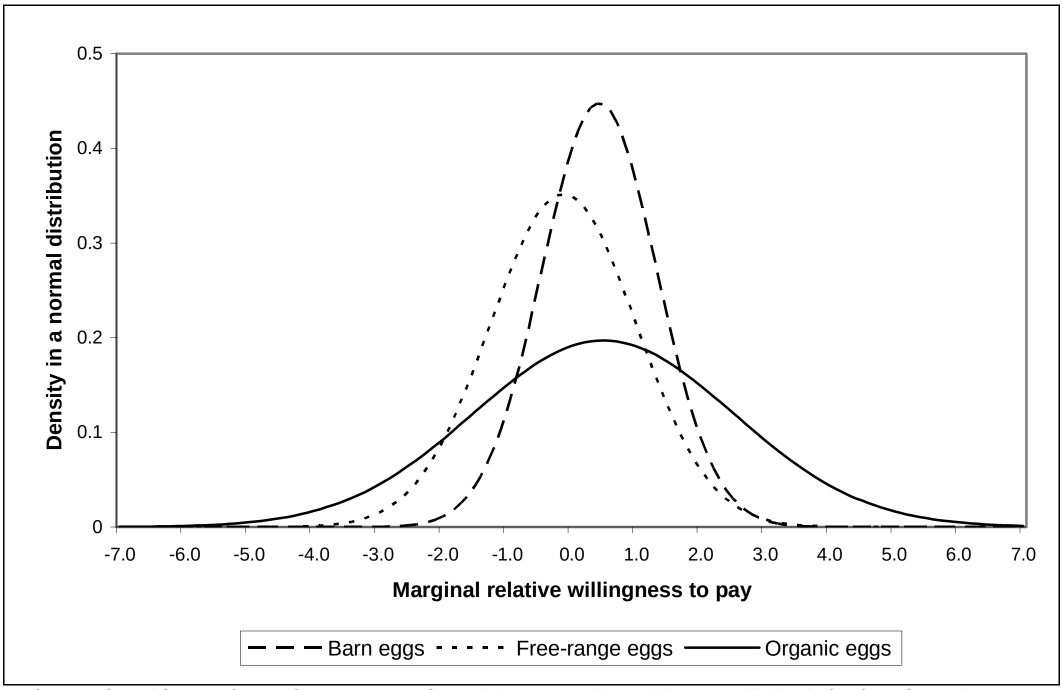
When estimating the conventional multinomial logit, it was important whether the estimated parameters were significantly different from zero, since this influenced the interpretation of the sign and the magnitude of marginal willingness to pay. In the mixed multinomial logit, whether the estimated mean of the distribution is significantly different from zero is no longer as important. Instead it is very important whether the estimated *standard deviations* might as well be zero, since this would mean that the distribution might as well be a one point distribution, indicating that mixing is unnecessary.

Figure 6.3 illustrates that the mean of the marginal willingness to pay might not be as important as the standard deviation. The huge difference in standard deviations means that if the reaction to each egg type follows a normal distribution, 69 percent of the families have positive marginal willingness to pay for barn eggs compared to battery eggs, 46 percent for free-range eggs and 60 percent for organic eggs; even though the mean of the marginal willingness to pay for organic eggs is higher than the mean of the marginal willingness to pay for barn eggs.[14]

[13] According to Revelt and Train (1998), the parameters of the mixed logit is generally expected to be numerically larger than the ones in the conventional logit. This difference reflects the fact that the mixed multinomial logit decomposes the unobserved portion of utility and normalises parameters on the basic part of the unobserved portion.
In the conventional multinomial logit, utility is given by $U_{ijt} = \beta' x_{ijt} + \xi_{ijt}$ where the $\beta$'s are normalised to allow $\xi$ to have the variance of an extreme value error. In the mixed multinomial logit some of $\xi$ is captured by the estimated variance of $\beta$. The variance in the error terms is therefore smaller in the mixed multinomial logit than in the conventional multinomial logit, and the normalisation makes the parameters of the mixed multinomial logit bigger (in absolute value) than in the conventional multinomial logit.
However, the marginal willingness to pay is calculated as the ratio of two parameters, which means that this effect cancels out. If both models are well specified the mean in the mixed logit should therefore equal the parameter in the conventional logit. If the mixed logit is assumed to be the 'true' model the conventional logit will to some extend be miss-specified and this may lead to differences.

[14] If the estimated parameters are taken as given and the standard errors are ignored, the normal distribution makes it easy to compute the percentage of all individuals (households) that have a positive marginal willingness to pay. As discussed in chapter 5, this is not the optimal way of using the parameters, but applying the information from the standard errors is beyond the scope of this study, and as in Revelt and Train (1998), Train (1998), Train (1999b) and Rouwendal and Meijer (2001) the standard errors are ignored in this study.

**Figure 6.3 Marginal relative willingness to pay for different types of eggs compared to battery eggs**




Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Only data from SuperBrugsen. Prices are relative which means that the relative marginal willingness to pay is measured as a percentage of the battery egg price divided by one hundred. Rationing is allowed.
Number of repetitions: 500. Starting values for the price and the means of the mixed parameters are taken from the conventional logit estimation in section 6.1, and 0.1 is used as starting value for the standard deviation of the reaction to egg types.

As mentioned in section 6.1, the labels 'barn eggs' and 'free-range eggs' mainly indicates increased animal welfare, whereas the 'organic' label indicates a more environmentally friendly production as well as a higher level of animal welfare. Some households may also perceive the organic eggs as being healthier than other egg types because the hens are fed with organic feed. The heterogeneity of marginal willingness to pay for organic eggs can therefore be induced by differences in the perception and evaluation of at least three different attributes, whereas the heterogeneity of marginal willingness to pay for barn and free-range eggs is expected to arise only from differences in perception and evaluation of animal welfare. This hypothesis is confirmed by the fact that the standard deviation of marginal relative willingness to pay (reported in Table 6.4) is much higher for organic eggs than for the two other egg types.

The likelihood ratio test established that the simple version of the mixed multinomial logit model is better than the simple version of the conventional multinomial logit model. It is therefore relevant to proceed with estimations under the mixed multinomial logit and expand the model by using some of the background information presented in chapter 4. This is done in the following section.

### *6.2.3 Expanding the model by using background information about the households*

As shown in chapter 4 the background data on households includes 800 variables not only about standard socio-demographic attributes of the households, but also about habits and attitudes. In the following models including geographical location of the residence, age of the main buyer or attitude to branded goods are estimated as examples of use of the background information. Geography and age are conventional socio-demographics whereas attitude to branded goods (‘mærkevarer’) is an example of one of the unique features of the GfK data.

As explained in chapter 3, variables that are choice invariant can only enter the model if they are crossed with variables that depend on choice. In this study it is assumed that the background variables only influence the reaction to *egg types*, but it could also be assumed that they influence the reaction to prices alone or to both prices and egg types. All three models have their virtues, and testing versions of all of them could be interesting in another study. In this study all customers in SuperBrugsen are assumed to have the same reaction to price, but customers e.g. from the capital area are allowed to have different preferences for the egg types than customers in rural municipalities in Jutland.

#### 6.2.3.1 Geography

The geographical location of the household residence is divided into five groups: Capital, island city-municipality, other city-municipality, Jutland city-municipality and Jutland other municipality. These geographical categories were also used in chapter 4 and are defined in appendix F.

The simple mixed multinomial logit model estimated in section 6.2.2 can be seen as a restricted version of the model allowing the parameter for each egg type to vary between geographical regions. It is therefore possible to conduct a likelihood ratio test to see if the use of geographical variation improves the model significantly. The test is presented in Table 6.7 and show that it can be rejected that the model without geography is just as good as the one using geography. Including geographical variation therefore has a significant effect on the preferences for barn eggs, free-range eggs and organic eggs in SuperBrugsen.

**Table 6.7 Testing the influence of geography on reaction to egg type**

| Log-likelihood value | | $-2*(\ln L_0 - \ln L_1)$ | Test-probability | Conclusion |
|---|---|---|---|---|
| $H_1$: With geography ($\ln L_1$) | $H_0$: Without geography ($\ln L_0$) | | $(\chi^2_{24})$ | |
| -1,931 | -1,972 | 82 | 0.000 | $H_0$ is rejected |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. The number of parameters in the model using no background variables is 1 (price) + 2(mean and standard deviation) * 3(egg types) = 7. Using geography leads to 1+2 * (3(egg types) * 5(geographical categories)) = 31 parameters so the degrees of freedom is 31 - 7 = 24.

Note that the log-likelihood values are lower than minus one *thousand*, not just one.

The most easily interpreted result of the estimations is the share of households with positive marginal willingness to pay for animal welfare related to eggs. If this share is less than 50 percent it means that the *mean* marginal willingness to pay is negative. However, these negative values of marginal willingness to pay will rarely be observed in the market unless battery eggs are offered at a higher unit price than non-battery eggs. As long as the price of battery eggs is lower than the price of other eggs the market will only observe that marginal willingness to pay is lower than the price difference. The estimated negative marginal willingness to pay is therefore in most cases an artificial problem, created by the functional form of the distribution of the parameters for egg types, and can be perceived as being zero.

The share of households with positive marginal willingness to pay for different types of eggs compared to battery eggs are presented in Table 6.8.[15] Customers in city municipalities (Island city municipality and to some extent Jutland city municipality) and especially in the capital more often have a positive marginal willingness to pay for non-battery eggs, than the customers in the rest of the country.

**Table 6.8 Share of households with positive marginal willingness to pay for different types of eggs in SuperBrugsen, by geographical location of household residence**

| Geography | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|
| Capital | 92 | 88 | 83 |
| Island city-municipality | 92 | 66 | 63 |
| Other Island municipality | 62 | 31 | 48 |
| Jutland city municipality | 62 | 44 | 62 |
| Jutland other municipality | 61 | 30 | 48 |
| *If geography is ignored:* | *69* | *46* | *60* |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated distribution of marginal willingness to pay therefore has no standard error.

The conclusion is therefore that marginal willingness to pay for organic eggs is generally higher in city municipalities than in the rest of the country, whereas the picture is less clear for barn eggs and free-range eggs.

#### *6.2.3.2 Age and attitude to branded goods*

Age of the person mainly responsible for the shopping is divided into the four groups: ‘Less than 30 years’ (approximately corresponding to ’21-29 years’), 30-44 years, 45-59 years and ‘60 years or more’. The result of testing the two models (with and without age) against each other is presented in Table 6.9.

**Table 6.9 Testing the influence of age of main buyer on reaction to egg type**

| Log-likelihood value | | $-2*(\ln L_0 - \ln L_1)$ | Test-probability | Conclusion |
|---|---|---|---|---|
| $H_1$: With age ($\ln L_1$) | $H_0$: Without age ($\ln L_0$) | | $\left(\chi^2_{18}\right)$ | |
| -1,963 | -1,972 | 17 | 0.549 | $H_0$ is accepted |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
The number of parameters in the model using no background variables is:
1(price) + 2(mean and standard deviation) * 3(egg types) = 7. Using age leads to 1+2*(3(egg types)*4(age categories)) =25 which means that the degrees of freedom is 18.

Note again, that the log-likelihood values are below minus one *thousand*.

The probability that the restricted model (ignoring the effect of age on preferences) is just as good as the unrestricted model (including the effect of age) is 54.9 percent. This means that the restricted model cannot be rejected. Age therefore has no significant effect on preferences for different egg types in SuperBrugsen.

[15] The estimation results are presented in Table G.5 in appendix G.

The effect of attitude to branded goods[16] is not significant either, but the probability that the restricted model (ignoring the effect of attitude to branded goods) is just as good as the unrestricted model (including the effect of attitude to branded goods) is only 8.1 percent, so it is close to being significant. Including the attitude to branded goods is significantly better than ignoring it in some of the other stores. This will be described in chapter 7.

When estimating mixed multinomial logit models using the program in Train et al. (1999), it is recommended by Train (1999c) that results of the conventional multinomial logit are used as starting values for the mean of the mixed parameters. The models including background data have therefore also been estimated with all parameters fixed.

As expected the log of the likelihood values for the mixed models are much higher than the ones using the conventional multinomial logit. As in section 6.2.2, likelihood ratio tests can be used to test whether the standard deviations differ significantly from zero. The tests show that mixed multinomial logit is a better model than the conventional multinomial logit for all three models including background data.

### *6.2.4 Conclusion on mixed multinomial logit modelling*

Using the program in Train et al. (1999) for estimation of the mixed multinomial logit requires that the conventional multinomial logit is estimated first, so that the parameter values from this estimation can be used as starting values for the simulated maximisation of the mixed multinomial logit.

When estimating mixed multinomial logit one must decide which parameters to mix, which distribution(s) to use and how many repetitions should be used in the simulation. The most important thing to remember when choosing which parameters to mix is that the derived marginal willingness to pay includes the inverted parameter of the price. The price must therefore not be defined on zero, since this would lead to undefined marginal willingness to pay. In the program in Train et al. (1999c) this can be avoided either by mixing with the lognormal distribution, or by keeping the price parameter fixed. In this study all price parameters are assumed to be fixed. This also solves the problem of aggregating marginal willingness to pay over consumers (mentioned when deriving mean marginal willingness to pay in chapter 2) since it means that all consumers have the same utility of money.

Likelihood ratio tests show that the mixed multinomial logit describes data significantly better than the conventional multinomial logit in all estimated models.

## 6.3 Conclusion on SuperBrugsen

The estimated distributions of marginal willingness to pay for the three non-battery egg types compared to battery eggs show that in SuperBrugsen 69 percent of the families

[16] 1,013 observations come from households that prefer branded goods to low prices, 1,354 observations come from households that prefer low prices to branded goods. No costumers in SuperBrugsen answered 'don't know' to this question.

have positive marginal willingness to pay for barn eggs compared to battery eggs, 46 percent of the families have for free-range eggs and 60 percent have for organic eggs.

The marginal willingness to pay for barn eggs exceeds the mean imputed relative price of these eggs for 60 percent of the population, and for organic eggs this is the case for almost 50 percent of the population. This corresponds with the result in Baltzer (2002). In Baltzer (2002), the marginal willingness to pay for free-range eggs lies below the observed price, which is also the case in the present study. The results obtained in the present study, using data from SuperBrugsen, therefore correspond in general terms with the results obtained in Baltzer (2002), using data from Kvickly.

Including background variables about customers in the modelling showed that marginal willingness to pay for organic eggs in SuperBrugsen is generally higher in city municipalities than in the non-city municipalities, whereas the picture is less clear for barn eggs and free-range eggs. Age of the main buyer has no significant effect on preferences for the different egg types, whereas attitude to branded goods is close to being significant (is significant at the ten percent level, but not on the 5 percent level).

In previous related studies the marginal willingness to pay has been assumed to be the same for the entire population. The present study uses mixed multinomial logit which also estimates a standard deviation of the willingness to pay, and this can be used as a measure of the degree of heterogeneity in the population. As expected, the heterogeneity of marginal willingness to pay for organic eggs was higher than for barn eggs and free-range eggs. This is explained by the fact that organic eggs encompass more attributes than the other egg types, and that variation may occur in the valuation of each of these attributes (animal welfare, environment, health and possibly more).

The fact that the estimated standard deviations can be seen as a measure of heterogeneity is one of the most attractive features of the mixed multinomial logit compared to the conventional multinomial logit, and it shows its strength in this case too.

# 7 Comparing results from different stores

In chapter 6 the results of estimations using data only from SuperBrugsen were presented. This chapter presents results from estimations using data from either (almost) the entire sample or from individual stores, and discusses the problems that occur.

Section 7.1 briefly presents the stores used in this chapter along with the results of conventional multinomial logit that are to be used as starting values in the mixed multinomial logit. In section 7.2 estimations are conducted using the mixed multinomial logit and only purchase data. The model is expanded by including background data on households in section 7.3. Possible improvements of the model are discussed in section 7.4, and section 7.5 concluded this chapter.

## 7.1 Brief description of the stores and the results of the conventional multinomial logit

The more than 100 different stores in the original data have been aggregated to the 20 different aggregated stores shown in Table 4.3, chapter 4.[1] Of these aggregated stores 15 are relevant in this study and have been used to further define the subsamples B and C, also in Table 4.3, chapter 4. Subsample A includes all 15 aggregated stores including sales directly from farms. Using these purchases is problematic, because the farms are expected to be very heterogeneous, and because the alternatives to the purchased egg type in many cases will be non-existent, and in most cases impossible to derive from data. Sales directly from farms are therefore excluded in Subsample B, and subsample A is disregarded in the estimations. In subsample C the aggregated stores 'Various grocers', 'Various discount stores', 'Greengrocers etc', and Irma have been excluded too. The first three aggregated stores are expected to be almost as heterogeneous as the sales directly from farms, although shortage of supply of individual egg types is expected to be less frequent than in sales directly from farms. Irma is excluded because the number of observations is very low. (Battery eggs are only observed purchased once in Irma, which means that all relative imputed prices are relative to this single observation).

Estimations using data from all stores (subsample B or C) lead to price parameters not significantly different from zero in both subsample B and C as long as absolute prices were used.[2] This meant that marginal willingness to pay could not be calculated since it includes the inverted value of the price parameter. Using relative prices solved the price parameter problem, but the resulting marginal relative willingness to pay compared to battery eggs was negative for all egg types. Estimations using data from one aggregated store at a time revealed huge differences in price parameters as well as parameters for

[1] The aggregation of the 15 aggregated stores used in this study are presented in detail in Table E.1.2 in appendix E.
[2] Estimation results are presented in Table G.6 in appendix G.

the three egg types.[3] This lead to considerable differences in marginal willingness to pay for the three egg types in different store aggregates. In some stores the marginal willingness to pay was negative for one or more of the egg types, and combined with high absolute values this posed a serious problem. Including rationing as discussed in chapter 5 improved the results in stores where rationing is frequent, but it did not solve the problem of negative willingness to pay. In some store aggregates the price parameter was not significantly different from zero and in theses stores marginal willingness to pay could not be calculated.[4]

As in chapter 6 the marginal willingness to pay for barn eggs, free-range eggs and organic eggs is compared to battery eggs and the estimation results from the conventional multinomial logit (see Table G.7 in appendix G) are used as starting values for the means of the mixed parameters. The three parameters for the egg types are mixed with the normal distribution, and the price parameter is not mixed. Separate estimations are conducted using data from different subsamples.

## 7.2 Estimations using only purchase data

The problem of insignificant price reaction is just as relevant when the mixed multinomial logit is used as when the conventional multinomial logit is used. If the price parameter is not significantly different from zero the marginal willingness to pay cannot be derived. The estimated price parameters are presented in Table G.9 in appendix G. The price parameter is not significantly different from zero in Irma, Fakta, Netto, Aldi, Favør, 'Various discount stores', 'Greengrocers etc' and 'Directly from farms'. Irma and 'Greengrocers etc' have very few observations, and very few purchases of battery eggs. Apart from 'Directly from farms' where the prices (as mentioned in chapter 4) collapse around one, the remaining stores are discount stores with very little price variation. The subsamples with price parameters significantly different from zero are the same as in the conventional multinomial logit (except Dagligbrugsen where the price parameter now becomes significant at the five percent level), and they all have negative signs as expected.

The estimated mean and standard deviation of the parameters for the three different egg types are presented in Table G.10 in appendix G. In subsamples where the price parameter is significantly different from zero all estimated standard deviations are also significantly different from zero, which indicates that the mixed parameter is better than the fixed one used in the conventional multinomial logit. As in chapter 6, it is possible to conduct a likelihood ratio test with three degrees of freedom, to see if the conventional multinomial logit is almost as good as the mixed multinomial logit. Applying this test to each of the stores with price parameter significantly different from

---

[3] Estimation results are presented in Table G.7 in appendix G.
[4] For store aggregates with price parameter significantly different from zero, the marginal relative willingness to pay with and without rationing is reported in Table G.8 in appendix G.

zero shows that the mixed multinomial logit is significantly better than the conventional multinomial logit in all cases.[5]

Some of the means of the parameters for different egg types are not significantly different from zero, but as mentioned in chapter 6, this is not as important in the mixed multinomial logit as in the conventional multinomial logit.

Figure 7.1 to Figure 7.3 graphically present the estimated distributions of willingness to pay for different types of eggs for subsample C, SuperBrugsen and 'Various grocers'. As in chapter 6, it is clear that the standard deviations may sometimes be more important than the means in a mixed multinomial logit.[6]

**Figure 7.1 Marginal willingness to pay for barn eggs**

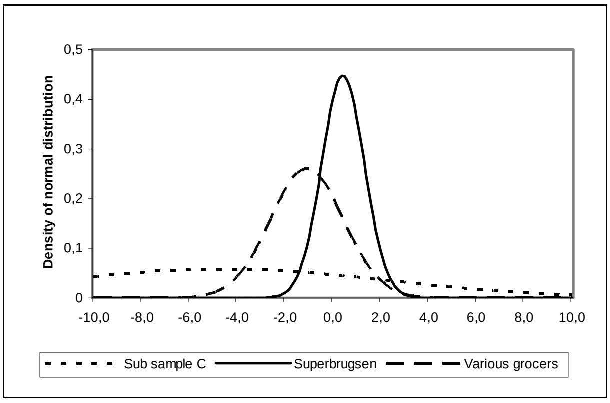



**Figure 7.2 Marginal willingness to pay for free-range eggs**

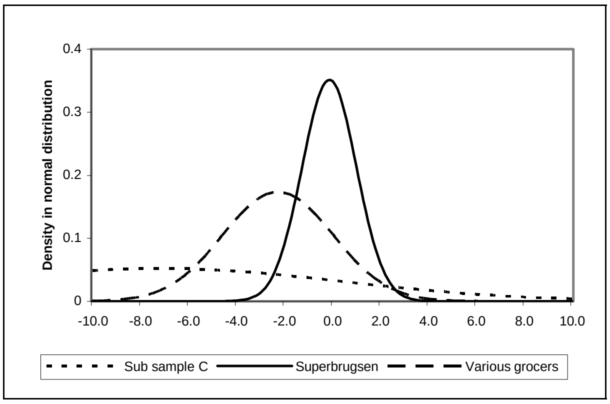



**Figure 7.3 Marginal willingness to pay for organic eggs**

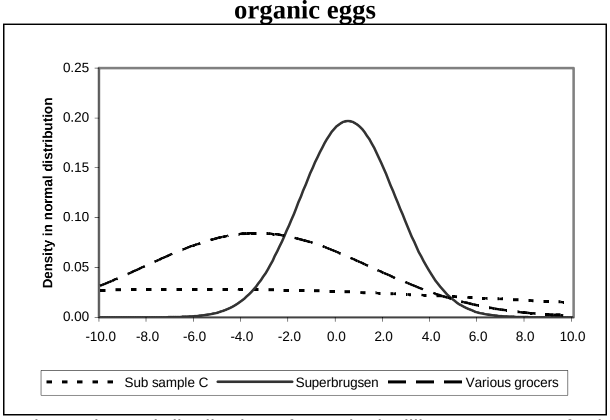



Figure 7.1 to Figure 7.3 show the estimated distribution of marginal willingness to pay for barn, free-range and organic eggs in different subsamples. (From Table G.11 in appendix G).

As expected the estimated standard deviation is much smaller in subsamples including only one chain of stores, than in subsamples that include more (and therefore more heterogeneous) chains of stores. The standard deviation is a measure of the degree of

[5] The tests are presented in Table G.12 in appendix G.

[6] Table G.11 in appendix G presents the distribution of marginal relative willingness to pay for different egg types in all store aggregates where the parameter for price is significantly different from zero.

heterogeneity among individuals, and using more homogeneous subsamples therefore decreases the standard deviation.

As also expected, the heterogeneity of marginal willingness to pay for organic eggs is higher than for barn eggs and free-range eggs (standard deviations for organic eggs are generally twice as high as for the two other egg types). As in chapter 6, this confirms the fact that organic eggs encompass more attributes than the other egg types, and that variation may occur in the valuation of each of these attributes (animal welfare, environment, health and possibly more).

The marginal willingness to pay for free-range eggs is generally lower than for barn eggs or organic eggs, just as it was found in chapter 6 and Baltzer (2002). This could be a small sample problem. As can be seen in table 4.14 in chapter 4, six of the 15 aggregated stores have less than ten percent sales of free-range eggs, whereas only two out of the 15 has less than ten percent sales of barn eggs. In many stores purchases of free-range eggs will therefore be very infrequent, and the method for detecting rationing in this study may no always perceive this rationing correctly.

It could, however, also be an indication that free-range eggs are redundant in the egg market. If people perceive the level of animal welfare in free-range eggs as close to the level of animal welfare in barn eggs (either because they are unable to distinguish the two types or because they actually perceive the extra animal welfare in the free-range eggs as marginal), and generally experience that free-range eggs are more expensive, barn eggs will be preferred to free-rang eggs. If people perceive the level of animal welfare in free-range eggs as being close to the level of animal welfare in organic eggs, and also perceive the organic eggs as encompassing additional attractive attributes aside from animal welfare, the fact that free-range eggs cost almost the same as organic eggs will lead to organic eggs being chosen over free-range eggs in many cases.

As mentioned in chapter 6 results in Statistics Denmark (2002) indicate that only 50 percent of the population have positive willingness to pay for organic goods in general. This means that the median willingness to pay is expected to be close to zero. As long as the mixing distribution of the evaluation of e.g. organic eggs does not allow half of the population to ascribe the value of zero to organic eggs, a positive median marginal willingness to pay in some store aggregates will lead to a negative median marginal willingness to pay in others.

In most of the stores the estimated mean marginal willingness to pay for the different non-battery egg types is negative. In general, the more heterogeneous the store aggregate is, the more negative the mean gets.

Negative marginal willingness to pay for different egg types does not necessarily mean that people have negative utility of e.g. the animal welfare related to the eggs. It may also be the case that people believe that there is no difference in the level of animal welfare related to different types of eggs, and that the different labels are simply a way of deceiving the consumers in order to gain a higher profit. This type of mistrust of the labels may lead to negative marginal willingness to pay. According to Økologisk Landsforening (2002), in 2000, 8.8 percent of the Danish population had no trust in

Danish organic goods, and 25.5 percent had little trust. This means that only two thirds of the population trusted that organic goods labelled with the 'Ø-label' were truly organic. However, the third of the population with little or no trust ought not to be enough to justify the mean negative marginal willingness to pay.

In practice, these negative values of marginal willingness to pay will rarely be observed in the market unless battery eggs are offered at a higher unit price than non-battery eggs. As long as the price of battery eggs is lower than the price of other eggs the market will only observe that marginal willingness to pay is lower than the price difference. The estimated negative marginal willingness to pay may therefore be seen as an artificial problem, created by the functional form of the distribution of the parameters for egg types, and perceived as being zero in stead of being negative.

As in chapter 6 it is possible to compute the percentage of all individuals (households) that have a positive marginal willingness to pay. If this share is less than 50 percent it means that the *mean* marginal willingness to pay is negative. The results are presented in Table 7.1.

Remember that these figures depend on the *chosen* distribution and that the standard errors of the estimated parameters are ignored. The most relevant thing to do with the numbers is therefore not to interpret them as the absolute truth, but rather to compare the share of customers with positive marginal willingness to pay in different subsamples. The subsamples B and C are much alike, but there are huge differences between the stores. Note that the more stores there are in the subsample, the lower is the percentage of customers with positive marginal willingness to pay in general. 'Dagligbrugsen' is a mixture of different independent stores, so they may also be quite heterogeneous.

**Table 7.1 Percentage of the households that have positive marginal willingness to pay for different egg types compared to battery eggs**

| | Percentage with positive marginal willingness to pay for | | |
|---|---|---|---|
| | **Barn eggs** | **Free-range eggs** | **Organic eggs** |
| Subsample B | 24 | 17 | 32 |
| Subsample C | 25 | 17 | 33 |
| SuperBrugsen | 69 | 46 | 60 |
| DagligBrugsen | 39 | 26 | 31 |
| Kvickly and OBS | 74 | 49 | 52 |
| Føtex | 50 | 45 | 51 |
| Prima | 40 | 26 | 44 |
| Various grocers | 24 | 16 | 24 |
| Bilka | 27 | 31 | 36 |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsamples B and C are defined in chapter 4, Table 4.3. Separate estimations on each subsample. Prices are relative which means that the relative marginal willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit. The starting values for the standard deviations are set to 0.1.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated distribution of marginal willingness to pay therefore has no standard error.

Likelihood ratio tests established that the simple version of the mixed multinomial logit model was better than the simple version of the conventional multinomial logit model in all cases. As in chapter 6 it is therefore relevant to proceed with estimations under the mixed multinomial logit and expand the model by using some of the background information presented in chapter 4. This is done in the following section.

## 7.3 Expanding the model by using background information about the households

As in chapter 6 models including geographical location of the residence, age of the main buyer or attitude to branded goods are estimated as examples of use of the background information. Geography and age are conventional socio-demographics and will only be discussed briefly. Attitude to branded goods ('mærkevarer') is an example of one of the unique features of the GfK data and will be presented in more detail.

Based on the estimated distributions of marginal willingness to pay, the estimations concentrate on

1. *Subsample C* as an example of a very heterogeneous subsample
2. *SuperBrugsen* as an example of a store with customers with positive mean marginal willingness to pay
3. *Føtex* as an example of a store with customers with mean marginal willingness to pay not significantly different from zero
4. *Bilka* as an example of a store with customers with negative mean marginal willingness to pay

As in chapter 6 it is assumed that the background variables only influence the reaction to *egg types*, not the reaction to price. The estimations are conducted separately on the different subsamples. This means that both the reaction to price and egg types differs from store aggregate to store aggregate. The reaction to price and egg types does therefore not have to be the same in Bilka as in SuperBrugsen. However, within SuperBrugsen all customers are assumed to have the same reaction to price, but customers e.g. from the capital area are allowed to have different preferences for the egg types than customers in rural municipalities in Jutland.

### *7.3.1 Geography*

Estimating models where reaction to egg type is allowed to depend on geographical location of household residence[7] means that the number of observations used to estimate each parameter decreases, as can be seen in Table 7.2 below.

[7] These geographical categories were also used in chapter 4 and are defined in appendix F.

**Table 7.2 Number of observations in four subsamples by geographical location of household residence**

| | Subsample C | SuperBrugsen | Føtex | Bilka |
|---|---|---|---|---|
| Capital | 5,104 | 507 | 434 | 103 |
| Island city-municipality | 3,264 | 355 | 175 | 347 |
| Other island municipality | 1,784 | 532 | 44 | 39 |
| Jutland city municipality | 3,636 | 402 | 880 | 261 |
| Jutland other municipality | 2,016 | 571 | 50 | 94 |
| *Total* | *15,804* | *2,367* | *1,583* | *844* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000.

The number of observations from 'other city-municipality' is very low in Føtex and in Bilka and this may be the reason why estimations using these two subsamples failed to converge.[8] Estimations using subsample C and SuperBrugsen converged and it is therefore possible to conduct a likelihood ratio test to see if the use of geographical variation improves the model significantly. The tests are presented in Table 7.3 and show that it can be rejected that the model without geographical variation is just as good as the one using geography. Geographical variation therefore has a significant effect on the preferences for different egg types in both subsample C and SuperBrugsen.

**Table 7.3 Testing the influence of geography on reaction to egg type**

| | Log-likelihood value | | -2*($\ln L_0 - \ln L_1$) | Test-probability | Conclusion |
|---|---|---|---|---|---|
| | $H_1$: With geog. ($\ln L_1$) | $H_0$: without geogr. ($\ln L_0$) | | $\left(\chi^2_{18}\right)$ | |
| Subsample C | -13,114 | -13,164 | 101 | 0.000 | $H_0$ is rejected |
| SuperBrugsen | -1,931 | -1,972 | 82 | 0.000 | $H_0$ is rejected |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsample C is defined in chapter 4, Table 4.3. The number of parameters in the model using no background variables is 1 (price) + 2(mean and standard deviation) * 3(egg types) = 7. Using geography leads to 1+2 * (3(egg types) * 5(geographical categories)) = 31parameters so the degrees of freedom is 31 - 7 = 24.

Note that the log-likelihood value in subsample C is minus thirteen-*thousand*, not thirteen.

The share of households with positive marginal willingness to pay for different types of eggs compared to battery eggs are presented in Table 7.4.[9] Customers in city municipalities (island city municipality and Jutland city municipality) and especially in the capital more often have a positive marginal willingness to pay for organic eggs, than the rest of the country. The pattern is, however, not completely clear since 'Jutland other municipality' have the highest fraction of households with positive marginal willingness to pay for barn eggs and the lowest for free-range eggs.

[8] When estimating mixed multinomial logit models using the program in Train et al. (1999), it is recommended by Train (1999c) that results of the conventional multinomial logit are used as starting values for the mean of the mixed parameters. The models including background data have therefore also been estimated with all parameters fixed. These estimations, using the conventional multinomial logit, succeeded for all four subsamples in all three cases, indicating that the small sample problems encountered above are worsened by the mixed multinomial logit. The mixed multinomial logit provides more information than the conventional multinomial logit, and it is therefore reasonable to expect the mixed multinomial logit model to be more vulnerable to small sample problems. The problems occurred when the number of observations within a group was lower than 50.

[9] The estimation results are presented in Table G.5 and G.13 in appendix G.

**Table 7.4 Share of households with positive marginal willingness to pay for different types of eggs in subsample C and SuperBrugsen, by geographical location of household residence**

| Geography | Subsample C | | | SuperBrugsen | | |
|---|---|---|---|---|---|---|
| | Barn eggs | Free-range eggs | Organic eggs | Barn eggs | Free-range eggs | Organic eggs |
| Capital | 20 | 18 | 44 | 92 | 88 | 83 |
| Island city-municipality | 20 | 17 | 35 | 92 | 66 | 63 |
| Other island municipality | 27 | 14 | 26 | 62 | 31 | 48 |
| Jutland city municip. | 30 | 19 | 37 | 62 | 44 | 62 |
| Jutland other municip. | 32 | 9 | 18 | 61 | 30 | 48 |
| *If geography is ignored:* | *25* | *17* | *33* | *69* | *46* | *60* |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsample C is defined in chapter 4, Table 4.3.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated distribution of marginal willingness to pay therefore has no standard error.

The conclusion must therefore be that marginal willingness to pay for organic eggs is generally higher in city municipalities than in the rest of the country, whereas the picture is less clear for barn eggs and free-range eggs.

### *7.3.2 Age*

Dividing the age of the person responsible for most of the shopping into four categories gives the following number of observations in each group:

**Table 7.5 Number of observations in four subsamples by age of main buyer**

| Age | Subsample C | SuperBrugsen | Føtex | Bilka |
|---|---|---|---|---|
| < 30 years | 1,095 | 119 | 210 | 43 |
| 30 - 44 years | 4,054 | 585 | 366 | 271 |
| 45 - 59 years | 5,387 | 785 | 534 | 332 |
| > 60 years | 5,268 | 878 | 473 | 198 |
| *Total* | *15,804* | *2,367* | *1,583* | *844* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000.

Here only the low number of observations from young people in Bilka cause problems and as expected the estimations using Bilka did not succeed. The results of testing the two models against each other in subsample C, SuperBrugsen and Føtex are presented in Table 7.6.

**Table 7.6 Testing the influence of age of main buyer on reaction to egg type**

| | Log-likelihood value | | -2*(lnL$_0$ –lnL$_1$) | Test-probability | Conclusion |
|---|---|---|---|---|---|
| | $H_1$: With age (lnL$_1$) | $H_0$: without age (lnL$_0$) | | $\left(\chi^2_{18}\right)$ | |
| Subsample C | -13,146 | -13,164 | 36 | 0.006 | $H_0$ is rejected |
| SuperBrugsen | -1,963 | -1,972 | 17 | 0.549 | $H_0$ is accepted |
| Føtex | -1,705 | -1,718 | 25 | 0.123 | $H_0$ is rejected |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsample C is defined in chapter 4, Table 4.3. The number of parameters in the model using no background variables is 1 (price) + 2 (mean and standard deviation) * 3(egg types) = 7. Using age leads to 1+2*(3(egg types)*4(age categories)) =25 which means that the degrees of freedom is 18.

Note again, that the log-likelihood values are below minus one *thousand*.

In SuperBrugsen and Føtex the probability that the restricted model (ignoring the effect of age on preferences) is just as good as the unrestricted model (including the effect of age) is 54.9 and 12.3 percent. This means that the restricted model cannot be rejected. Age therefore has no significant effect on preferences for the egg types in SuperBrugsen and Føtex. However, the restricted model is rejected in subsample C, which means that age influences the preferences for egg types in subsample C.

The share of households with positive marginal willingness to pay for the each egg type[10] in subsample C is presented in Table 7.4 and shows that in subsample C young people and old people have a slight tendency to have a higher marginal willingness to pay than persons between 30 and 60, but the pattern is not very clear.

**Table 7.7 Share of households with positive marginal willingness to pay for different types of eggs in subsample C, by age of main buyer**

| Age of main buyer | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|
| < 30 years | 21 | 15 | 37 |
| 30 - 44 years | 22 | 15 | 32 |
| 45 - 59 years | 26 | 14 | 29 |
| > 60 years | 32 | 19 | 38 |
| *Ignoring age:* | *25* | *17* | *33* |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsample C is defined in chapter 4, Table 4.3.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated distribution of marginal willingness to pay therefore has no standard error.

### *7.3.3 Attitude to branded goods*

One of the unique features of the GfK data is the background information about attitudes. This study presents only one result based on attitudes but there are numerous possibilities.

In one question the person answering the background questionnaire (usually the person responsible for most of the shopping) is asked to choose between two statements about brand labels ('mærkevarer'):

- 'I prefer brand labels to cheaper products, to be sure to get good quality' (In the following tables: Brand > low price)
- 'No-name products are often just as good as brand labels. I buy cheaper no-names just as often as brand labels' (In the following tables: Low price > brand)

Persons not answering or answering 'don't know' are not included in the estimations.

Preferences for branded goods indicate that other attributes than price are important when deciding which good to purchase. In this study it is therefore expected that people who prefer branded goods are willing to pay more for the non-battery egg types. In Table 7.8 the observations in the different store aggregates are distributed by attitudes to branded goods.

[10] The estimation results are presented in Table G.14 in appendix G.

**Table 7.8 Number of *observations* (No. obs.) and distribution of observations (%) in four different subsamples by attitude to branded goods**

| | Subsample C | | SuperBrugsen | | Føtex | | Bilka | |
|---|---|---|---|---|---|---|---|---|
| | No. obs. | % | No. obs. | % | No. obs. | % | No. obs. | % |
| Brand > low price | 4,535 | 28.7 | 1,013 | 42.8 | 625 | 39.5 | 214 | 25.4 |
| Low price > brand | 11,189 | 70.8 | 1,354 | 57.2 | 953 | 60.2 | 627 | 74.3 |
| Don't know/not answered | 80 | 0.5 | 0 | 0.0 | 5 | 0.3 | 3 | 0.4 |
| *Total* | *15,804* | *100.0* | *2,367* | *100.0* | *1,583* | *100.0* | *844* | *100.0* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000.

Small sample problems only occur in the 'don't know' category and since households with these answers are excluded in the estimation there should be no small sample problems.

The number of purchases in a given store may differ from household to household, and the households may appear as customers in more than one store. The distribution of *households* with different attitudes may therefore differ from the distribution of *purchases* made by these households. The results in Table 7.9 are the same as in chapter 4, and differ from the results in Table 7.8.

**Table 7.9 Distribution of attitudes to branded goods by *households* in four different subsamples**

| | Subsample C | SuperBrugsen | Føtex | Bilka |
|---|---|---|---|---|
| Brand > low price | 26.9 | 33.7 | 33.0 | 27.0 |
| Low price > brand | 72.5 | 66.3 | 66.6 | 72.3 |
| Don't know/ not answered | 0.6 | 0.0 | 0.5 | 0.7 |
| *Total* | *100* | *100* | *100* | *100* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000.

SuperBrugsen and Føtex have very similar customers (based on this attitude question) and Bilka is close to the average achieved in subsample C. The log-likelihood functions of the two models with and without attitude to branded good are presented in Table 7.10, along with the results of the likelihood ratio test for significance of attitude to branded goods.

**Table 7.10 Testing the influence of attitude to branded goods on reaction to egg type**

| | Log-likelihood value | | -2*(lnL$_0$ –lnL$_1$) | Test-probability $\left(\chi^2_{18}\right)$ | Conclusion |
|---|---|---|---|---|---|
| | $H_1$: With geog. (lnL$_1$) | $H_0$: without geogr. (lnL$_0$) | | | |
| Subsample C | -13,151 | -13,164 | 26 | 0.000 | $H_0$ is rejected |
| SuperBrugsen | -1,966 | -1,972 | 11 | 0.081 | $H_0$ is accepted |
| Føtex | -1,705 | -1,718 | 26 | 0.000 | $H_0$ is rejected |
| Bilka | -843 | -852 | 18 | 0.007 | $H_0$ is rejected |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsample C is defined in chapter 4, Table 4.3.
The number of parameters in the model using no background variables is 1 (price) + 2 (mean and standard deviation) * 3(egg types) = 7. Using attitude to branded goods leads to 1+2 * (3(egg types) * 2(attitude categories)) = 13 so the degrees of freedom is 6

In subsample C, Føtex and Bilka the probability that the model ignoring the effect of attitude to branded goods on preferences for eggs, is as good as the model including this attitude, is lower than one percent. This means that the model ignoring the attitude to branded goods can be rejected. Attitude to branded goods therefore influence preferences for the egg types in subsample C, Føtex and Bilka.

As it was seen in chapter 6, the probability that the two models are equally good in SuperBrugsen is 8.1 percent, which means that the model ignoring the attitude to branded goods *cannot* be rejected. Attitude to branded goods does therefore not influence the preferences for egg types in SuperBrugsen. Customers in SuperBrugsen and Føtex had the same distribution of answers to the attitude question, yet the influence of attitudes is significant in Føtex but not in SuperBrugsen. However the marginal willingness to pay was generally positive in SuperBrugsen even without using the attitude question and that may mean that the extra effect of the attitude is very minor, and therefore not significant. The share of households with positive marginal willingness to pay using the attitude to branded goods is shown for subsample C, Føtex and Bilka in Table 7.11.[11]

**Table 7.11 Share of households with positive marginal willingness to pay for different types of eggs in subsample C, Føtex and Bilka, by attitude to branded goods**

| Attitude to branded goods | Subsample C | | | Føtex | | | Bilka | | |
|---|---|---|---|---|---|---|---|---|---|
| | Barn eggs | Free-range eggs | Org. eggs | Barn eggs | Free-range eggs | Org. eggs | Barn eggs | Free-range eggs | Org. eggs |
| Brand > low price | 36 | 27 | 47 | 49 | 50 | 64 | 40 | 57 | 51 |
| No-name > brand | 22 | 14 | 29 | 50 | 43 | 46 | 24 | 26 | 29 |
| *Ignoring the attitude:* † | *25* | *27* | *33* | *50* | *45* | *51* | *27* | *31* | *36* |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsample C is defined in chapter 4, Table 4.3. †: Taken from Table 7.1.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated distribution of marginal willingness to pay therefore has no standard error.

This question apparently hits the heart of the matter, particularly in subsample C and Bilka. People who prefer branded goods are more likely to have positive marginal willingness to pay for all types of eggs than people who would be just as happy to buy no-name products to save money. In Føtex the effect is smaller than in subsample C but it still means that the mean marginal willingness to pay for organic eggs becomes positive and significantly different from zero.

The effect in Bilka is very strong and the fraction of households with positive marginal willingness to pay is close to twice as high for households who prefer branded goods compared to households who do not.

The results from Bilka are presented graphically in Figure 7.4 to Figure 7.6.

Two thirds of the families in Bilka prefer no-name products to branded goods. Therefore only 214 observations are used to estimate the reaction to egg types for the 72 families that prefer branded goods, while there are 627 observations used to estimate the reaction for the 193 families that prefer no-name products. This might explain why the standard deviation is much bigger for the families who prefer no-name products. On the

[11] The estimation results are presented in Table G.15 to Table G.18 in appendix G.

other hand, it could also actually be the case that people who prefer branded goods are more homogeneous than those who do not.

**Figure 7.4 Marginal willingness to pay for barn eggs in Bilka**


**Figure 7.5 Marginal willingness to pay for free-range eggs in Bilka**

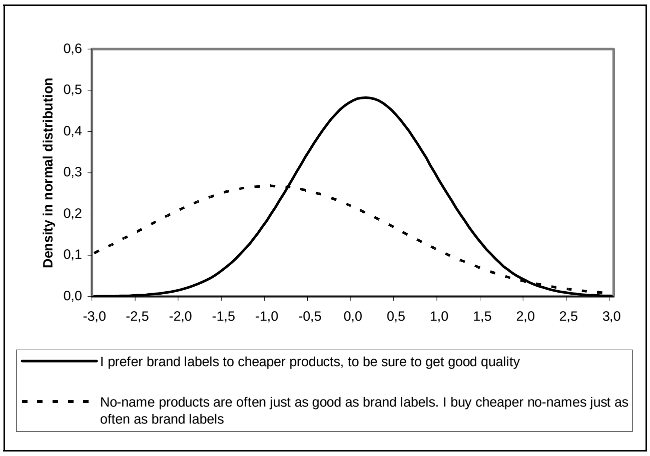
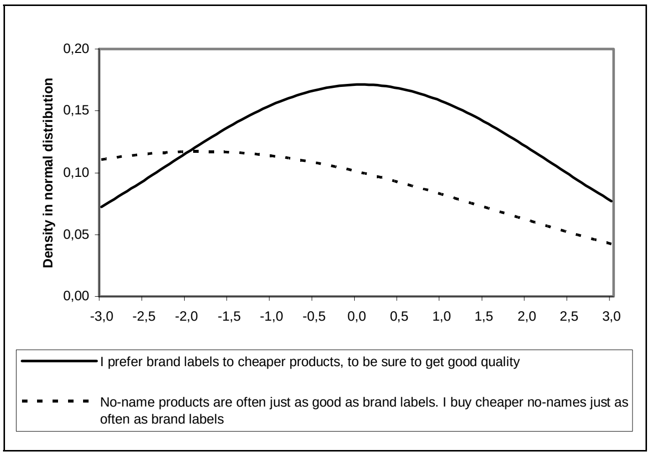



**Figure 7.6 Marginal willingness to pay for organic eggs in Bilka**




Estimations using GfK purchase data on eggs in Bilka from 26 June 1999 to 30 June 2000.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated distribution of marginal willingness to pay therefore has no standard error.

To conclude, it is clear that people who generally prefer branded goods to no-name products are willing to pay more for all types of eggs compared to battery eggs than people who are just as happy to buy no-name products to save money. The fact that the differences between the two groups are so pronounced indicates that utility of e.g. animal welfare is closely correlated with utility of quality of other goods. This is very interesting, but not unexpected.

### *7.3.4 Summing up the results of using background variables*

As expected the log of the likelihood values for the mixed models are much higher than the ones using the conventional multinomial logit. As in section 7.2, likelihood ratio tests can be used to test whether all standard deviations differ significantly from zero. The tests show that mixed multinomial logit is a better model than the conventional multinomial logit for all subsamples for which the mixed multinomial logit could be estimated, in all three models including background data.

Using geography leads to a better model for both subsample C and SuperBrugsen. The model could not be estimated for Føtex and Bilka. Using the age of the main buyer only improves the model significantly when estimating using data from subsample C. The model could not be estimated for Bilka. Surprisingly enough the attitude to branded goods is not a significant improvement in SuperBrugsen, even though the parameters varied systematically and visibly. (These results were not presented). However, SuperBrugsen was the store aggregate where the influence was weakest, and for the three other subsamples using this attitude is a significant improvement.

Customers in SuperBrugsen and Føtex had the same distribution of answers to the attitude question, yet the influence of attitudes is significant in Føtex but not in SuperBrugsen. However, the marginal willingness to pay was generally positive in SuperBrugsen even without using the attitude question and that might mean that the extra effect of the attitude is very minor, and therefore not significant.

### 7.4 Possible improvements of the results

The results of the estimations supports the hypothesis that marginal willingness to pay for barn eggs, free-range eggs and organic eggs compared to battery eggs varies between stores and that the heterogeneity of marginal willingness to pay for organic eggs is higher than for barn eggs and free-range eggs. In the conventional multinomial logit positive results were only achieved in SuperBrugsen, 'Kvickly and OBS' and, for organic eggs, in Føtex. In some cases the mean marginal wiliness to pay was *lower* than the price of a battery egg, meaning that more than 50 percent of the costumers in this particular store aggregate actually ought to demand money in order to accept e.g. a free-range egg. Introducing the mixed multinomial logit did not improve the results as far as the mean marginal willingness to pay was concerned. In some cases the negative mean of a parameter was several times higher (numerically) than the parameter in the conventional multinomial logit.

However, the standard *deviations* (not the standard errors) of the parameters in the mixed multinomial logit yielded new and interesting information compared to the conventional multinomial logit. First of all the size of the standard deviations indicated that aggregates consisting of many different chains of stores lead to a very high degree of heterogeneity. This unobserved heterogeneity might be the reason for the counter-intuitive results.

As discussed in chapter 5 many different decisions had to be made before proceeding with the actual estimations. It is reasonable to expect both the size and the freshness of the eggs offered to influence the choice between different types of eggs. This information is not available, just as the information about the actual size of the egg.

Information about freshness cannot be achieved in any way, but perhaps information about the general distribution over egg sizes for each type of egg could be gained for each of the egg producers recorded in the data. This might be used to improve the imputed prices. As mentioned in chapter 5, information about geography might also be used when imputing prices, in a more refined way than the simple means used in this study. Using information about tray size might also improve the imputed prices.

There is no doubt that the method for detecting rationing used in this study is very crude. One way of improving the method could be to use the information about weekday and hour of day recorded for each purchase. Rationing might be more frequent on a Saturday afternoon than a Monday morning and this could be used to define a *probability* of rationing, and then use this probability to detect rationing more precisely.

A recent study by Bjørner et al. (2002) investigates the effect of the Nordic Swan label on consumers' choice. The Nordic Swan is an eco-label, indicating that the product is less harmful to environment. The data used in Bjørner et al. was purchases of toilet paper, paper towels and detergents in the exact same GfK data set as used in the present study. Estimation results in Bjørner et al. persistently yielded positive marginal willingness to pay for the Swan label, especially for detergents. The main difference between the study by Bjørner et al. and the present study is the good used to reveal altruistic preferences. Detergents are far easier to describe than eggs. There are no differences in freshness, the unit is defined by 'one standard wash' and the risk of experiencing rationing is far lower than for eggs. Comparing the results of the estimations using detergents in Bjørner et al. (2002) with the results obtained in this study indicates that unobserved heterogeneity contributes to the negative marginal willingness to pay for the three egg types. Therefore, improved detection of rationing and better imputed prices might reduce the unobserved heterogeneity and lead to better results than the ones obtained in this study.

Accepting the results as they are, the standard deviations of parameters in the mixed multinomial logit lead to result that even if the mean marginal willingness to pay is negative there will always be a share of the population that have positive marginal willingness to pay for each egg type. The shares vary between stores and between households with different background attributes, but in most stores at least 25 percent have positive marginal willingness to pay. This leaves 75 percent with a negative marginal willingness to pay since the normal distribution assigns the probability zero to any point, including zero. However, a negative marginal willingness to pay will hardly ever be observable on a market since it requires that a person chooses a battery egg even though it is more expensive than a non-battery egg. This situation does occur in data, but not very frequently and probably as a result of errors in the imputed prices. The normal distribution will always assume that some share of the population has negative marginal willingness to pay for each egg type. However, in the real world this share will usually be perceived as having a marginal willingness to pay of zero. As mentioned before the problem of negative marginal willingness to pay can therefore be seen as induced by the normal distribution rather than the observed data. Finding a better functional form than the ones available in Train et al. (1999) might therefore also improve the results a great deal.

## 7.5 Conclusion on comparing results from different stores

The mixed multinomial logit describes data significantly better than the conventional multinomial logit in all estimated models. The estimated distribution of marginal willingness to pay for egg types varies from store to store and the standard deviations of the distributions decrease when the heterogeneity of the stores in the subsample decreases. The differences are not surprising since simple examinations of the data (in chapter 4) revealed that the customers in different stores vary not only on conventional socio-demographics, but also on habits and attitudes. Furthermore, the variety and price structure vary from store to store. In stores that focus on high quality rather than price (e.g. SuperBrugsen), marginal willingness to pay for non-battery eggs is generally positive. In stores that focus on price rather than quality (e.g. Bilka), marginal willingness to pay for non-battery eggs is generally negative.

As expected, the heterogeneity of marginal willingness to pay for organic eggs was generally higher than for barn eggs and free-range eggs. This is explained by the fact that organic eggs encompass more attributes than the other egg types, and that variation may occur in the valuation of each of these attributes (animal welfare, environment, health and possibly more).

It is also found that allowing the preferences for different egg types to depend on background variables such as geographical location of household residence improves the model significantly, and that allowing the preferences for different egg types to depend on attitudes, such as attitude to branded goods, also improves the model. The effect of the age of the main buyer was, in most cases, not significant.

Marginal willingness to pay for non-battery eggs compared to battery eggs is highest in SuperBrugsen and lowest in Bilka (when aggregates including very heterogeneous stores are excluded). Marginal willingness to pay is generally higher in city municipalities and among households that prefer branded goods to be sure to get good quality.

Not only does the mixed multinomial logit model describe data significantly better than the conventional multinomial logit model, it also provides new and interesting results about the distribution of marginal willingness to pay in the population. Looking at the percentage of households that are expected to have positive marginal willingness to pay sometimes reveals more (or at least different) information than just looking at the mean.

# 8 Conclusion

In SuperBrugsen, the estimated distributions of marginal willingness to pay for the three non-battery egg types compared to battery eggs show that 69 percent of the families have positive marginal willingness to pay for barn eggs compared to battery eggs, 46 percent of the families have for free-range eggs and 60 percent have for organic eggs.

Comparing results of separate estimations using data from one store at a time shows that the estimated distribution of marginal willingness to pay for egg types varies from store to store, and that the standard deviations of the distributions decrease when the heterogeneity of the stores in the subsample decreases. The differences are not surprising since simple examinations of the data revealed that the customers in different stores vary not only on conventional socio-demographics, but also on habits and attitudes. Furthermore, the variety and price structure vary from store to store. In stores that focus on high quality rather than price (e.g. SuperBrugsen), marginal willingness to pay for non-battery eggs is generally positive. In stores that focus on price rather than quality (e.g. Bilka), marginal willingness to pay for non-battery eggs is generally negative.

In general the heterogeneity of marginal willingness to pay for organic eggs was higher than for barn eggs and free-range eggs. This is seen as a result of the fact that organic eggs encompass more attributes than the other egg types, and that variation may occur in the valuation of each of these attributes (animal welfare, environment, health and possibly more).

It is also found that allowing the preferences for different egg types to depend on background variables such as geographical location of household residence improves the model significantly, and that allowing the preferences for different egg types to depend on attitudes, such as attitude to branded goods, also improves the model. The effect of the age of the main buyer was, in most cases, not significant.

Marginal willingness to pay for non-battery eggs compared to battery eggs is highest in SuperBrugsen and lowest in Bilka (when aggregates including very heterogeneous stores are excluded). Marginal willingness to pay is generally higher in city municipalities and among households that prefer branded goods to be sure to get good quality.

The detailed documentation of the entire GfK data set in chapter 4 provides an opportunity for other researchers to gain information about this new and very extensive data set. The documentation demonstrates that the GfK data covers many different aspects of food consumption and that estimating the value of animal welfare only exploits a fraction of the potential of the data. In particular, the vast amount of background information calls for further study.

This study has demonstrated that the GfK data can be applied when valuing labels indicating non-market goods such as animal welfare related to eggs or environmentally friendly production. The virtue of the data is the vast amount of information that is available, but this is also the Achilles' heel.

The practical problems revealed and solved in chapter 5 show that working with detailed panel data on purchases of food is not to be taken lightly, but requires great consideration and thoroughness. The solutions provided by this study may, in many cases, be transferred to other studies, and the proposals for refinement of the methods may lead to better results than the ones obtained here.

The statistical basis of the new and advanced econometric technique known as the mixed multinomial logit was presented in chapter 3 along with an outline of the simulation technique used when estimating the model. This study has demonstrated that estimations using mixed multinomial logit are computationally feasible and have many attractive features, of which the heterogeneity of the estimated marginal willingness to pay was the main one exploited here.

In principle, the parameters in the mixed multinomial logit may follow any distribution, but the estimation programs available today only include a limited number of functional forms. In real life, many people are actually indifferent between, e.g., two types of eggs, but as long as continuous distributions are used alone, the special case of zero marginal willingness to pay will always have zero probability, and in most cases counter-intuitively high negative values of marginal willingness to pay will have positive probability. Expanding the set of distributions to allow functional forms that are more behaviourally realistic would be a natural extension of this study.

Summing up, the present study provides a theoretical foundation for the concept of marginal willingness to pay and the mixed multinomial logit estimation technique. It also demonstrates that the new and promising mixed multinomial logit model can be combined with the GfK data set that provides huge amounts of complicated information. The combination raises several practical problems, but the study proposes solutions to these, and the solutions are used in the actual estimations using the data. These estimations show that mixing improves the explanatory power of the models estimated using the GfK data, and that the mixed multinomial logit estimations are computationally feasible. The mixed multinomial logit and the GfK data can, therefore, be recommended for use in further studies.

# References


- *Baltzer, K. (2002):* Efterspørgslen efter fødevarekvalitet og –sikkerhed: Et pilot-studie af danske forbrugeres efterspørgsel efter æg. Working paper from Danish Research Institute of Food Economics (FØI). Available at www.foi.dk.
- *Ben-Akiva, M., D. Bolduc and M. Bradley (1993):* Estimation of travel choice models with randomly distributed values of time. Transportation Research Record, **1413**: 88-97.
- *Ben-Akiva, M. and D. Bolduc (1996):* Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working paper, Department of Civil Engineering, Massachusetts Institute of Technology.
- *Ben-Akiva, M. and S. R. Lerman (1985):* Discrete choice analysis. Massachusetts Institute of Technology, Cambridge, MA, USA.
- *Bhat, C. (1996):* Accommodating variations in responsiveness to level-of-service measures in travel model choice modelling. Working paper, Department of Civil Engineering, University of Massachusetts at Amherst.
- *Bjørner, T. B., L. G. Hansen, C. S. Russel and T. Olsen (2002):* The effect of the Nordic Swan Label on Consumers' choice. AKF memo, Institute of Local Government Studies – Denmark. Available at www.akf.dk.
- *Bockstael, N. E. (1995):* Travel cost models. Chapter 29, pp 655-671, *in* The handbook of environmental economics, edited by D. W. Bromley. Blackwell Publ. Ltd., Oxford, UK.
- *Boyd, J. and J. Mellman (1980):* The effect of fuel economy standards on the U.S. automotive market: a hedonic demand analysis. Transportation Research, **14A** (5-6): 367-378.
- *Brownstone, D. and K. Train (1999):* Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics, **89:** 109-129.
- *Cardell, N, and F. Dunbar (1980):* Measuring the societal impacts of automobile downsizing. Transportation Research, **14A** (5-6): 423-434.
- *Danish Plant Directorate ('Plantedirektoratet') (2000):* Vejledning om økologisk jordbrugsproduktion. August, ISSN 1600-4760.
- *Danish Poultry Association ('Dansk Fjerkræråd').* Available at www.poultry.dk/.
- *Deaton, A. and J. Muellbauer (1980):* Economics and consumer behaviour. Cambridge University Press, Cambridge, UK.
- *European Commission (1991):* Rådets forordning (EOEF) nr 1274/91 af 15. maj 1991 om gennemførelsesbestemmelser til forordning (EOEF) nr 1907/90 om handelsnormer for æg (EF-Tidende nr. L 121 af 16/05/1991 page 0011-0024).

- *GfK (2001):* GfK–ConsumerScan (Dansk husstandspanel) metode og standardinformationer.
- *Greene, W. H. (2000):* Econometric analysis, fourth edition. Prentice Hall, Upper Saddle River, New Jersey, USA.
- *Hanemann, W. M. (1984):* Welfare evaluations in contingent valuation experiments with discrete responses. American Journal of Agricultural Economics, **66** (3): 332-341.
- *Hensher, D. A. and W. H. Greene (2001):* The mixed logit model: The state of practice and warnings for the unwary. 28 November 2001. Available at www.stern.nyu.edu/~wgreene/MixedLogitSOP.pdf.
- *Layton, D. F. and G. Brown (2000):* Heterogeneous preferences regarding global climate change. The Review of Economics and Statistics, **82** (4): 616-624.
- *LØJ ('Landsforeningen Økologisk Jordbrug'), Danish Association of Organic Farming (2000):* Avlsgrundlag og forskelsregler for økologisk jordbrug, August 24.
- *McFadden, D. (1973):* Conditional logit analysis of qualitative choice behaviour, pp. 105-142, *in* Frontiers in econometrics, edited by P. Zarembka, Academic Press, New York. USA.
- *McFadden, D. and K. Train (2000):* Mixed MNL models for discrete response. Journal of Applied Econometrics, **15**: 447-470.
- *Ministry of Agriculture ('Landbrugsministeriet') (1992):* Bekendtgørelse om økologisk jordproduktion. Bekendtgørelse nr. 720 August 19. 1992, amended September 11, 1992.
- *Morey, E., W. D. Shaw, and R. Rowe (1991):* A discrete-choice model of recreational participation, site choice and activity valuation when complete trip data are not available. Journal of Environmental Economics and Management, **22:** 181-201.
- *Økologisk Landsforening (2002):* Forbrugernotat (Consumer note). Århus, Danmark.
- *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.
- *Rouwendal, J. and E. Meijer (2001):* Preferences for housing, jobs and commuting: A mixed logit analysis. Journal of Regional Science, **41 (**3): 475-505.
- *Statistics Denmark*, Statistikbanken, www.dst.dk
- *Statistics Denmark (2002):* NYT fra Danmarks Statistik, Emnegruppe: Miljø og energi. Familiernes miljøvaner 2002. **379**, 25. september 2002.

- *Train, K. (1998):* Recreation demand models with taste differences over people. Land Economics, **72**(2): 230-239.
- *Train, K. (1999a):* Halton sequences for mixed logit. Working paper, Department of Economics, University of California, Berkeley.
- *Train, K. (1999b):* Mixed logit models for recreation demand, pp. 121-140, *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.
- *Train, K. (1999c):* The file 'manual.txt'. Included in the file 'train0299.zip' from Train's MMNL homepage elsa.berkeley.edu/Software/abstracts/train0296.html.
- *Train, K., D. Revelt and R. Ruud (1999):* The program 'mxlg.g'. Included in the file 'train0299.zip' from Train's MMNL homepage elsa.berkeley.edu/Software/abstracts/train0296.html.

## Appendix A  GfK commodity groups in the years 1998 to 2001

| Vg. | 1998 | Vg. | 1999 | Vg. | 2000 | Vg. | 2001 |
|---|---|---|---|---|---|---|---|
| 1 | Smør og blandingsprodukter | 1 | Smør og blandingsprodukter | 1 | Smør og blandingsprodukter | 1 | Smør og blandingsprodukter |
| 2 | Margarine | 2 | Margarine | 2 | Margarine | 2 | Margarine |
| 3 | Fast ost i skiver og stk. | | Flytter til varegruppe 33 i løbet af 1998 | | | | |
| | | 4 | Fløde & Pulverfløde | 4 | Fløde & Pulverfløde | | |
| 5 | Dessertoste | 5 | Dessertoste | 5 | Dessertoste | 5 | Dessertoste |
| 6 | Æg | 6 | Æg | 6 | Æg | 6 | Æg |
| 7 | Alle former for mælk | 7 | Alle former for mælk | 7 | Alle former for mælk | 7 | Alle former for mælk |
| 8 | Alle former for surmælksprodukter | 8 | Alle former for surmælksprodukter | 8 | Surmælksprodukter & Creme fraiche | 8 | Surmælksprodukter & Creme fraiche |
| 9 | Rugbrød | 9 | Rugbrød | 9 | Rugbrød | 9 | Rugbrød |
| 10 | Hvedebrød | 10 | Hvedebrød | 10 | Hvedebrød | 10 | Hvedebrød |
| 11 | Dybfrosne Pizzaer | 11 | Dybfrosne Pizzaer | 11 | Dybfrosne Pizzaer | | |
| 12 | Instant Kaffe og Kakao mv. | 12 | Instant Kaffe og Kakao mv. | 12 | Instant Kaffe og Kakao mv. | 12 | Instant Kaffe og Kakao mv. |
| 13 | Mejeri snacks | 13 | Mejeri snacks | 13 | Mejeri snacks | 13 | Mejeri snacks |
| 15 | Alm. -/ Koffeinfri Kaffe | 15 | Alm. -/ Koffeinfri Kaffe | 15 | Alm. -/ Koffeinfri Kaffe | 15 | Alm. -/ Koffeinfri Kaffe |
| 16 | Knækbrød | 16 | Knækbrød | 16 | Knækbrød | 16 | Knækbrød |
| 17 | The | 17 | The | 17 | The | 17 | The |
| 18 | Alle former for isthe | 18 | Alle former for isthe | 18 | Alle former for isthe | 18 | Alle former for isthe |
| 19 | Sukker, stødt melis | 19 | Sukker, stødt melis | 19 | Sukker & sødemidler | 19 | Sukker |
| 20 | Hvedemel | 20 | Hvedemel | 20 | Mel | 20 | Mel |
| 21 | Alle former for cerealer | 21 | Alle former for cerealier | 21 | Alle former for cerealier | 21 | Alle former for cerealier |
| | | 22 | Is,Isdesserter,Sorbet & Yoghurt-is | 22 | Is,Isdesserter,Sorbet & Yoghurt-is | 22 | Is,Isdesserter,Sorbet & Yoghurt-is |
| 23 | Kiks & krackers, marengs, vafler, søde bisquits & sammenlagte kiks | 23 | Kiks & krackers, marengs, vafler, søde bisquits & sammenlagte kiks | 23 | Kiks & krackers, marengs, vafler, søde bisquits & sammenlagte kiks | 23 | Kiks & krackers, marengs, vafler, søde bisquits & sammenlagte kiks |
| 24 | Fiskekonserves | 24 | Fiskekonserves | 24 | Fiskekonserves | | |
| 25 | Flåede tomater og surprod. i glas | 25 | Flåede tomater og surprod. i glas | 25 | Survarer, Frugt- & Grøntkonserves | | |

| Vg. | 1998 | Vg. | 1999 | Vg. | 2000 | Vg. | 2001 |
|---|---|---|---|---|---|---|---|
| 26 | Koldprod., butterprod., honningsnittter, makroner og kokosmakroner | 26 | Koldprod., butterprod., honningsnittter, makroner og kokosmakroner | 26 | Koldprod., butterprod., honningsnittter, makroner og kokosmakroner | 26 | Koldprod., butterprod., honningsnittter, makroner og kokosmakroner |
| 27 | Bouillon / Suppe / Krydderterninger | 27 | Bouillon / Suppe / Krydderterninger | 27 | Bouillon / Suppe / Krydderterninger | 27 | Bouillon / Suppe / Krydderterninger |
| 28 | Traditionelle småkager, kransekager, kammerjunkere og julesmåkager | 28 | Traditionelle småkager, kransekager, kammerjunkere og julesmåkager | 28 | Traditionelle småkager, kransekager, kammerjunkere og julesmåkager | 28 | Traditionelle småkager, kransekager, kammerjunkere og julesmåkager |
| 29 | Alle former for bagemix | 29 | Alle former for bagemix | 29 | Alle former for bagemix | | |
| 30 | Marmelade, syltetøj og gele | 30 | Marmelade, syltetøj og gele | 30 | Marmelade, syltetøj og gele | 30 | Marmelade, syltetøj og gele |
| 31 | Kyllinger og andet fjerkræ | 31 | Kyllinger og andet fjerkræ | 31 | Kyllinger og andet fjerkræ | 31 | Kyllinger og andet fjerkræ |
| 32 | Is,Isdesserter,Sorbet & Yoghurt-is | 32 | Is,Isdesserter,Sorbet & Yoghurt-is | 32 | Dybfrost grønsager & grøntsagsblandinger | 32 | Dybfrost grønsager & grøntsagsblandinger |
| 33 | Fast ost i skiver og stk. | 33 | Fast ost i skiver og stk. | 33 | Fast ost i skiver og stk. | 33 | Fast ost i skiver og stk. |
| 34 | Skærekager, roulader og tærter | 34 | Skærekager, roulader og tærter | 34 | Skærekager, roulader og tærter | 34 | Skærekager, roulader og tærter |
| 35 | Spaghetti, pasta og nudler | 35 | Spaghetti, pasta og nudler | 35 | Spaghetti, pasta og nudler | 35 | Spaghetti, pasta og nudler |
| 36 | Løse ris og grødris | 36 | Løse ris og grødris | 36 | Løse ris og grødris | 36 | Løse ris og grødris |
| 37 | Alle former for sovs og kryddermix | 37 | Alle former for sovs og kryddermix | 37 | Saucer & Kryddermix | 37 | Saucer & Kryddermix |
| 38 | Pålægs- og smørechokolade | 38 | Pålægs- og smørechokolade | 38 | Pålægs- og smørechokolade | 38 | Pålægs- og smørechokolade |
| 39 | Sennep, ketchup og tomatpure | 39 | Sennep, ketchup og tomatpure | 39 | Sennep, ketchup og tomatpure | 39 | Ketchup |
| 40 | Bordvin- Rød/ Hvid / Rose | 40 | Bordvin- Rød/ Hvid / Rose | 40 | Bordvin- Rød/ Hvid / Rose | 40 | Bordvin- Rød/ Hvid / Rose |
| 41 | Juice og saft | 41 | Juice og saft | 41 | Juice og saft | 41 | Juice og saft |
| 42 | Sodavand | 42 | Sodavand | 42 | Sodavand | 42 | Sodavand |
| 43 | Øl | 43 | Øl | 43 | Øl | 43 | Øl |
| 44 | Hedvin, Aperitif, champagne mv. | 44 | Hedvin, Aperitif, champagne mv. | 44 | Hedvin, aperitif, champagne, cider og Lambrusco. | 44 | Hedvin, aperitif, champagne, cider og Lambrusco. |
| 45 | Snaps | 45 | Snaps | 45 | Spiritus | | |
| 46 | Æbler,bananer,tomater,frugt | 46 | Æbler,bananer,tomater,frugt | 46 | Frugt | 46 | Frugt |
| 47 | Kartofler, løg, gulerødder, grønsag. | 47 | Kartofler, løg, gulerødder, grønsag. | 47 | Grøntsager | 47 | Grøntsager |
| 48 | Spegepølse, kødpålæg og bacon | 48 | Spegepølse, kødpålæg,bacon & leverpostej | 48 | Leverpostej, bacon & paté | 48 | Leverpostej, bacon & paté |
| 49 | Mayonnaise, remoulade og salater | 49 | Mayonnaise, remoulade og salater | 49 | Mayonnaise, remoulade og salater | | |
| | | 50 | Mad- & Spiseolie | 50 | Mad- & Spiseolie | 50 | Mad- & Spiseolie |

| Vg. | 1998 | Vg. | 1999 | Vg. | 2000 | Vg. | 2001 |
|---|---|---|---|---|---|---|---|
| 51 | Salatdressinger | 51 | Salatdressinger | 51 | Salatdressinger | 51 | Salatdressinger |
| 52 | Pølser, lamme-, svine-, okse- & hakket kød | 52 | Pølser, lamme-, svine-, okse- & hakket kød | 52 | Pølser, lamme-, svine-, okse- & hakket kød | 52 | Pølser, lamme-, svine-, okse- & hakket kød |
| 53 | Konfekture | 53 | Konfekture | 53 | Frugtgrød | | |
| | | 54 | Kødpålæg | 54 | Kødpålæg | 54 | Kødpålæg |
| 55 | Alle former for fisk og skaldyr | 55 | Alle former for fisk og skaldyr | 55 | Alle former for fisk og skaldyr | 55 | Alle former for fisk og skaldyr |
| 56 | Pasta-,nudel- og risretter | 56 | Pasta-,nudel- og risretter | 56 | Pasta-,nudel- og risretter | 56 | Pasta-,nudel- og risretter |
| 57 | Vaskepulver | 57 | Vaskepulver | 57 | Vaskepulver | 57 | Vaskepulver |
| | | | | 58 | Wienerbrød | 58 | Wienerbrød |
| 59 | Afkalkningsmidler | 59 | Afkalkningsmidler | 59 | Afkalkningsmidler | 59 | Afkalkningsmidler |
| 60 | Skyllemidler | 60 | Skyllemidler | 60 | Skyllemidler | 60 | Skyllemidler |
| 61 | Opvaskemidler | 61 | Opvaskemidler | 61 | Opvaskemidler | 61 | Opvaskemidler |
| 62 | Alle former for rengøringsmidler | 62 | Alle former for rengøringsmidler | 62 | Alle former for rengøringsmidler | 62 | Alle former for rengøringsmidler |
| 64 | Toiletpapir | 64 | Toiletpapir | 64 | Toiletpapir | 64 | Toiletpapir |
| 65 | Køkkenruller | 65 | Køkkenruller | 65 | Køkkenruller | 65 | Køkkenruller |
| 66 | Menstruationsmidler | 66 | Menstruationsmidler | 66 | Menstruationsmidler | 66 | Menstruationsmidler |
| 68 | Blade og magasiner | 68 | Blade og magasiner | 68 | Blade og magasiner | 68 | Blade og magasiner |
| 70 | Bodyshampoo og skumbad | 70 | Bodyshampoo og skumbad | 70 | Bodyshampoo og skumbad | 70 | Bodyshampoo og skumbad |
| 71 | Hånd- og toiletsæbe | 71 | Hånd- og toiletsæbe | 71 | Hånd- og toiletsæbe | 71 | Hånd- og toiletsæbe |
| 72 | Tandpasta og mundskyllevand | 72 | Tandpasta og mundskyllevand | 72 | Tandpasta og mundskyllevand | 72 | Tandpasta og mundskyllevand |
| 73 | Hårshampoo | 73 | Hårshampoo | 73 | Hårshampoo | 73 | Hårshampoo |
| | | | | 75 | Smørrebrøds & Bagepapir | 75 | Smørrebrøds & Bagepapir |
| 77 | Alle former for hårstyling | 77 | Alle former for hårstyling | 77 | Alle former for hårstyling | 77 | Alle former for hårstyling |
| 79 | Hårbalsam | 79 | Hårbalsam | 79 | Hårbalsam | 79 | Hårbalsam |
| 80 | Potteplanter,blomster og buketter | 80 | Potteplanter,blomster og buketter | 80 | Potteplanter,blomster og buketter | 80 | Potteplanter,blomster og buketter |
| | | | | 81 | Kosmetik | 81 | Kosmetik |
| | | | | 82 | Kildevand | 82 | Kildevand |
| 83 | Blomster og urte-/grønsagsfrø | 83 | Blomster og urte-/grønsagsfrø | | | | |
| 87 | Deodoranter | 87 | Deodoranter | 87 | Deodoranter | 87 | Deodoranter |
| 88 | Alle former for creme og lotion | 88 | Alle former for creme og lotion | 88 | Hudplejeprodukter | 88 | Hudplejeprodukter |
| | | 89 | Hårfarve & Toning | 89 | Hårfarve & Toning | 89 | Hårfarve & Toning |
| 90 | Vitaminer, mineraler og kosttilskud | 90 | Vitaminer, mineraler og kosttilskud | 90 | Vitaminer, mineraler og kosttilskud | 90 | Vitaminer, mineraler og kosttilskud |
| 99 | Distrib | 99 | Distrib | 99 | Distrib | 99 | Distrib |

# Appendix B Background questionnaire, overview



**Table B.1 Background information about the household in general, collected once a year by GfK 1997-2001**

| |
|---|
| Household number |
| Municipality ('kommune') |
| County ('amt') |
| Type of home |
| Ownership of home |
| Gender and age of persons in household |
| Number of years in school for mother and father |
| Length of post-school education for mother and father |
| Degree of occupation for mother and father |
| Type of current occupation for mother and father |
| Annual household income in DKK<br>• 1997 – 1999: 0-50.000, 50.000-100.000 … 400.000-450.000, more than 450.000<br>• 2000 – 2001 : 0-100.000, 100.000-150.000 … 550.000-600.000, more than 600.000 |
| Membership of clubs or organisations<br>• New clubs added and other clubs removed in 1998 and 2000 |
| Electronic and other equipment in household<br>• New items added in 2000 |
| How does the household receive TV signals? |
| Which TV-channels does the household receive?<br>• New channels added in 1998, 2000 and 2001. |

Source: Background questionnaires from GfK 1997 to 2001

**Table B.2 Background information about the person answering the questionnaire (probably the the person mainly responsible for shopping ('dagligvareindkøberen')), collected once a year by GfK, 1997-2001**

| |
|---|
| Gender of the person mainly responsible for shopping (independent of who is answering the questionnaire) |
| Name of grocery store where the household does most of its shopping |
| Distribution of budget for everyday necessities (on types of stores, not names) |
| *Sunday* papers (11 different names):<br>• How many editions out of 6 do you read? (for each paper)<br>• Which one would you be most reluctant to do without? |
| *Daily* papers (13 different names):<br>• How many editions out of 6 do you read? (for each paper)<br>• Which one would you be most reluctant to do without? |
| *Weekly* magazines (9 – 13 different names):<br>1. How many editions out of 6 do you read? (for each magazine)<br>2. Which one would you be most reluctant to do without?<br>• New magazines added in 1998 and 2000 |
| Magazines (15 – 28 different names):<br>1. How many editions out of 6 do you read? (for each magazine)<br>2. Which one would you be most reluctant to do without?<br>• New magazines added in 1999 and 2000 |
| How many days in a week do you watch the following TV-channels? |
| How many of the catalogues/flyers sent to you by the shops do you read? |
| How important are the special offers you see in the catalogues/flyers to the way you shop? |
| Do you read catalogues/flyers at random or do you always read catalogues/flyers from the same stores? |
| Do you prefer brand labels to cheaper products? |
| Do you look for special offers when shopping? |
| Are low prices important for your choice of store? |

Source: Background questionnaires from GfK 1997 to 2001

**Table B.3 Background information about the main income provider, collected once a year by GfK 1997-2001**

| |
|---|
| Who is the main income provider?<br>• Changes from father/mother to Diary keeper/spouse in 2000 |
| Current (and for retired, previous) occupation for main income provider |
| Type of work for main income provider |

Source: Background questionnaires from GfK 1997 to 2001

**Table B.4 Background information about the household in general, collected once a year by GfK only for parts of the 1997-2001 period**

| **(October 1997 – September 1998 etc.)** | **1997** | **1998** | **1999** | **2000** | **2001** |
|---|---|---|---|---|---|
| Do you have access to e-mail?<br>2000:Yes/No, 2001: At home/at work | | | | X | X |
| Approximate amount of household budget per month (available for housekeeping, clothes, pleasures, and buying new things after paying rent, loans, insurance and phone) | X | X | X | | |
| How do you use the Internet? (separate answers for daily shopper, spouse and children) | | | | X | X |
| How many hours a week do you spend on the Internet? (separate answers for daily shopper, spouse and children) | | | | X | X |
| How many cars are there in the household? | | | X | | |
| Date of birth of main user of each car | | | X | | |
| Is the car privately owned or is it a company car? (for each car) | | | X | | |
| How often does the household use the following convenience products? | | | X | X | X |
| Why do you use these products? | | | X | X | X |
| Why are these products not used more often? | | | X | X | X |
| How do you and your family feel about cooking? | | | X | X | X |
| How often does the household bake bread, cakes etc.? | | | X | X | X |
| How often does your household eat breakfast together? | | | | X | X |
| How often does your household eat dinner together? | | | | X | X |
| How many days in a week do the household cook its own dinner? | | | | X | X |

Source: Background questionnaires from GfK 1997 to 2001

**Table B.5 Background information about the person answering the questionnaire (probably the the person mainly responsible for shopping ('dagligvareindkøberen')), collected once a year by GfK only for parts of the 1997-2001 period**

| **(October 1. 1997 – September 30. 1998 etc.)** | **1997** | **1998** | **1999** | **2000** | **2001** |
|---|---|---|---|---|---|
| Name of grocery store *closest to home* | | | | X | X |
| Grade 25 different shopping chains on: | | | | | |
| Interior and atmosphere | | | | X | X |
| Service | | | | X | X |
| Assortment | | | | X | X |
| Quality | | | | X | X |
| Parking | | | | X | X |
| Price level | | | | X | X |
| Location of the store | | | | X | X |
| General assessment | | | | X | X |
| Which attribute is *most* important to you when deciding how satisfied you are with a shop? | | | | X | X |
| Which attribute is the *second most* important to you when deciding how satisfied you are with a shop? | | | | X | X |
| How often do you go to the cinema during a 6-month period? | X | X | X | | |
| Interest in different types of television programme | X | X | X | | |
| Which TV-channels have you watched for at least 15 minutes within the last week? | X | X | X | | |
| How many catalogues/flyers from food shops does your household receive every week? | | | | X | X |
| How many of these catalogues/flyers do you read or browse each week? | | | | X | X |
| How much time do you spend reading catalogues/flyers each week? | | | | X | X |
| Why do you primarily read the catalogues/flyers? | | | | X | X |
| How do you experience catalogues/flyers? | | | | X | X |
| Which of the following competitions in which you are supposed to collect points or parts of the packing<br>1. Are you *aware of?*<br>2. Have your household *participated in?* | | X | X | | |
| How much time do you spend (average) cooking dinner on a weekday? (Do not include time for shopping, marinating etc.) | | | X | X | X |

Source: Background questionnaires from GfK 1997 to 2001

# Appendix C Background questionnaire, details



**Table C.1 Detailed background information about the household in general**

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| Household number | | X | X | X | X | X |
| Municipality (kommune) | | X | X | X | X | X |
| County (amt) | • København Amtskommune, Københavns Kommune and Frederiksberg Kommune<br>• Frederiksborg Amt<br>• Roskilde Amt<br>• Vestsjællands Amt<br>• Storstrøms Amt<br>• Bornholms Amt<br>• Fyns Amt<br>• Sønderjyllands Amt<br>• Ribe Amt<br>• Vejle Amt<br>• Ringkøbing Amt<br>• Århus Amt<br>• Viborg Amt<br>• Nordjyllands Amt | X | X | X | X | X |
| Type of home | • One-family house/ terraced house<br>• Farm<br>• Two-family house<br>• Flat or three-family house<br>• Flat divided into bed-sitting rooms (‘klubværelse’) or rented room | X | X | X | X | X |
| Ownership of home | • Tenant (‘bor til leje’)<br>• Owns house or farm<br>• Free or official residence<br>• Flat under a multi-ownership scheme (‘andelslejlighed’)<br>• Owner-occupied flat | X | X | X | X | X |
| Gender and date of birth for all persons in household | | X | X | X | X | X |
| Do you have access to e-mail? | • Yes/no | | | | X | X |
| | • Yes, at work<br>• Yes, at home<br>• Yes, both at work and at home<br>• No | | | | | X |
| Number of years in school for mother and father | • 7th grade<br>• 8th/9th grade<br>• 10th grade or ‘Mellemskole/real-eksamen’<br>• High school/HF/HH | X | X | X | X | X |
| Length of post-school education for mother and | • Name of education<br>• Length of education beyond school | X | | | | |

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| father | 1. 'erhvervsrettet uddannelse' (e.g. EFG, Elev, Lærling, Laborant, Sygehjælper)<br>2. Short advanced studies (e.g. Pædagogiske, tekniske, politibetjent)<br>3. Medium long advanced studies (e.g. Folkeskolelærer, HD/HA, sygeplejerske)<br>4. Long advanced studies (e.g. University, Psykolog, læge, Cand. Merc.) | | X | X | X | X |
| Degree of occupation for mother and father | 0. On leave<br>1. Full time (at least 30 hours per week)<br>2. Part time (16-29 hours per week)<br>3. Part time (less than 15 hours per week)<br>4. Self-employed or assisting spouse<br>5. Homemaker/housewife<br>6. Old age pensioner<br>7. Other pensioner<br>8. Unemployed<br>9. Student | X | X | X | X | X |
| Type of current occupation for mother and father | 1. Self-employed in primary sector<br>2. Self employed in other sectors<br>3. White collar<br>4. Blue collar | X | X | X | X | X |
| Annual household income | 0. 0-49.999 DKK<br>1. 50.000-99.999<br>2. 100.000-149.999<br>3. 150.000-199.999<br>4. 200.000-249.999<br>5. 250.000-299.999<br>6. 300.000-349.999<br>7. 350.000-399.999<br>8. 400.000-449.999<br>9. 450.000 or more | X | X | X | | |
| | 1. 0-99.999 DKK<br>2. 100.000-149.999<br>3. 150.000-199.999<br>4. 200.000-249.999<br>5. 250.000-299.999<br>6. 300.000-349.999<br>7. 350.000-399.999<br>8. 400.000-449.999<br>9. 450.000-499.999<br>10. 500.000-549.999<br>11. 550.000-599.000<br>12. 600.000 or more | | | | X | X |
| Approximate amount at disposal per month (available for housekeeping, clothes, pleasures, and buying new things after paying rent, loans, insurance and phone) | 0. 0-1000 DKK<br>1. 1001-2000<br>2. 2001-3000<br>3. 3001-4000<br>4. 4001-5000<br>5. 5001-6000<br>6. 6001-7000<br>7. 7001-8000<br>8. 8001 or more | X | X | X | | |

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| Membership of clubs or organizations | • Number 1 (Dagrofa's indkøbsklub) | X | | | | |
| | • Sparklubben (pr. 1.11.1996) | X | X | | | |
| | 1. FDB/brugsforeningerne<br>2. LIC (Lærernes Indkøbs Central)<br>3. Mini-Klubben (Kvickly)<br>4. Ronald McDonald's Fødselsdagsklub<br>5. Fætter BR-klubben<br>6. Anders And's bogklub<br>7. Børnenes trafikklub<br>8. Libero-klubben<br>9. Club-Dillen (Legekæden)<br>10. Kræftens Bekæmpelse | X | X | X | X | X |
| | • Forbrugsforeningen af 1866<br>• FDM<br>• Ikea Family Club<br>• Statoil Premium Club<br>• TV2-Lorrys medlemsklub<br>• Other | | X | X | X | X |
| Electronic equipment in household | 1. TV<br>2. VCR<br>3. Dishwasher<br>4. Washing machine<br>5. Tumble drier<br>6. Microwave oven<br>7. PC<br>8. Access to Internet<br>9. Modem<br>10. CD-rom drive<br>11. Printer<br>12. Telefax<br>13. Answering machine<br>14. Mobile Phone | X | X | X | X | X |
| | • Freezer | X | | | | |
| | • Freezer separated from refrigerator | | X | X | X | X |
| | • Laptop<br>• Scanner for PC<br>• DVD player | | | | X | X |
| How do you use the Internet? (separate answers for daily shopper, spouse and children) | 1. Shopping<br>2. Private economy (banking)<br>3. Games/searching for information/hobby<br>4. Surfing/chatting<br>5. Working or studying<br>6. Never uses the Internet | | | | X | X |
| How many hours a week do you spend on the Internet? (separate answers for daily shopper, spouse and children) | Number of hours | | | | X | X |
| How does the household receive TV signals? | 1. Community antenna<br>2. 'Hybridnet'/cable<br>3. Own ordinary antenna<br>4. Own 'parabol' | X | X | X | X | X |

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| Which TV-channels does the household receive? | • ZTV/TV3+ | X | | | | |
| | 1. DR1<br>2. DR2<br>3. TV2<br>4. TV3 | X | X | X | X | X |
| | • TV3+ | | X | X | X | X |
| | • TvDanmark | | X | X | X | |
| | • DK4<br>• Eurosport<br>• MTV<br>• CNN International | | | | X | X |
| | • TV2 Zulu<br>• TvDanmark1<br>• TvDanmark2 | | | | | X |
| How many cars are there in the household? | | | | X | | |
| Date of birth for the main user of each car | | | | X | | |
| Is the car privately owned or is it a company car? (for each car) | | | | X | | |
| How often does the household use the following convenience products? | • Frozen lasagna/pizza<br>• Instant sauces<br>• Instant soup<br>• Frozen pasta/rice meals<br>• Instant bouillon ('Bouillonterninger')<br>• Spice mixes like 'Knorr mexican meal spice mix' | | | X | X | X |
| Why do you use these products? | 1. Easy and quick<br>2. Gives variation<br>3. I/my family likes the taste of it<br>4. It is a good emergency solution | | | X | X | X |
| Why are these products not used more often? | Open question, translated by GfK to closed categories | | | X | X | X |
| How do you and your family feel about cooking? | 1. Bad<br>2. Neither good nor bad<br>3. Good<br>4. Quite good<br>5. Really good | | | X | X | X |
| How often does the household bake bread, cakes etc.? | 1. Once a week<br>2. Once every two weeks<br>3. Once a month<br>4. Once every 2-3 months<br>5. Once every 6 months<br>6. Never | | | X | X | X |
| How often does your household eat *breakfast* together? | 1. Never/rarely ever<br>2. 1-2 times a week<br>3. 3-4 times a week<br>4. Daily (5-7 times a week)<br>5. Only on weekends | | | | X | X |
| How often does your household eat *dinner* together? | As above | | | | X | X |
| How many days in a week does the household *cook* its own dinner? | As above | | | | X | X |

Source: Background questionnaires from GfK 1997 to 2001

**Table C.2 Detailed background information About the person mainly responsible for shopping, 'dagligvareindkøberen'** (or the person answering the questionnaire)

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| Gender | 1. Female<br>2. Male | X | X | X | X | X |
| Name of grocery store where the household does *most of its shopping* | 160 Different store names | X | X | X | X | X |
| Name of grocery store *closest to home* | 160 Different store names | | | | X | X |
| Distribution of budget for everyday necessities | % spent in:<br>• Discount stores (Bilka, OBS)<br>• Discount (e.g. Aldi, Alta, Fakta, Netto, Remma 1000)<br>• Department stores (e.g. Føtex, Kvickly, Mega, S&E)<br>• Large supermarkets (e.g. Super Brugsen, Irma, Favør, Prima)<br>• Minimarkets/small supermarkets (e.g. DagliBrugsen, Oceka, Super 1, neighborhood shops, local grocers)<br>• 'Specialbutikker' (Butcher, greengrocer, fishmonger, baker etc.)<br>• Other shopping places (e.g. kiosks, gas stations, wholesale etc.) | X | X | X | X | X |

| **Background information** | **Codes** | **1997** | **1998** | **1999** | **2000** | **2001** |
|---|---|---|---|---|---|---|
| Grade 25 different shopping chains on: | • Aktiv Super<br>• Aldi<br>• Bilka<br>• Dagli'Brugsen<br>• De friske Butikker<br>• Edeka<br>• Fakta<br>• Favør<br>• Focus<br>• Føtex<br>• Irma<br>• Iso<br>• Kvickly<br>• LokalBrugsen<br>• Løvbjerg<br>• Merko<br>• Netto<br>• OBS!<br>• Prima<br>• REMA 1000<br>• Spar<br>• Suma<br>• SuperBest<br>• SuperBrugsen<br>• SuperNærkøb | | | | | |
| Interior and atmosphere | 1. Very good<br>2. Good<br>3. Neither good nor bad<br>4. Bad<br>5. Very bad | | | | X | X |
| Service | As above | | | | X | X |
| Assortment | As above | | | | X | X |
| Quality | As above | | | | X | X |
| Parking | As above | | | | X | X |
| Price level | As above | | | | X | X |
| Location of the store | As above | | | | X | X |
| General assessment | As above | | | | X | X |
| Which attribute is *most* important for you when deciding how satisfied you are with a shop? | 1. Interior and atmosphere<br>2. Service<br>3. Assortment<br>4. Quality<br>5. Price level<br>6. Parking<br>7. Location of the store<br>8. General assessment of the store | | | | X | X |
| Which attribute is the *second most* important for you when deciding how satisfied you are with a shop? | As above | | | | X | X |

| **Background information** | **Codes** | **1997** | **1998** | **1999** | **2000** | **2001** |
|---|---|---|---|---|---|---|
| How often do you go to the cinema during a 6-month period? | 1. >12<br>2. 8-12<br>3. 5-7<br>4. 2-4<br>5. 1<br>6. <1<br>7. 0 | X | X | X | | |
| Interest in different types of television programs | 1. Quiz programs<br>2. Talk shows<br>3. Danish entertainment shows ('Greven af hittegodset' etc.)<br>4. Classical music<br>5. Pop/Rock<br>6. News<br>7. Debates (debatprogrammer)<br>8. Documentaries<br>9. Youth-series (Beverly Hills 90210, Melrose Place etc.)<br>10. Glamour-series (Dollars, Dallas, Glamour etc.)<br>11. Detective/action-series (New York Blues, In the Heat of the Night etc.)<br>12. Cartoons<br>13. Programs for children<br>14. Specific sport events (e.g. a game)<br>15. Programs about sports in general<br>16. Comedies (Mash, Roseanne, Keeping up Appearances etc.)<br>17. Action/suspense movies, thrillers<br>18. Family/love films<br>19. Erotic films/programs<br>20. Consumer programs<br>21. Cultural/historical programs (Galleri 11 etc.)<br>22. Nature programs<br>23. Local/regional programs<br>24. Food programs<br>25. Educational programs | X | X | X | | |
| *Sunday* papers:<br>• How many editions out of 6 do you read? (for each paper)<br>• Which one would you be most reluctant to do without? | • Berlingske Tidende Søndag<br>• B.T. Søndag<br>• Ekstra Bladet Søndag<br>• Morgenposten (Fyns Stiftstidende Søndag)<br>• Jyllandsposten Søndag<br>• Jydske Vestkysten Søndag<br>• Politiken Søndag<br>• Weekendavisen (fredag)<br>• NordJyskSøndag (Aalborg Stiftstidende Søndag)<br>• Århus Stiftstindende Søndag<br>• Søndagsavisen (gratis omdelt) | X | X | X | X | X |

| **Background information** | **Codes** | **1997** | **1998** | **1999** | **2000** | **2001** |
|---|---|---|---|---|---|---|
| *Daily* papers:<br>• How many editions out of 6 do you read? (for each paper)<br>• Which one would you be most reluctant to do without? | • Berlingske Tidende<br>• B.T.<br>• Børsen<br>• Det fri Aktuelt (1997, hereafter:) Aktuelt<br>• Ekstra Bladet<br>• Fyens Stiftstidende<br>• Jyllandsposten<br>• Jydske Vestkysten<br>• Politiken<br>• Aalborg Stiftstidende<br>• Århus stiftstidende<br>• The local paper in my area (not free papers)<br>• Information | X | X | X | X | X |
| *Weekly* magazines:<br>• How many editions out of 6 do you read? (for each magazine)<br>• Which one would you be most reluctant to do without? | • Alt for Damerne<br>• Billed Bladet<br>• Familie Journalen<br>• Femina<br>• Hendes verden<br>• Hjemmet<br>• Se & Hør<br>• Ude & Hjemme<br>• Ugemagasinet Søndag (Søndags B.T.) | X | X | X | X | X |
| | • Her & Nu<br>• Tæt På<br>• Kig Ind | | X | X | X | X |
| | • Anders And | | | | X | X |

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| Magazines:<br>• How many editions out of 6 do you read? (for each magazine)<br>• Which one would you be most reluctant to do without? | • Alt om Mad (4-6 times a year)<br>• Bo Bedre (monthly)<br>• Det Bedste (monthly)<br>• Mad og Bolig-magasinet (4-6 times a year)<br>• Månedsbladet IN (monthly)<br>• Samvirke (monthly)<br>• I Form (monthly)<br>• Haven/Alt om haven (monthly)<br>• Helse (monthly)<br>• Idé-Nyt (4-6 times a year)<br>• Forældre og Børn (4-6 times a year)<br>• Damernes Verden (monthly)<br>• Illustreret Videnskab (monthly)<br>• Tidens Kvinder (4-6 times a year)<br>• Mit Livs Novelle (monthly) | X | X | X | X | X |
| | • Eurowoman (monthly)<br>• Euroman (monthly)<br>• Men's Health (monthly) | | | X | X | X |
| | • Ud og se med DSB<br>• Bilen Motor og sport<br>• Motor<br>• Bådnyt<br>• Gør det selv<br>• Computer for alle<br>• Alt om håndarbejde<br>• Vi Unge<br>• Mix<br>• Chilli | | | | X | X |
| Which TV-channels have you watched for at least 15 minutes within the last week? | • ZTV/TV3+ | X | | | | |
| | 1. DR1<br>2. DR2<br>3. TV2<br>4. TV3 | X | X | X | | |
| | • TV3+ | | X | X | | |
| | • TvDanmark | | X | X | | |
| How many days in a week do you watch the following TV-channels? | • ZTV/TV3+ | X | | | | |

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| | • DR1<br>• DR2<br>• TV2<br>• TV2 regional TV<br>• TV3 | X | X | X | X | X |
| | • TV3+ | | X | X | X | X |
| | • TvDanmark | | X | X | X | |
| | • DK4 | | | | X | X |
| | • TV2 Zulu<br>• TvDanmark1<br>• TvDanmark2 | | | | | X |
| How many of the flyers sent to you by the shops do you read? | 1. All of them<br>2. Most of them<br>3. About half of them<br>4. Less than half of them<br>5. Almost none of them<br>6. Never read them | X | X | X | X | X |
| | • Does not receive flyers ('reklamer nej tak') | | | | X | X |
| How important are the special offers you see in the flyers to the way you shop? | 1. Very important<br>2. Some importance<br>3. Small importance<br>4. Not important | X | X | X | X | X |
| Do you read flyers at random or do you always read flyers from the same stores? | 1. At random<br>2. From the same stores | X | X | X | X | X |
| How many flyers from food shops does your household receive every week? | 0-10 or more | | | | X | X |
| How many of these flyers do you read or browse each week? | 0-10 or more | | | | X | X |
| How much time do you spend reading flyers each week? | 1. Less than 2 minutes<br>2. 2-5 minutes<br>3. 6-10 minutes<br>4. 11-20 minutes<br>5. 21-29 minutes<br>6. 30-59 minutes<br>7. 1-1½ hours<br>8. 1½-2 hours<br>9. More than 2 hours | | | | X | X |
| Why do you primarily read the flyers? | 1. I read flyers from the shop I usually use<br>2. I look for inspiration for my shopping<br>3. I plan my shopping to save money | | | | X | X |
| How do you experience flyers? | 1. I enjoy reading flyers ('hygger mig')<br>2. I use them to save money<br>3. I browse them to check if my usual store has a reasonable price level | | | | X | X |
| Do you prefer brand labels to cheaper products? | 1. Yes<br>2. No | X | X | X | X | X |
| Do you look for special offers when shopping? | 1. Yes<br>2. No | X | X | X | X | X |
| Are low prices important for your choice of store? | 1. Yes<br>2. No | X | X | X | X | X |

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| Which of the following competitions (1996 and 1997) in which you are supposed to collect points or parts of the packing<br>• are you *aware of ?*<br>• have your household *participated in?* | • Daim Chokolade<br>• Pepsi/Pepsi Max/Seven Up<br>• Nutella | | X | | | |
| | • Coca-Cola/Fanta/Sprite<br>• Merrilds Kaffe<br>• Kims Chips<br>• Gevalia Kaffe<br>• Kelloggs<br>• MD Foods, Lillebror | | X | X | | |
| How much time do you spend (average) cooking dinner on a weekday? (do not include time for shopping, marinating etc.) | 1. Less than ½ hour<br>2. ½ hour<br>3. App. 45 minutes<br>4. 1 hour<br>5. 1½ hour<br>6. 2 hours<br>7. More than 2 hours<br>8. Differs too much, can't give an average<br>9. Don't know<br>10. I don't cook | | | X | X | X |

Source: Background questionnaires from GfK 1997 to 2001

**Table C.3 Detailed background information about the main income provider**

| Background information | Codes | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|---|
| Main income provider | 1. Father<br>2. Mother<br>3. Other | X | X | X | | |
| | 1. Diary keeper/person main responsible for shopping<br>2. Spouse<br>3. Other | | | | X | X |

Source: Background questionnaires from GfK 1997 to 2001

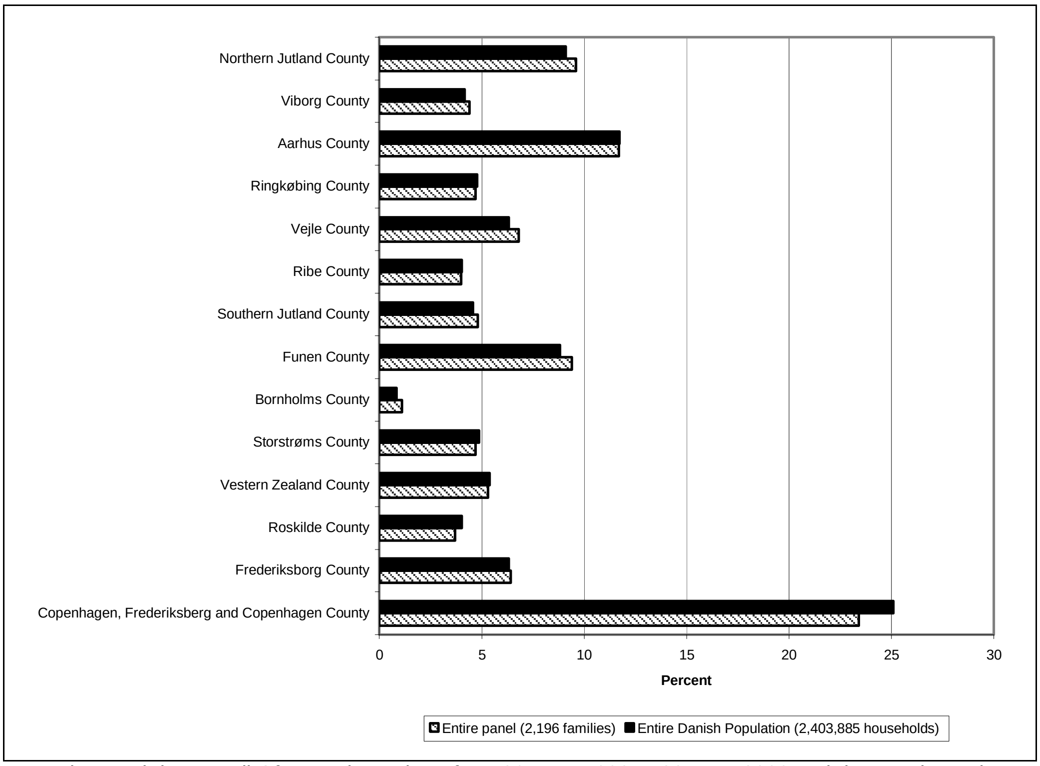
# Appendix D Illustration of background data



Note that some Statistics Denmark data is based on households, other on families. One household may consist of more than one family which means that the number of households does not equal the number of families. Not all families (in both GfK data and Statistics Denmark data) consist of a man and a woman. Data on only one of the two genders will therefore be based on a subsample of the entire sample. The number of families/households may therefore vary from figure to figure.

# D.1 Comparing the panel with the entire Danish population

**Figure D.1.1 Distribution of families on Danish Counties**




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 and data on the entire population in 1999 from Statistics Denmark (BOL4).

**Figure D.1.2 Distribution of families on age of the *male* member of the household**

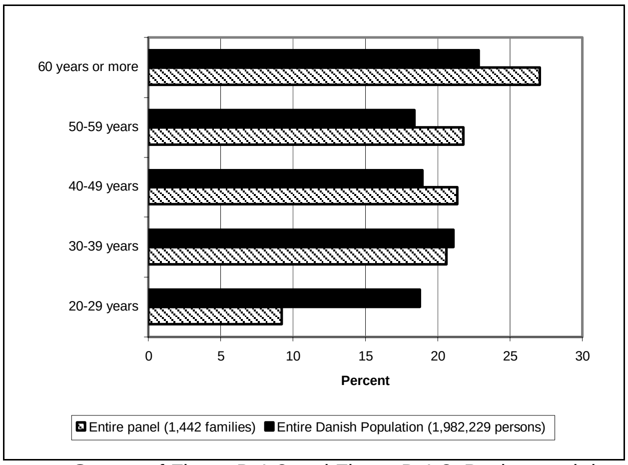
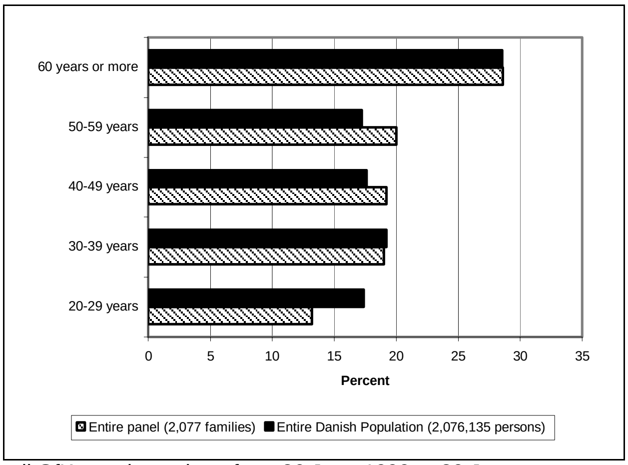
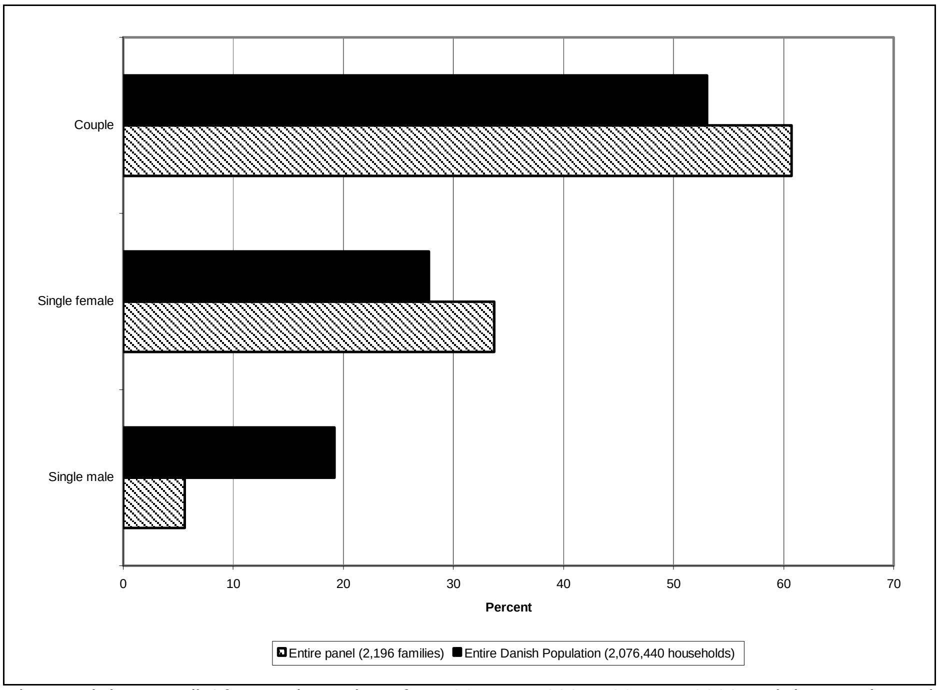



**Figure D.1.3 Distribution of families on age of the *female* member of the household**




Source of Figure D.1.2 and Figure D.1.3: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 and data on the entire population in 1999 from Statistics Denmark (FAM8).
Note for Figure D.1.2 and Figure D.1.3: In the GfK data the lowest category is just '29 or less' but it is used only on 'the father' or 'the mother', so 20-29 is a reasonable approximation, but the group might contain a few individuals under the age of 20. The corresponding age category in Statistics Denmark is 20-29.

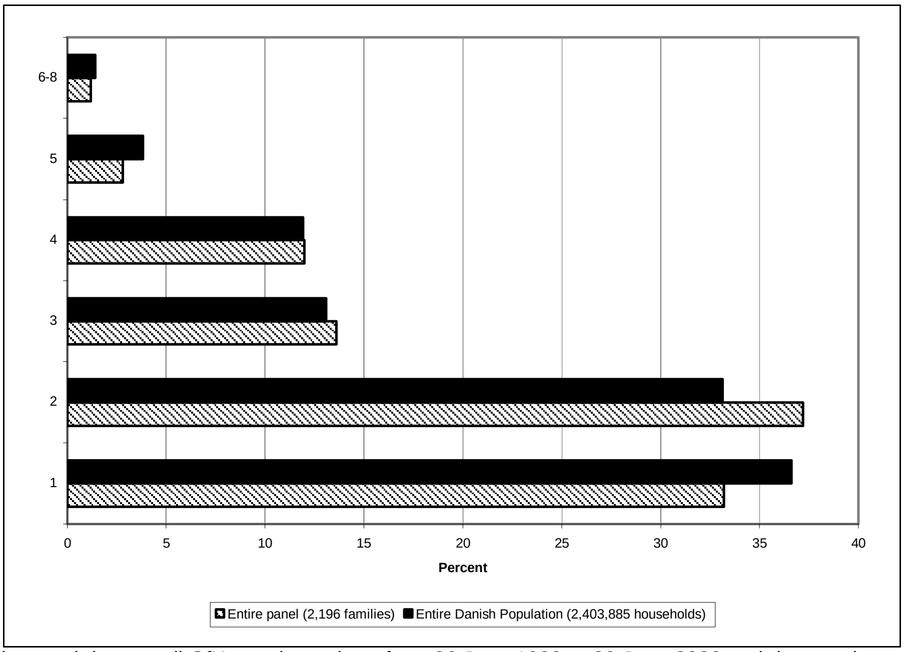
**Figure D.1.4 Distribution of families on 'marital' status of the household**[1]




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 and data on the entire population in 1999 from Statistics Denmark (BOL4).

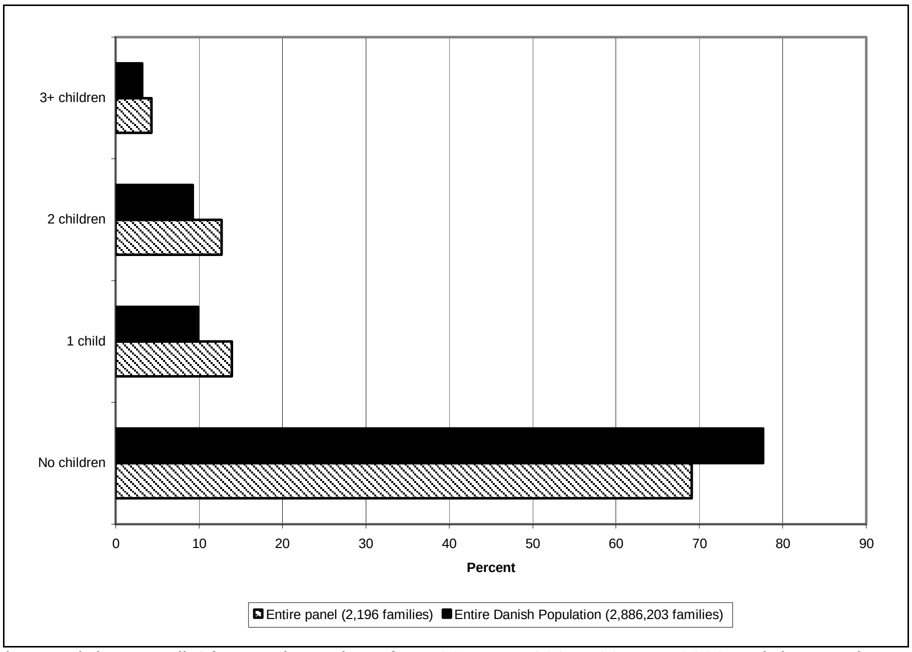
[1] Statistics Denmark participate the households into couples (married and unmarried), singles (males and females), 'children under 18 not living at home', 'households with adult children living at home' and 'other households consisting of more than one family'. The last three categories are inconsistent with the data I have constructed based on information about the individuals in the GfK panel, and I therefore ignore them when calculating the distribution in the entire population. Together the three groups cover 13.6 per cent of all households in Denmark, and ignoring them basically means that I assume that the distribution of single males, single females and couples are the same in these three groups as in the rest of the population. In this case the entire Danish population is represented by 2,093,733 households in stead of 2,423,208.

**Figure D.1.5 Distribution of families on number of persons in the household**




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 and data on the entire population in 1999 from Statistics Denmark (BOL51).

**Figure D.1.6 Distribution of families on number of children in the household**[2]




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 and data on the entire population in 1999 from Statistics Denmark (FAM4).

[2] Notice that data for the entire population is based on families not on households (one household may consist of more than one family in the Statistics Denmark definition). Notice also that a 'child' must be less than 18 years old in the Statistics Denmark definition, but just less than 21 years old in the GfK definition. This contributes to the fact that more families in the panel have children than in the population in general.

**Figure D.1.7 Distribution of families on working status for the male member of the household.**

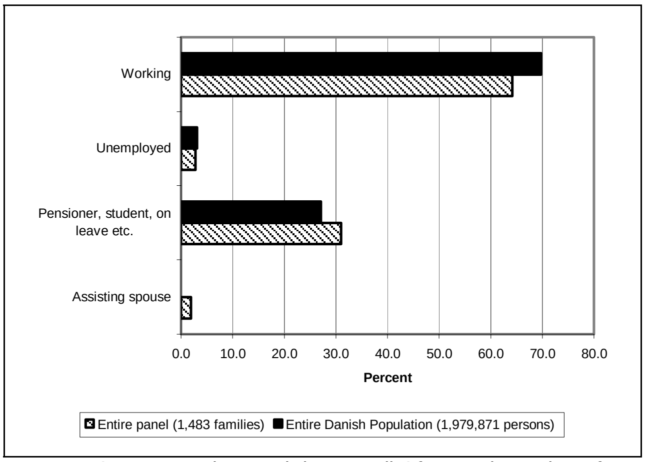
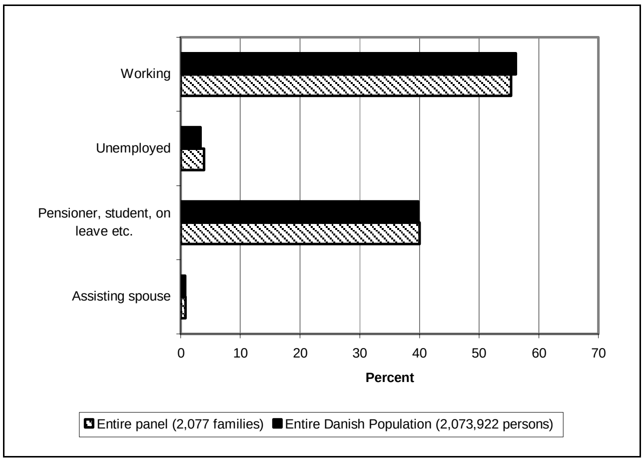
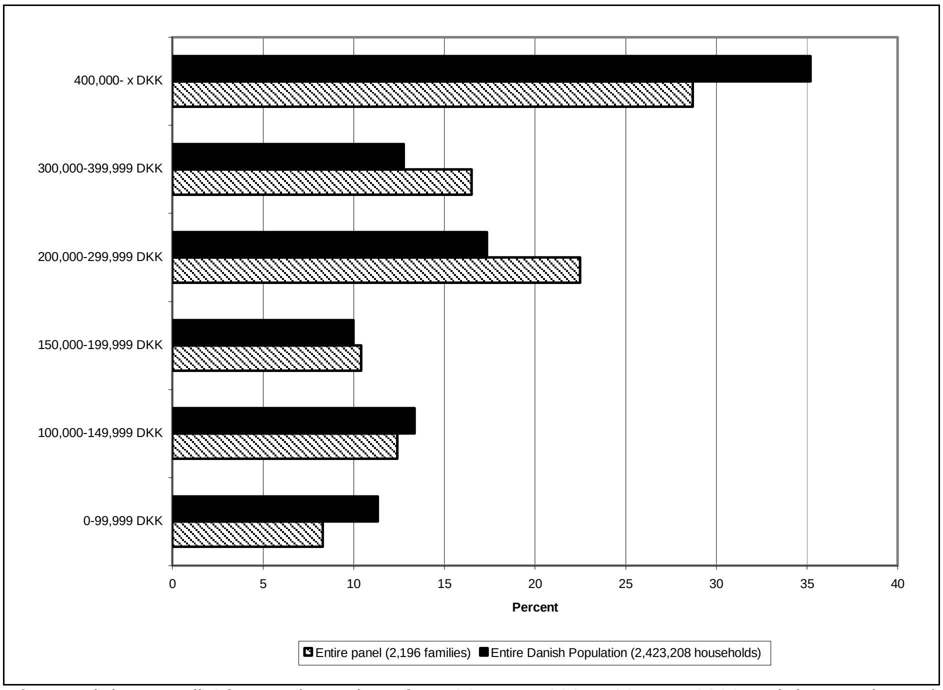



**Figure D.1.8 Distribution of families on working status for female member of the household.**




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 and data on the entire population in 1999 from Statistics Denmark (RAS11 for persons older than 19 years).

**Figure D.1.9 Distribution of families on household income**

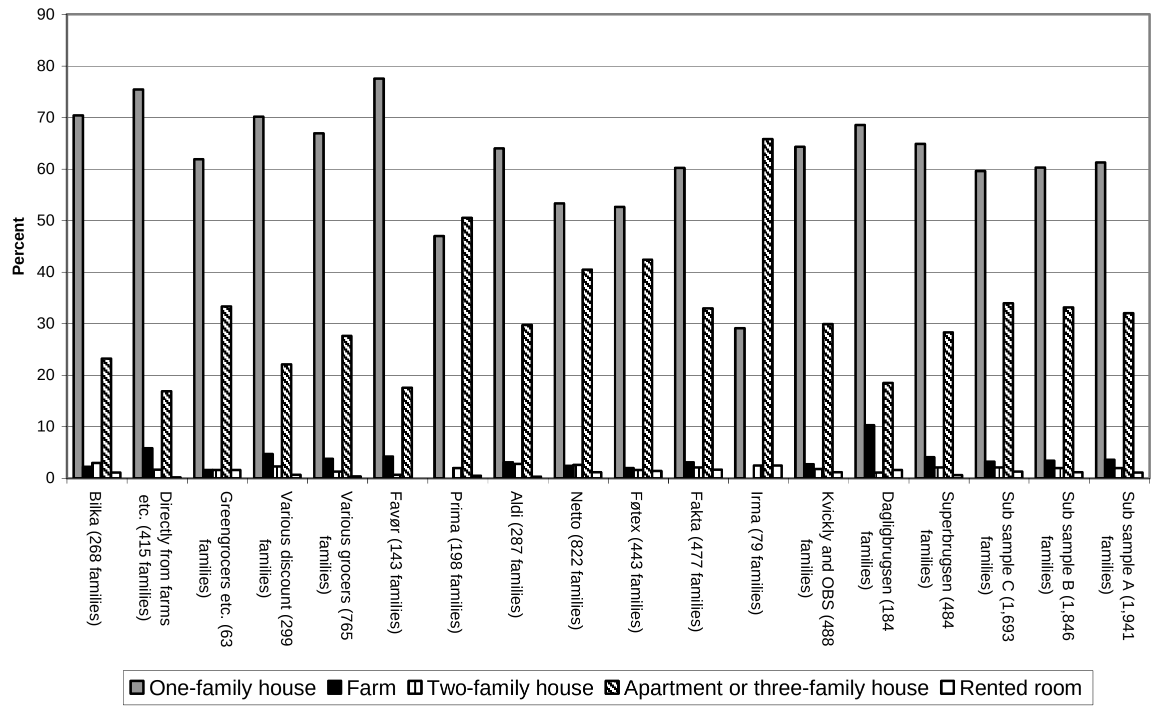



Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 and data on the entire population in 1999 from Statistics Denmark (BIL4).

## D.2 Comparing segments of the data

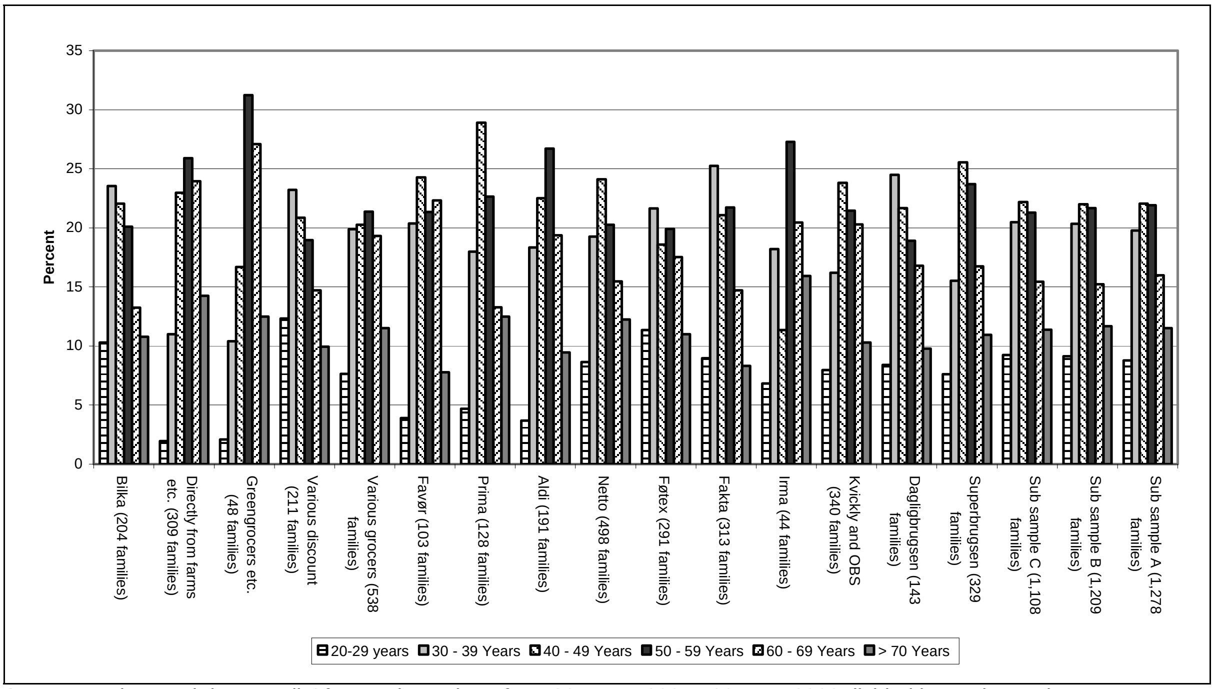
Remember that a family may use several different stores and may therefore be represented more than once in a diagram. All numbers are percent, and since the number of families that use a store varies a great deal the number of families on which the percentages are based is stated in the figures. Keep in mind that the percentages in a small store aggregate like Irma (79 families) is more likely to differ from the total sample (subsample A, 1,941 families) than a big store aggregate like Netto (822 families), simply because each household represents more than one percentage point in Irma but less than 1/8 percentage point in Netto.

**Figure D.2.1 Distribution of families in different subsamples of the GfK data on type of home**



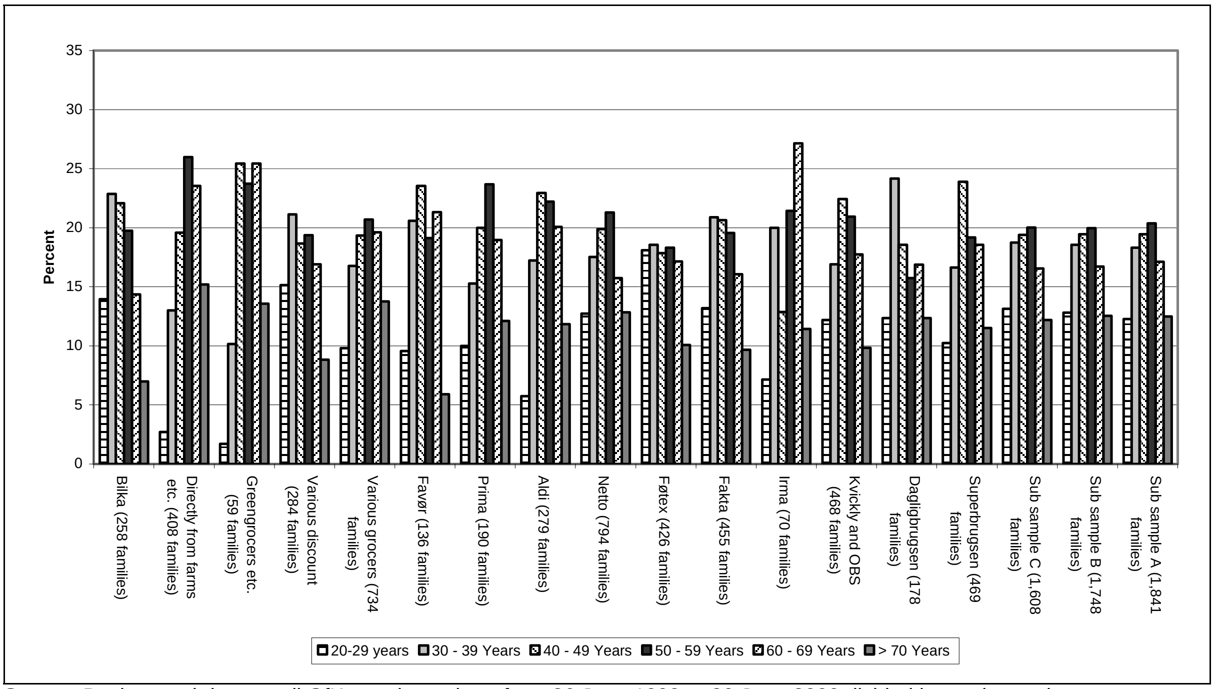

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.2 Distribution of families in different subsamples of the GfK data on father's age**



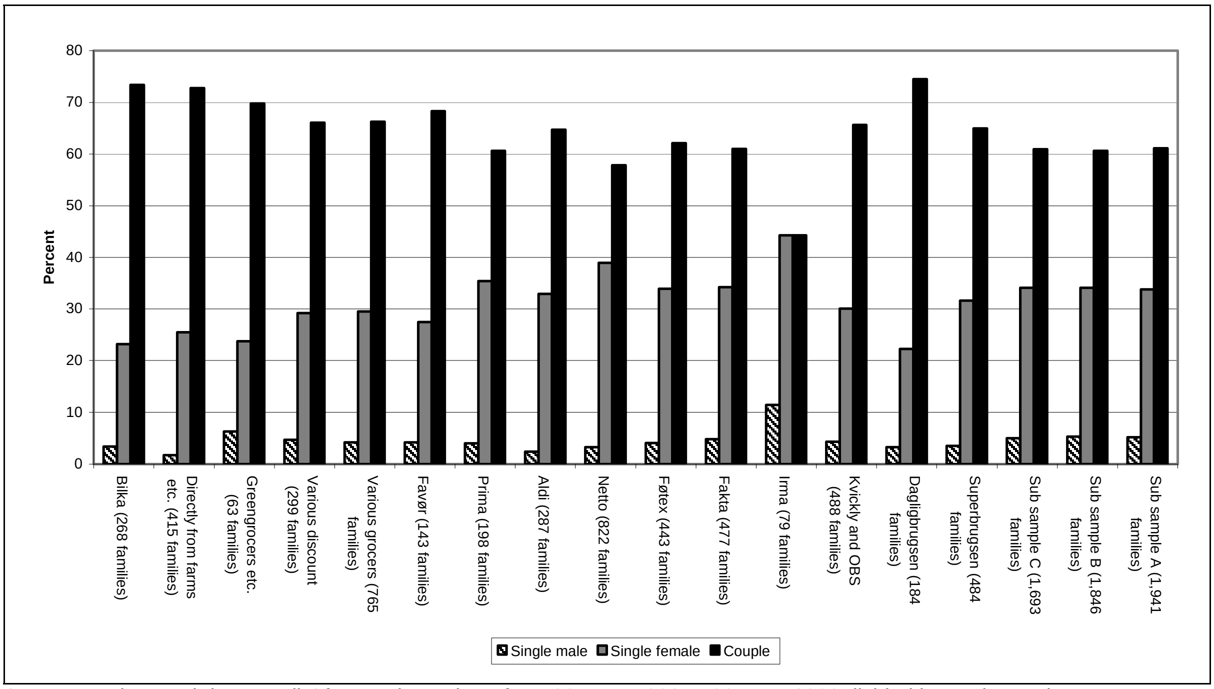

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.3 Distribution of families in different subsamples of the GfK data on mother's age**



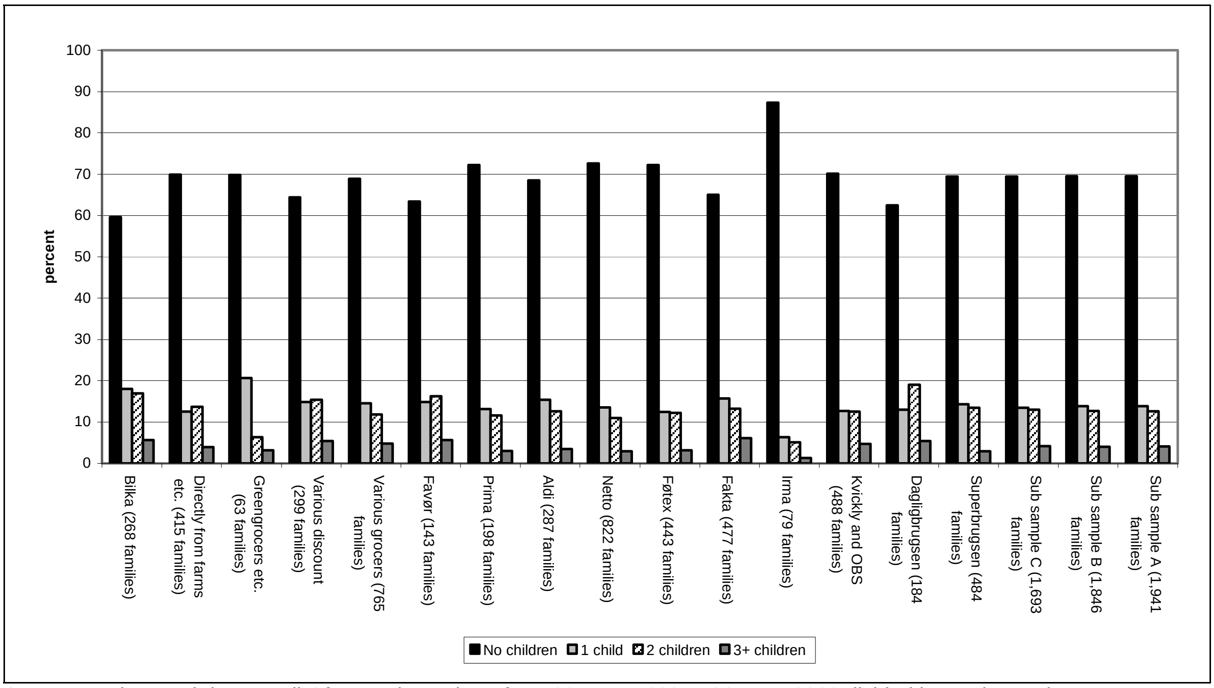

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.4 Distribution of families in different subsamples of the GfK data on 'marital' status**



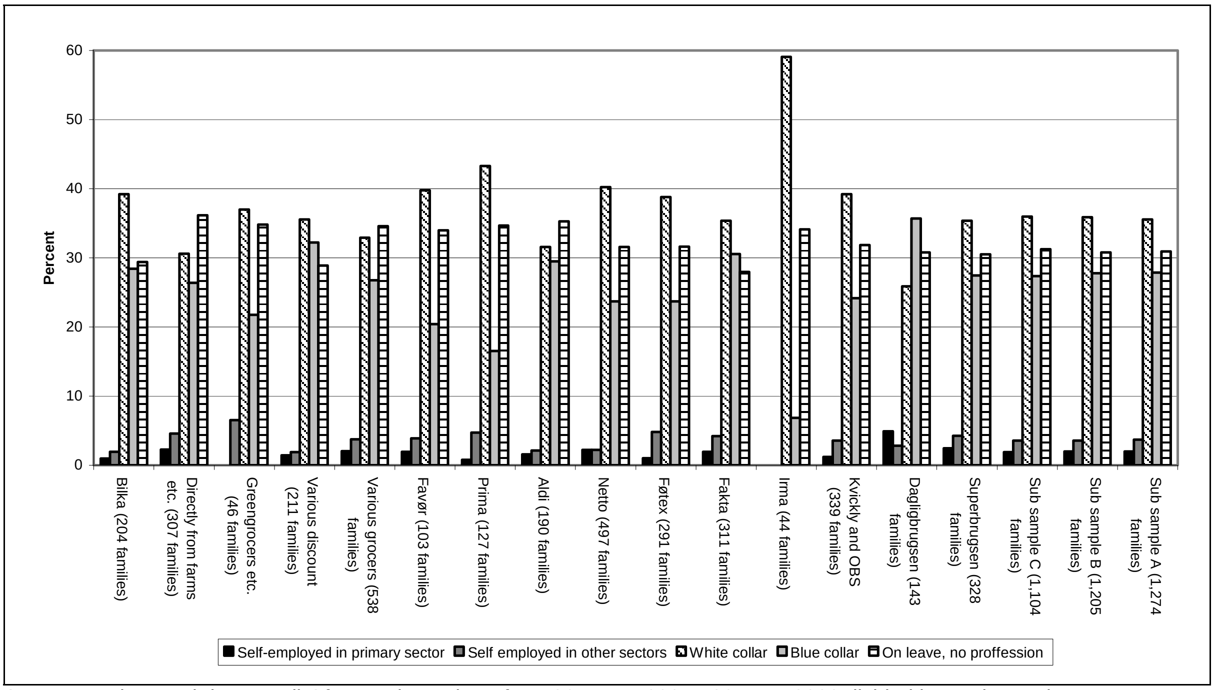

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.5 Distribution of families in different subsamples of the GfK data on number of children in the household**



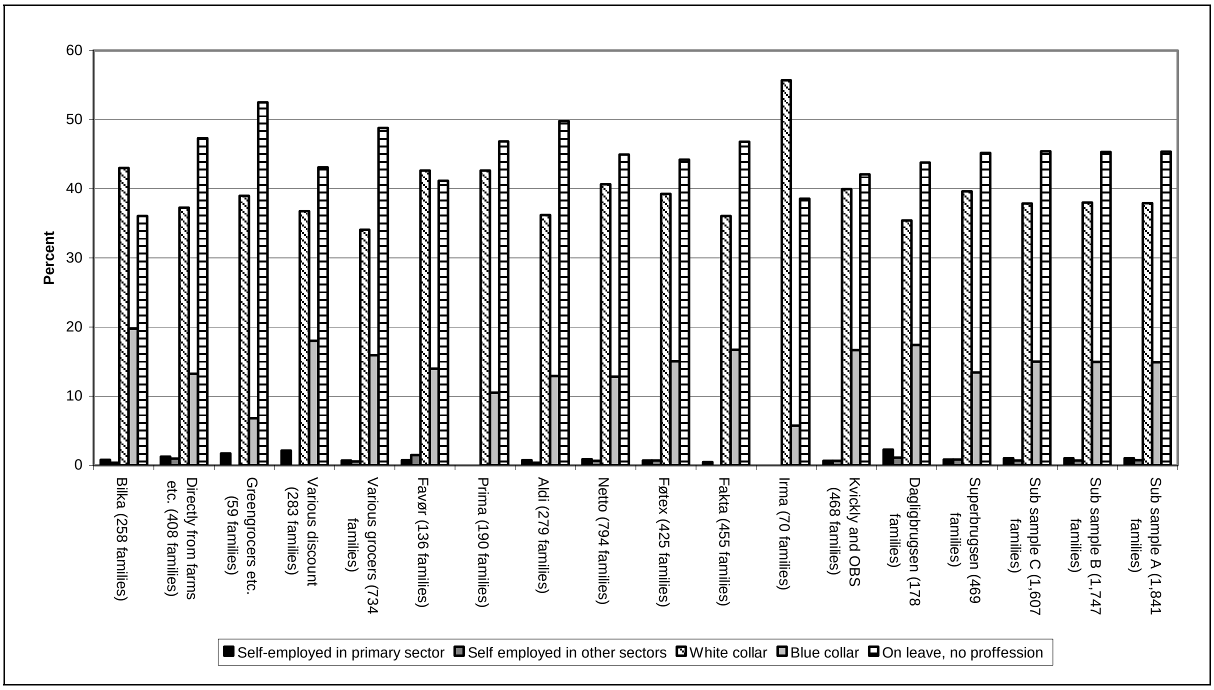

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.6 Distribution of families in different subsamples of the GfK data on father's profession**



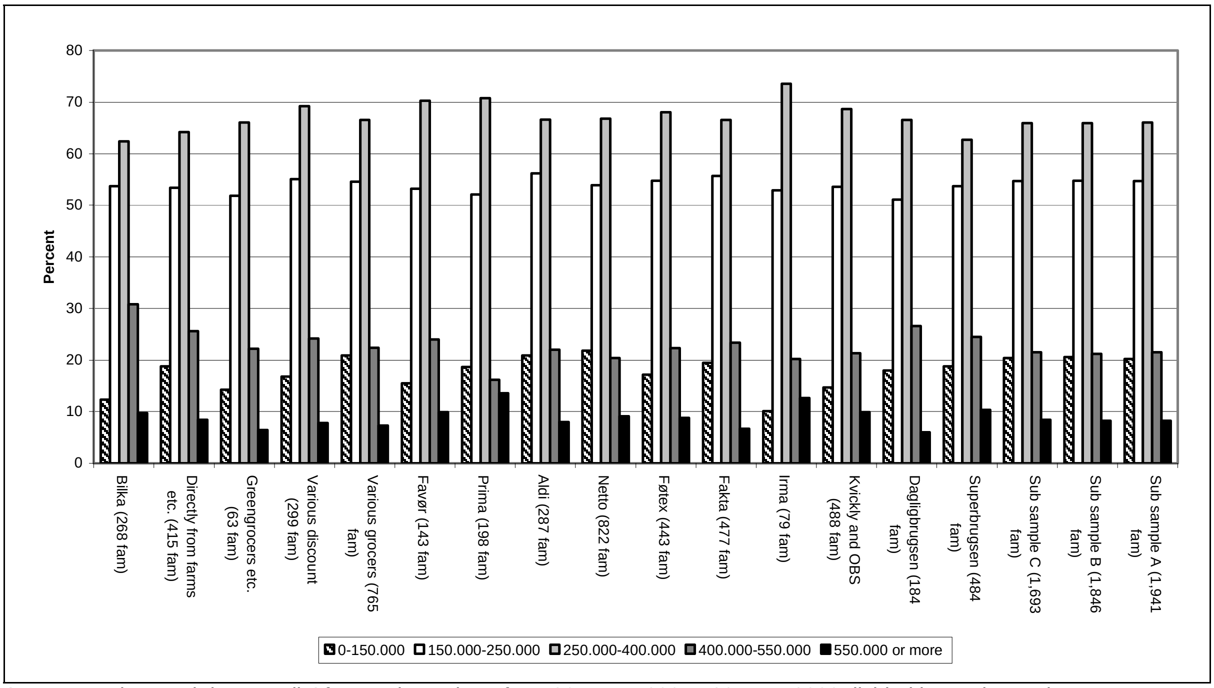

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store.
Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.7 Distribution of families in different subsamples of the GfK data on mother's profession**



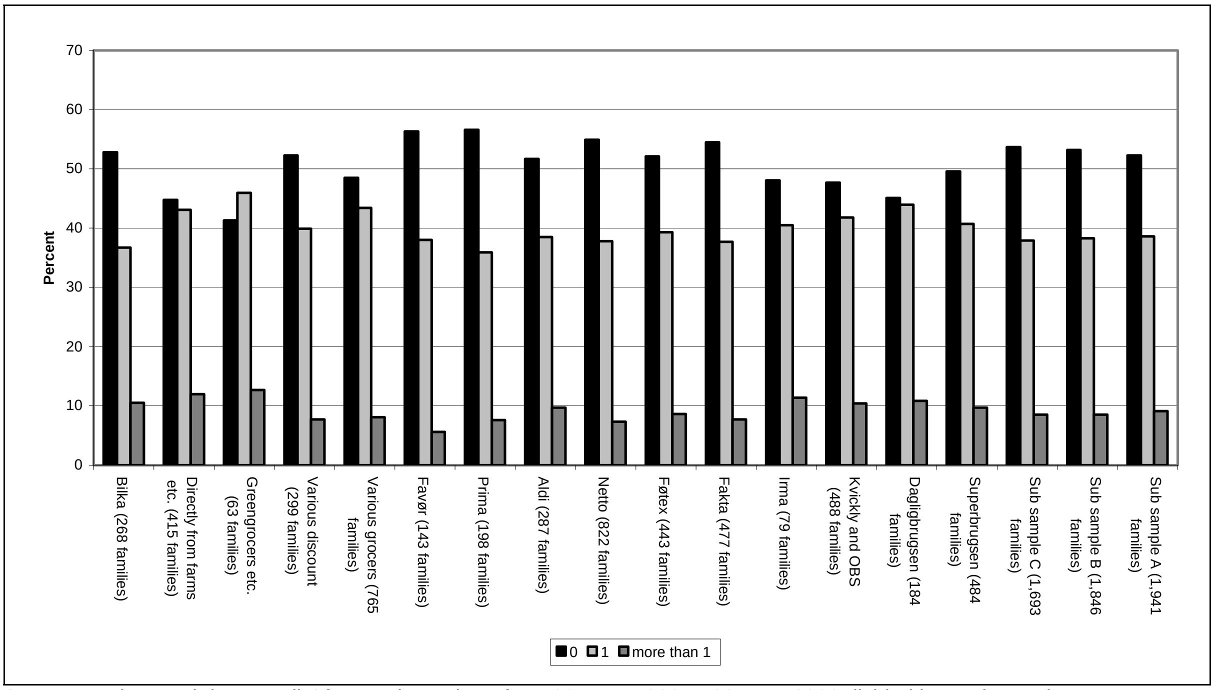

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store.
Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.8 Distribution of families in different subsamples of the GfK data on household income**



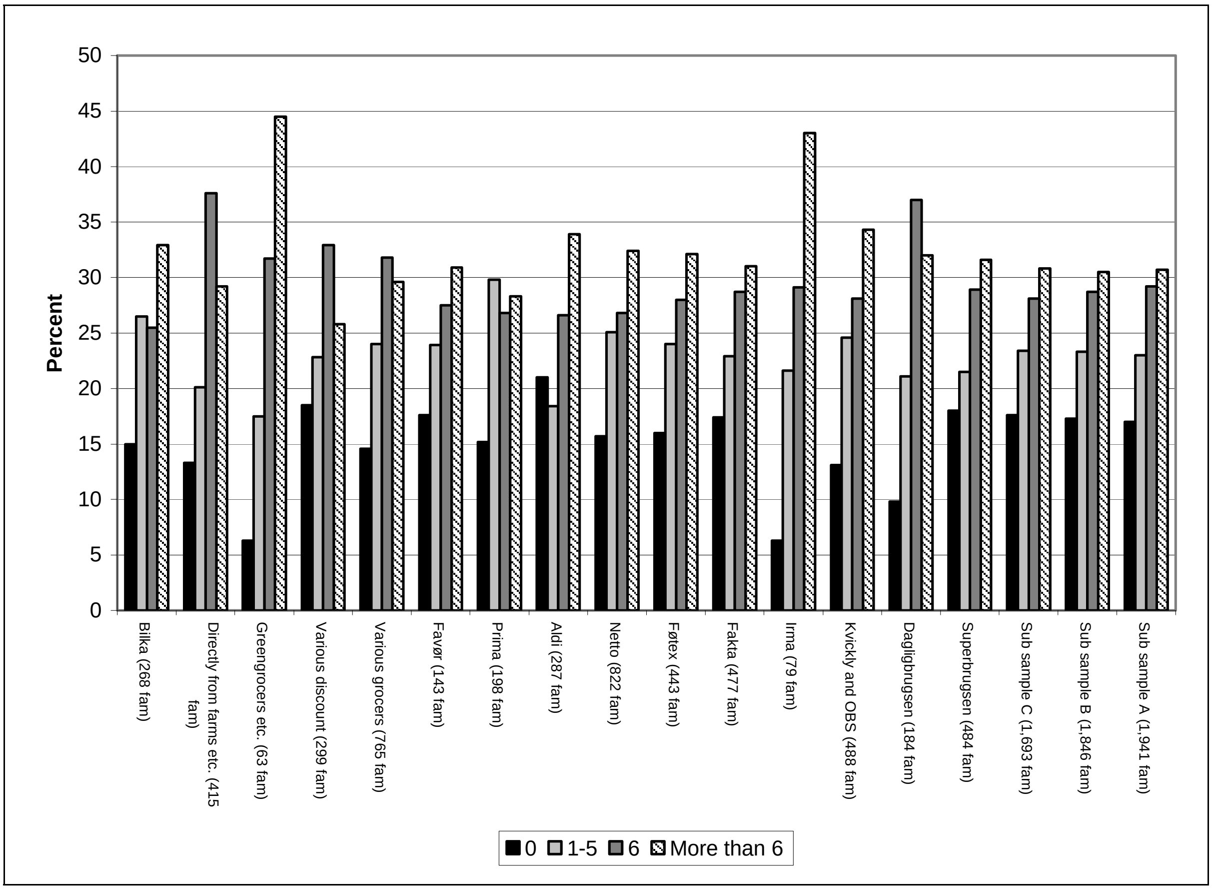

Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store.
Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.9 Distribution of families in different subsamples of the GfK data on number of newspaper subsciptions in the household, weekdays**




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store.
Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

**Figure D.2.10 Distribution of families in different subsamples of the GfK data on number of papers read within the last week (on weekdays)**




Source: Background data on all GfK panel members from 26 June 1999 to 30 June 2000 divided into subsamples.
Note: A family may appear in more than one subsample since purchases can be made in more than one store. Subsample A to C represents nested subsamples of the panel, and subsample A can be used as a measure of the average distribution including all subsamples.

# Appendix E Purchase data on eggs

## E.1 Aggregation of stores

**Table E.1.1 Distribution of sales on aggregated stores**

| Aggregated store | Number of sales in this store | Number of eggs sold in this store | Percentage of all eggs sold to the panel |
|---|---|---|---|
| Superbrugsen | 2379 | 24905 | 8.63 |
| Dagligbrugsen | 636 | 7593 | 2.63 |
| Kvikly and OBS | 1787 | 17838 | 6.18 |
| Irma | 223 | 1734 | 0.60 |
| Fakta (Discount) | 1875 | 19125 | 6.63 |
| Føtex | 1591 | 16510 | 5.72 |
| Netto (Discount) | 4636 | 43863 | 15.20 |
| Aldi (Discount) | 957 | 12138 | 4.21 |
| Prima | 646 | 6811 | 2.36 |
| Favør | 515 | 6352 | 2.20 |
| Various grocers | 3632 | 40245 | 13.94 |
| Various discount stores | 1204 | 16084 | 5.57 |
| Greengrocers etc. | 225 | 2600 | 0.90 |
| Directly from farms | 2798 | 60980 | 21.13 |
| Corner store/ petrol station | 87 | 678 | 0.23 |
| Other stores | 99 | 1174 | 0.41 |
| Bilka | 857 | 10005 | 3.47 |
| *Total* | *24147* | *288635* | *100.00* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, possibly organic too.

**Table E.1.2 Stores in store aggregates**

| Detailed stores in each aggregated store | Number of sales | Number of eggs sold in this store (to the panel) | Percentage of all eggs sold in the aggregated store (to the panel) |
|---|---|---|---|
| **Superbrugsen** | | | |
| Superbrugsen | 2379 | 24905 | 100.0 |
| **Dagligbrugsen** | | | |
| DagliBrugsen | 452 | 5610 | 73.9 |
| Lokalbrugsen (opr. 1/1995) | 164 | 1766 | 23.3 |
| Various stores with less than 50 purchases | 20 | 217 | 2.9 |
| **Kvikly and OBS** | | | |
| Kvickly | 1283 | 12637 | 70.8 |
| OBS | 504 | 5201 | 29.2 |
| **Irma** | | | |
| Irma Supermarked | 223 | 1734 | 100.0 |
| **Fakta (Discount)** | | | |
| Fakta | 1875 | 19125 | 100.0 |
| **Føtex** | | | |
| Føtex | 1591 | 16510 | 100.0 |
| **Netto (Discount)** | | | |
| Netto | 4636 | 43863 | 100.0 |
| **Aldi (Discount)** | | | |
| Aldi | 957 | 12138 | 100.0 |
| **Prima** | | | |
| Prima | 646 | 6811 | 100.0 |

| **Detailed stores in each aggregated store** | **Number of sales** | **Number of eggs sold in this store (to the panel)** | **Percentage of all eggs sold in the aggregated store (to the panel)** |
|---|---|---|---|
| **Favør** | | | |
| Favør | 515 | 6352 | 100.0 |
| **Various grocers** | | | |
| SuperBest Vest for storebælt tidl. Centralkøb | 486 | 6170 | 15.3 |
| Spar/Sparmarked/Kwik Spar/Dankøb (minimarked) | 306 | 3685 | 9.2 |
| Edeka Aktiv Super | 275 | 3460 | 8.6 |
| SuperBest Øst for storebælt tidl. Centralkøb | 284 | 2892 | 7.2 |
| Br. Dreisler | 239 | 2435 | 6.1 |
| Merko | 225 | 2322 | 5.8 |
| Super 1 | 195 | 2308 | 5.7 |
| Iso | 283 | 2263 | 5.6 |
| KC Storkøb (Dagrofa lagerhotel) | 176 | 1946 | 4.8 |
| Løvbjerg | 134 | 1657 | 4.1 |
| Supermarked | 154 | 1571 | 3.9 |
| Super Spar (supermarked) | 120 | 1497 | 3.7 |
| De Friske Butikker | 118 | 1443 | 3.6 |
| Farvorit (Tidl. Centralkøb) | 143 | 1396 | 3.5 |
| Minimarked | 97 | 922 | 2.3 |
| Øst for storebælt (tidl. Davli, Centralkøb) | 70 | 797 | 2.0 |
| Diverse tidl. Centralkøb | 64 | 668 | 1.7 |
| Focus (tidl. Chr. Hansen) | 53 | 496 | 1.2 |
| Various stores with less than 50 purchases | 210 | 2317 | 5.8 |
| **Various discount stores** | | | |
| Rema 1000 | 494 | 6364 | 39.6 |
| Suma Discount | 219 | 3214 | 20.0 |
| Alta (Jaco discount), Super Alta | 246 | 2723 | 16.9 |
| ABC Lavpris (ej Jaco) | 170 | 2722 | 16.9 |
| Super Store. Navn ændret til Gobi Super Store 01/2001 | 72 | 1021 | 6.3 |
| Various stores with less than 50 purchases | 3 | 40 | 0.2 |
| **Greengrocers etc.** | | | |
| Greengrocer(Grønthandler) | 130 | 1766 | 67.9 |
| Cheese shop | 50 | 404 | 15.5 |
| Various stores with less than 50 purchases | 45 | 430 | 16.5 |
| **Directly from farms** | | | |
| Farm, Bought in the country | 2527 | 56628 | 92.9 |
| Market (torv) | 174 | 2739 | 4.5 |
| Various stores with less than 50 purchases | 97 | 1613 | 2.6 |
| **Bilka** | | | |
| Bilka | 857 | 10005 | 100.0 |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, possibly organic too.

## E.2 Producers and distribution

**Table E.2.1 Egg producers**

| **Producer** | **Number of sales** | **Number of eggs** | **Percentage of all eggs sold to the panel** |
|---|---|---|---|
| Æg Fra Friske Burhøns | 949 | 9492 | 3.29 |
| Danæg | 6683 | 76274 | 26.43 |
| Nemli | 249 | 2982 | 1.03 |
| Heslegård | 154 | 1803 | 0.62 |
| Hedegaard/Farmæg | 5446 | 56404 | 19.54 |
| Møllebjerggård | 277 | 2192 | 0.76 |
| Skov | 2 | 24 | 0.01 |
| Dueholm | 679 | 4620 | 1.60 |
| Brd. Honum | 1006 | 11107 | 3.85 |
| Økologisk balance æg | 116 | 858 | 0.30 |
| F.D.B. incl. e.g. Danæg, Natura | 4353 | 44990 | 15.59 |
| Iso Dagsfriske | 7 | 42 | 0.01 |
| Alta æg | 42 | 556 | 0.19 |
| Farmer/farmgate selling | 2699 | 59099 | 20.48 |
| Unknown producer | 1485 | 18192 | 6.30 |
| *Total* | *24147* | *288635* | *100.00* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, possibly organic too.

**Table E.2.2 Aggregated stores per egg producer**

| **Aggregated store** | **Number of eggs** | **Percentage of all eggs produced by this producer (and sold to the panel)** |
|---|---|---|
| **Æg Fra Friske Burhøns** | | |
| Føtex | 4959 | 52.24 |
| Bilka | 4303 | 45.33 |
| Other stores | 66 | 0.70 |
| Stores with less than 50 purchases | 164 | 1.73 |

| Aggregated store | Number of eggs | Percentage of all eggs produced by this producer (and sold to the panel) |
|---|---|---|
| **Danæg** | | |
| Various grocers | 25968 | 34.05 |
| Fakta (Discount) | 18506 | 24.26 |
| Aldi (Discount) | 11410 | 14.96 |
| Various discount stores | 6758 | 8.86 |
| Prima | 4823 | 6.32 |
| Favør | 4249 | 5.57 |
| Kvikly and OBS | 1245 | 1.63 |
| Superbrugsen | 747 | 0.98 |
| Irma | 694 | 0.91 |
| Dagligbrugsen | 627 | 0.82 |
| Corner store/petrol station | 313 | 0.41 |
| Other stores | 308 | 0.40 |
| Føtex | 242 | 0.32 |
| Netto (Discount) | 199 | 0.26 |
| Greengrocers etc. | 136 | 0.18 |
| Stores with less than 50 purchases | 49 | 0.06 |
| **Nemli** | | |
| Various grocers | 1652 | 55.40 |
| Favør | 1330 | 44.60 |
| **Heslegård** | | |
| Prima | 1102 | 61.12 |
| Directly from farms | 343 | 19.02 |
| Other stores | 177 | 9.82 |
| Various grocers | 123 | 6.82 |
| Stores with less than 50 purchases | 58 | 3.22 |

| Aggregated store | Number of eggs | Percentage of all eggs produced by this producer (and sold to the panel) |
|---|---|---|
| **Hedegaard/Farmæg** | | |
| Netto (Discount) | 40982 | 72.66 |
| Føtex | 8135 | 14.42 |
| Bilka | 4098 | 7.27 |
| Various grocers | 1944 | 3.45 |
| Superbrugsen | 198 | 0.35 |
| Kvikly and OBS | 152 | 0.27 |
| Dagligbrugsen | 133 | 0.24 |
| Favør | 127 | 0.23 |
| Various discount stores | 106 | 0.19 |
| Prima | 104 | 0.18 |
| Aldi (Discount) | 100 | 0.18 |
| Fakta (Discount) | 96 | 0.17 |
| Greengrocers etc. | 74 | 0.13 |
| Directly from farms | 66 | 0.12 |
| Stores with less than 50 purchases | 89 | 0.16 |
| **Møllebjerggård** | | |
| Various grocers [1] | 2089 | 95.30 |
| Stores with less than 50 purchases | 103 | 4.70 |
| **Skov** | | |
| Netto (Discount) | 18 | 75.00 |
| Various grocers | 6 | 25.00 |
| **Dueholm** | | |
| Føtex | 2160 | 46.75 |
| Netto (Discount) | 1230 | 26.62 |
| Bilka | 974 | 21.08 |
| Prima | 152 | 3.29 |
| Stores with less than 50 purchases | 104 | 2.25 |

[1] 94% of these eggs are sold in ISO and 5% in Super Best.

| Aggregated store | Number of eggs | Percentage of all eggs produced by this producer (and sold to the panel) |
|---|---|---|
| | **Brd. Honum** | |
| Various discount stores [2] | 6434 | 57.83 |
| Various grocers [3] | 3699 | 33.25 |
| Aldi (Discount) | 228 | 2.05 |
| Dagligbrugsen | 184 | 1.65 |
| Superbrugsen | 173 | 1.56 |
| Corner store/petrol station | 160 | 1.44 |
| Kvikly and OBS | 89 | 0.80 |
| Greengrocers etc. | 52 | 0.47 |
| Stores with less than 50 purchases | 88 | 0.79 |
| | **Økologisk balance æg** | |
| Irma | 846 | 98.60 |
| Stores with less than 50 purchases | 12 | 1.40 |
| | **F.D.B. incl. e.g. Danæg, Natura** | |
| Superbrugsen | 22980 | 51.08 |
| Kvikly and OBS | 15699 | 34.89 |
| Dagligbrugsen | 6175 | 13.73 |
| Irma | 96 | 0.21 |
| Stores with less than 50 purchases | 40 | 0.09 |
| | **Iso Dagsfriske** | |
| Various grocers | 42 | 100.00 |
| | **Alta æg** | |
| Various discount stores | 516 | 92.81 |
| Stores with less than 50 purchases | 40 | 7.19 |
| | **Farmer/farmgate selling** | |
| Directly from farms | 58124 | 98.35 |
| Greengrocers etc. | 975 | 1.65 |

[2] 50% of these are sold in Edeka Aktiv Super
[3] 93% of these are sold in Rema 1000

| Aggregated store | Number of eggs | Percentage of all eggs produced by this producer (and sold to the panel) |
|---|---|---|
| **Unknown producer** | | |
| Various grocers | 4658 | 25.60 |
| Directly from farms | 2397 | 13.18 |
| Various discount stores | 2248 | 12.36 |
| Netto (Discount) | 1320 | 7.26 |
| Greengrocers etc. | 1333 | 7.33 |
| Føtex | 1002 | 5.51 |
| Superbrugsen | 754 | 4.14 |
| Favør | 636 | 3.50 |
| Kvikly and OBS | 611 | 3.36 |
| Prima | 590 | 3.24 |
| Other stores | 582 | 3.20 |
| Bilka | 549 | 3.02 |
| Fakta (Discount) | 475 | 2.61 |
| Dagligbrugsen | 446 | 2.45 |
| Aldi (Discount) | 382 | 2.10 |
| Corner store/petrol station | 147 | 0.81 |
| Irma | 62 | 0.34 |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, possibly organic too.

**Table E.2.3 Producers per aggregated store**

| Aggregated store | Producer | Number of purchases | Percentage of all purchases in this store |
|---|---|---|---|
| Superbrugsen | F.D.B. incl. e.g. Danæg, Natura | 22,980 | 92 |
| | Other producers | 1,925 | 8 |
| Dagligbrugsen | F.D.B. incl. e.g. Danæg, Natura | 6,175 | 81 |
| | Other producers | 1,418 | 19 |
| Kvikly and OBS | F.D.B. incl. e.g. Danæg, Natura | 15,699 | 88 |
| | Other producers | 2,139 | 12 |
| Irma | Økologisk balance æg | 846 | 49 |
| | Danæg | 694 | 40 |
| | Other producers | 194 | 11 |
| Fakta (Discount) | Danæg | 18,506 | 97 |
| | Other producers | 619 | 3 |
| Føtex | Hedegaard/Farmæg | 8,135 | 49 |
| | Æg Fra Friske Burhøns | 4,959 | 30 |
| | Dueholm | 2,160 | 13 |
| | Other producers | 1,256 | 8 |
| Netto (Discount) | Hedegaard/Farmæg | 40,982 | 93 |
| | Other producers | 2,881 | 7 |
| Aldi (Discount) | Danæg | 11,410 | 94 |
| | Other producers | 728 | 6 |
| Prima | Danæg | 4,823 | 71 |
| | Heslegård | 1,102 | 16 |
| | Other producers | 886 | 13 |
| Favør | Danæg | 4,249 | 67 |
| | Nemli | 1,330 | 21 |
| | Other producers | 773 | 12 |
| Various grocers | Danæg | 25,968 | 65 |
| | Other producers | 14,277 | 35 |
| Various discount stores | Danæg | 6,758 | 42 |
| | Brd. Honum | 6,434 | 40 |
| | Other producers | 2,892 | 18 |
| Greengrocer etc. | Unknown producer | 1,333 | 51 |
| | Farmer/farmgate selling | 975 | 38 |
| | Other producers | 292 | 11 |
| Directly from farms | Farmer/farmgate selling | 58,124 | 95 |
| | Other producers | 2,856 | 5 |
| Bilka | Æg Fra Friske Burhøns | 4,303 | 43 |
| | Hedegaard/Farmæg | 4,098 | 41 |
| | Other producers | 1,604 | 16 |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, possibly organic too.

## E.3 Prices

### Table E.3.1 Egg types

| Egg type | Number of eggs purchased | Percentage of all eggs purchased by the panel | Number of purchases | Percentage of all purchases made by the panel |
|---|---|---|---|---|
| Battery eggs | 128425 | 44.5 | 9842 | 40.8 |
| Barn eggs | 44392 | 15.4 | 3917 | 16.2 |
| Free-range eggs | 57097 | 19.8 | 3875 | 16.0 |
| Organic eggs | 55793 | 19.3 | 6369 | 26.4 |
| Free-range eggs, possibly organic too | 2928 | 1.0 | 144 | 0.6 |
| *Total* | *288635* | *100* | *24147* | *100* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, possibly organic too.

### Table E.3.2 Prices of different egg types

| Egg type | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price | Standard deviation divided by mean |
|---|---|---|---|---|---|---|---|
| Battery eggs | 9842 | 128425 | 0.25 | 2.99 | 1.25 | 0.279 | 0.223 |
| Barn eggs | 3917 | 44392 | 0.28 | 3.33 | 1.56 | 0.414 | 0.265 |
| Free-range eggs | 3875 | 57097 | 0.27 | 3.16 | 1.50 | 0.545 | 0.363 |
| Organic eggs | 6369 | 55793 | 0.33 | 4.33 | 1.96 | 0.460 | 0.235 |
| Free-range eggs, possibly organic too | 144 | 2928 | 0.50 | 2.79 | 1.21 | 0.467 | 0.385 |
| *Total* | *24147* | *288635* | *0.27* | *4.33* | *1.53* | *0.497* | *0.325* |

Source: GfK purchase data on eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Including data on eggs that are free-range, possibly organic too. Prices are in DKK per egg.

### Table E.3.3 Details about price of battery eggs for each producer

| Producer | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price | Standard deviation divided by mean |
|---|---|---|---|---|---|---|---|
| Æg Fra Friske Burhøns | 466 | 5137 | 0.72 | 2.53 | 1.44 | 0.200 | 0.139 |
| Danæg | 3680 | 51789 | 0.25 | 2.43 | 1.19 | 0.284 | 0.238 |
| Nemli | 117 | 1501 | 0.33 | 1.58 | 1.18 | 0.215 | 0.182 |
| Heslegård | 105 | 1262 | 0.58 | 2.99 | 1.36 | 0.465 | 0.342 |
| Hedegaard/Farmæg | 3397 | 39838 | 0.46 | 2.83 | 1.25 | 0.246 | 0.196 |
| Møllebjerggård | 39 | 264 | 1.00 | 2.33 | 2.05 | 0.296 | 0.145 |
| Skov | 1 | 18 | 1.83 | 1.83 | 1.83 | | |
| Brd. Honum | 662 | 8591 | 0.50 | 2.33 | 1.23 | 0.294 | 0.239 |
| F.D.B. incl. e.g. Danæg, Natura | 980 | 13710 | 0.33 | 2.33 | 1.36 | 0.195 | 0.144 |
| Alta æg | 21 | 310 | 1.06 | 1.40 | 1.14 | 0.060 | 0.053 |
| Farmer/farmgate selling | 50 | 1377 | 0.40 | 1.33 | 0.98 | 0.146 | 0.149 |
| Unknown producer | 324 | 4628 | 0.39 | 2.58 | 1.22 | 0.376 | 0.308 |
| *Total* | *9842* | *128425* | *0.25* | *2.99* | *1.25* | *0.279* | *0.223* |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.4 Details about prices of barn eggs for each producer**

| Producer | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price | Standard deviation divided by mean |
|---|---|---|---|---|---|---|---|
| Æg Fra Friske Burhøns | 249 | 2282 | 0.83 | 2.56 | 2.00 | 0.312 | 0.155 |
| Danæg | 1046 | 10105 | 0.50 | 2.83 | 1.56 | 0.420 | 0.270 |
| Nemli | 130 | 1456 | 0.60 | 2.00 | 1.50 | 0.274 | 0.183 |
| Heslegård | 12 | 143 | 0.87 | 2.20 | 1.56 | 0.401 | 0.257 |
| Hedegaard/Farmæg | 211 | 2446 | 0.83 | 2.83 | 1.45 | 0.299 | 0.206 |
| Møllebjerggård | 39 | 422 | 0.90 | 2.33 | 1.65 | 0.365 | 0.221 |
| Dueholm | 2 | 16 | 1.35 | 2.25 | 1.80 | 0.640 | 0.356 |
| Brd. Honum | 92 | 724 | 0.83 | 2.74 | 1.67 | 0.407 | 0.244 |
| F.D.B. incl. e.g. Danæg, Natura | 1297 | 14214 | 0.60 | 2.66 | 1.65 | 0.326 | 0.198 |
| Iso Dagsfriske | 7 | 42 | 1.66 | 3.33 | 2.28 | 0.507 | 0.222 |
| Alta æg | 18 | 211 | 1.13 | 1.70 | 1.44 | 0.211 | 0.147 |
| Farmer/farmgate selling | 279 | 6026 | 0.33 | 2.00 | 0.99 | 0.268 | 0.270 |
| Unknown producer | 535 | 6305 | 0.28 | 3.33 | 1.48 | 0.412 | 0.277 |
| *Total* | *3917* | *44392* | *0.28* | *3.33* | *1.56* | *0.414* | *0.265* |

Source: GfK purchase data on barn eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.5 Details about prices of free-range eggs for each producer**

| Producer | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price | Standard deviation divided by mean |
|---|---|---|---|---|---|---|---|
| Æg Fra Friske Burhøns | 211 | 1881 | 1.08 | 2.97 | 2.18 | 0.394 | 0.181 |
| Danæg | 596 | 5330 | 0.50 | 2.67 | 1.92 | 0.365 | 0.190 |
| Nemli | 2 | 25 | 0.86 | 1.40 | 1.13 | 0.376 | 0.333 |
| Heslegård | 27 | 322 | 0.83 | 2.83 | 1.80 | 0.619 | 0.345 |
| Hedegaard/Farmæg | 431 | 3828 | 0.67 | 3.16 | 1.67 | 0.254 | 0.152 |
| Møllebjerggård | 20 | 208 | 1.00 | 2.16 | 1.78 | 0.208 | 0.117 |
| Dueholm | 1 | 18 | 1.83 | 1.83 | 1.83 | | |
| Brd. Honum | 57 | 472 | 0.91 | 2.49 | 1.66 | 0.385 | 0.233 |
| Økologisk balance æg | 1 | 10 | 1.80 | 1.80 | 1.80 | | |
| F.D.B. incl. e.g. Danæg, Natura | 530 | 4896 | 1.00 | 2.83 | 2.07 | 0.308 | 0.148 |
| Alta æg | 2 | 25 | 1.13 | 1.85 | 1.49 | 0.509 | 0.342 |
| Farmer/farmgate selling | 1632 | 35553 | 0.33 | 2.50 | 1.02 | 0.216 | 0.211 |
| Unknown producer | 365 | 4529 | 0.27 | 2.83 | 1.51 | 0.521 | 0.346 |

Source: GfK purchase data on free-range eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.6 Details about prices of organic eggs for each producer**

| Producer | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price | Standard deviation divided by mean |
|---|---|---|---|---|---|---|---|
| Æg Fra Friske Burhøns | 14 | 120 | 1.25 | 2.76 | 2.13 | 0.518 | 0.243 |
| Danæg | 1361 | 9050 | 0.83 | 3.33 | 2.03 | 0.365 | 0.180 |
| Heslegård | 10 | 76 | 1.11 | 2.66 | 2.12 | 0.469 | 0.221 |
| Hedegaard/Farmæg | 1406 | 10282 | 1.16 | 4.33 | 1.82 | 0.211 | 0.116 |
| Møllebjerggård | 179 | 1298 | 1.10 | 3.33 | 2.09 | 0.459 | 0.220 |
| Skov | 1 | 6 | 2.42 | 2.42 | 2.42 | | |
| Dueholm | 676 | 4586 | 0.41 | 3.15 | 2.07 | 0.333 | 0.161 |
| Brd. Honum | 194 | 1314 | 1.30 | 3.33 | 2.05 | 0.431 | 0.210 |
| Økologisk balance æg | 115 | 848 | 1.38 | 2.75 | 2.41 | 0.338 | 0.141 |
| F.D.B. incl. e.g. Danæg, Natura | 1546 | 12170 | 0.80 | 3.66 | 2.27 | 0.297 | 0.131 |
| Alta æg | 1 | 10 | 1.40 | 1.40 | 1.40 | | |
| Farmer/farmgate selling | 623 | 13503 | 0.33 | 3.00 | 1.10 | 0.282 | 0.258 |
| Unknown producer | 243 | 2530 | 0.67 | 3.33 | 1.98 | 0.534 | 0.270 |

Source: GfK purchase data on organic eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.7 Mean prices of all eggs per producer**

| Producer | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|
| Æg Fra Friske Burhøns | 1.44 | 2.00 | 2.18 | 2.13 |
| Danæg | 1.19 | 1.56 | 1.92 | 2.03 |
| Nemli | 1.18 | **1.50** | **1.13** | . |
| Heslegård | 1.36 | 1.56 | 1.80 | 2.12 |
| Hedegaard/Farmæg | 1.25 | 1.45 | 1.67 | 1.82 |
| Møllebjerggård | 2.05 | 1.65 | 1.78 | 2.09 |
| Skov | 1.83 | . | . | 2.42 |
| Dueholm | . | 1.80 | 1.83 | 2.07 |
| Brd. Honum | 1.23 | **1.67** | **1.66** | 2.05 |
| Økologisk balance æg | . | . | 1.80 | 2.41 |
| F.D.B. incl. e.g. Danæg, Natura | 1.36 | 1.65 | 2.07 | 2.27 |
| Iso Dagsfriske | . | 2.28 | . | . |
| Alta æg | 1.14 | 1.44 | 1.49 | 1.40 |
| Farmer/farmgate selling | 0.98 | 0.99 | 1.02 | 1.10 |
| Unknown producer | 1.22 | 1.48 | 1.51 | 1.98 |
| *Total* | *1.25* | *1.56* | *1.50* | *1.96* |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.8 Details about prices of battery eggs per aggregated store**

| **Aggregated store** | **Number of purchases** | **Number of eggs purchased** | **Minimum price** | **Maximum price** | **Mean price** | **Standard deviation of price** | **Standard deviation divided by mean** |
|---|---|---|---|---|---|---|---|
| Superbrugsen | 484 | 6800 | 0.66 | 2.33 | 1.36 | 0.191 | 0.141 |
| Dagligbrugsen | 282 | 4473 | 0.66 | 2.58 | 1.25 | 0.216 | 0.173 |
| Kvikly and OBS | 362 | 4610 | 0.33 | 2.33 | 1.38 | 0.247 | 0.18 |
| Irma | 1 | 12 | 1.06 | 1.06 | 1.06 | | |
| Fakta (Discount) | 841 | 11874 | 0.9 | 2 | 1.17 | 0.164 | 0.141 |
| Føtex | 487 | 6345 | 0.5 | 2.53 | 1.26 | 0.327 | 0.26 |
| Netto (Discount) | 2928 | 31985 | 0.67 | 2.6 | 1.29 | 0.221 | 0.172 |
| Aldi (Discount) | 677 | 10267 | 0.53 | 1.99 | 1.02 | 0.073 | 0.071 |
| Prima | 348 | 4192 | 0.25 | 2.99 | 1.29 | 0.37 | 0.288 |
| Favør | 311 | 4384 | 0.25 | 2.33 | 1.06 | 0.317 | 0.299 |
| Various grocers | 1920 | 24316 | 0.42 | 2.83 | 1.3 | 0.349 | 0.268 |
| Various discount stores | 699 | 11127 | 0.5 | 2 | 1.19 | 0.258 | 0.217 |
| Greengrocers etc. | 15 | 178 | 1 | 2 | 1.45 | 0.364 | 0.251 |
| Directly from farms | 71 | 1777 | 0.4 | 1.8 | 1.06 | 0.232 | 0.219 |
| Corner store/ petrol station | 39 | 313 | 1 | 2.33 | 1.68 | 0.363 | 0.216 |
| Other stores | 24 | 367 | 0.39 | 1.99 | 1.1 | 0.496 | 0.452 |
| Bilka | 353 | 5405 | 0.72 | 2.3 | 1.21 | 0.297 | 0.246 |
| *Total* | *9842* | *128425* | *0.25* | *2.99* | *1.25* | *0.279* | *0.223* |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.9 Details about prices of barn eggs per aggregated store**

| **Aggregated store** | **Number of purchases** | **Number of eggs purchased** | **Minimum price** | **Maximum price** | **Mean price** | **Standard deviation of price** | **Standard deviation divided by mean** |
|---|---|---|---|---|---|---|---|
| Superbrugsen | 767 | 8534 | 0.6 | 3.33 | 1.62 | 0.319 | 0.197 |
| Dagligbrugsen | 139 | 1430 | 0.63 | 2.66 | 1.84 | 0.35 | 0.19 |
| Kvikly and OBS | 572 | 6343 | 0.8 | 2.49 | 1.68 | 0.309 | 0.184 |
| Irma | 55 | 478 | 1 | 2.66 | 2.04 | 0.497 | 0.243 |
| Fakta (Discount) | 326 | 2773 | 1 | 2.58 | 1.31 | 0.121 | 0.092 |
| Føtex | 318 | 3264 | 0.83 | 2.56 | 1.8 | 0.423 | 0.235 |
| Netto (Discount) | 101 | 1075 | 0.67 | 2.8 | 1.42 | 0.264 | 0.186 |
| Aldi (Discount) | 6 | 76 | 0.8 | 2.83 | 1.34 | 0.745 | 0.555 |
| Prima | 100 | 968 | 0.28 | 2.49 | 1.77 | 0.575 | 0.324 |
| Favør | 92 | 1027 | 0.6 | 2.49 | 1.65 | 0.41 | 0.248 |
| Various grocers | 742 | 7706 | 0.6 | 3.33 | 1.56 | 0.423 | 0.272 |
| Various discount stores | 217 | 2381 | 0.5 | 1.99 | 1.37 | 0.269 | 0.196 |
| Greengrocers etc. | 30 | 269 | 0.9 | 2.5 | 1.78 | 0.468 | 0.263 |
| Directly from farms | 310 | 6647 | 0.33 | 2.22 | 1.01 | 0.269 | 0.266 |
| Corner store/petrol station | 12 | 88 | 1 | 2.74 | 2.13 | 0.584 | 0.274 |
| Other stores | 9 | 77 | 1.31 | 2.49 | 1.81 | 0.464 | 0.256 |
| Bilka | 121 | 1256 | 0.94 | 2.51 | 1.72 | 0.354 | 0.206 |
| *Total* | *3917* | *44392* | *0.28* | *3.33* | *1.56* | *0.414* | *0.265* |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.10 Details about prices of free-range eggs per aggregated store**

| Aggregated store | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price | Standard deviation divided by mean |
|---|---|---|---|---|---|---|---|
| Superbrugsen | 277 | 2768 | 1 | 2.83 | 1.97 | 0.31 | 0.158 |
| Dagligbrugsen | 91 | 834 | 1 | 2.66 | 2.26 | 0.378 | 0.167 |
| Kvikly and OBS | 202 | 1751 | 1 | 2.49 | 2.07 | 0.291 | 0.141 |
| Irma | 50 | 384 | 1 | 2.66 | 2.32 | 0.4 | 0.173 |
| Fakta (Discount) | 196 | 1284 | 1.06 | 2.67 | 1.79 | 0.157 | 0.087 |
| Føtex | 290 | 2969 | 0.83 | 2.97 | 1.91 | 0.498 | 0.261 |
| Netto (Discount) | 276 | 1861 | 0.67 | 2.3 | 1.78 | 0.147 | 0.083 |
| Aldi (Discount) | 12 | 151 | 1 | 1.83 | 1.45 | 0.397 | 0.273 |
| Prima | 58 | 568 | 1.27 | 2.83 | 2.19 | 0.354 | 0.162 |
| Favør | 38 | 387 | 0.8 | 2.4 | 1.75 | 0.442 | 0.253 |
| Various grocers | 359 | 3745 | 0.5 | 3.16 | 1.77 | 0.46 | 0.26 |
| Various discount stores | 110 | 1246 | 0.8 | 2.53 | 1.77 | 0.463 | 0.261 |
| Greengrocers etc. | 123 | 1591 | 0.9 | 2.5 | 1.48 | 0.572 | 0.387 |
| Directly from farms | 1655 | 35941 | 0.33 | 2.33 | 1.02 | 0.194 | 0.19 |
| Corner store /petrol station | 6 | 69 | 0.93 | 2.66 | 1.65 | 0.638 | 0.387 |
| Other stores | 23 | 420 | 0.27 | 2.33 | 1.21 | 0.377 | 0.31 |
| Bilka | 109 | 1128 | 1.08 | 2.82 | 1.86 | 0.456 | 0.245 |
| *Total* | *3875* | *57097* | *0.27* | *3.16* | *1.50* | *0.545* | *0.363* |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.11 Details about prices of organic eggs per aggregated store**

| Aggregated store | Number of purchases | Number of eggs purchased | Minimum price | Maximum price | Mean price | Standard deviation of price | Standard deviation divided by mean |
|---|---|---|---|---|---|---|---|
| Superbrugsen | 851 | 6803 | 0.99 | 2.83 | 2.24 | 0.266 | 0.119 |
| Dagligbrugsen | 124 | 856 | 0.99 | 2.99 | 2.52 | 0.371 | 0.147 |
| Kvikly and OBS | 651 | 5134 | 0.8 | 3.66 | 2.25 | 0.31 | 0.138 |
| Irma | 117 | 860 | 1.38 | 2.75 | 2.39 | 0.344 | 0.144 |
| Fakta (Discount) | 512 | 3194 | 1 | 2.83 | 1.79 | 0.137 | 0.076 |
| Føtex | 489 | 3866 | 0.41 | 3.15 | 2.17 | 0.331 | 0.152 |
| Netto (Discount) | 1331 | 8942 | 0.67 | 4.33 | 1.8 | 0.176 | 0.098 |
| Aldi (Discount) | 262 | 1644 | 1.42 | 2.66 | 1.87 | 0.102 | 0.054 |
| Prima | 140 | 1083 | 1 | 2.99 | 2.31 | 0.399 | 0.173 |
| Favør | 74 | 554 | 0.87 | 2.99 | 2.27 | 0.421 | 0.186 |
| Various grocers | 610 | 4472 | 0.83 | 3.33 | 2.19 | 0.445 | 0.203 |
| Various discount stores | 175 | 1284 | 1 | 2.66 | 1.97 | 0.302 | 0.153 |
| Greengrocers etc. | 48 | 475 | 0.83 | 3.33 | 2.15 | 0.541 | 0.252 |
| Directly from farms | 644 | 13938 | 0.33 | 3 | 1.1 | 0.287 | 0.26 |
| Corner store/ petrol station | 29 | 198 | 1.17 | 3.33 | 2.76 | 0.59 | 0.214 |
| Other stores | 43 | 310 | 0.8 | 3.33 | 2.3 | 0.377 | 0.164 |
| Bilka | 269 | 2180 | 1.1 | 2.66 | 1.88 | 0.284 | 0.151 |
| *Total* | *6369* | *55793* | *0.33* | *4.33* | *1.96* | *0.460* | *0.235* |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.12 Mean price of all eggs per aggregated store**

| Aggregated store | Battery eggs | Barn eggs | Free-range eggs | Organic eggs |
|---|---|---|---|---|
| Superbrugsen | 1.36 | 1.62 | 1.97 | 2.24 |
| Dagligbrugsen | 1.25 | 1.84 | 2.26 | 2.52 |
| Kvikly and OBS | 1.38 | 1.68 | 2.07 | 2.25 |
| Irma | 1.06 | 2.04 | 2.32 | 2.39 |
| Fakta (Discount) | 1.17 | 1.31 | 1.79 | 1.79 |
| Føtex | 1.26 | 1.80 | 1.91 | 2.17 |
| Netto (Discount) | 1.29 | 1.42 | 1.78 | 1.80 |
| Aldi (Discount) | 1.02 | 1.34 | 1.45 | 1.87 |
| Prima | 1.29 | 1.77 | 2.19 | 2.31 |
| Favør | 1.06 | 1.65 | 1.75 | 2.27 |
| Various grocers | 1.30 | 1.56 | 1.77 | 2.19 |
| Various discount stores | 1.19 | 1.37 | 1.77 | 1.97 |
| Greengrocers etc. | 1.45 | **1.78** | **1.48** | 2.15 |
| Directly from farms | 1.06 | 1.01 | 1.02 | 1.10 |
| Corner store/petrol station | 1.68 | **2.13** | **1.65** | 2.76 |
| Other stores | 1.10 | **1.81** | **1.21** | 2.30 |
| Bilka | 1.21 | 1.72 | 1.86 | 1.88 |
| *Total* | *1.25* | *1.56* | *1.50* | *1.96* |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z. Prices are in DKK per egg.

**Table E.3.13 Ordering of imputed prices in different stores**

| Aggregated store | Order of prices[4] | Count | Percent |
|---|---|---|---|
| Superbrugsen | P(battery), P(barn), P(free-range), P(organic) | 1853 | 78.3 |
| | P(battery), P(barn), P(organic), P(free-range) | 218 | 9.2 |
| | P(battery), P(free-range), P(barn), P(organic) | 96 | 4.1 |
| | P(barn), P(battery), P(free-range), P(organic) | 160 | 6.8 |
| | P(free-range), P(battery), P(barn), P(organic) | 40 | 1.7 |
| Dagligbrugsen | P(battery), P(barn), P(free-range), P(organic) | 423 | 66.7 |
| | P(battery), P(barn), P(organic), P(free-range) | 115 | 18.1 |
| | P(battery), P(free-range), P(barn), P(organic) | 48 | 7.6 |
| | P(battery), P(organic), P(barn), P(free-range) | 11 | 1.7 |
| | P(battery), P(organic), P(free-range), P(barn) | 10 | 1.6 |
| | P(free-range), P(battery), P(barn), P(organic) | 27 | 4.3 |
| Kvickly and OBS | P(battery), P(barn), P(free-range), P(organic) | 1293 | 72.8 |
| | P(battery), P(barn), P(organic), P(free-range) | 403 | 22.7 |
| | P(battery), P(free-range), P(barn), P(organic) | 22 | 1.2 |
| | P(barn), P(battery), P(free-range), P(organic) | 57 | 3.2 |
| Irma | Two or more prices are equal | 29 | |
| | P(battery), P(barn), P(free-range), P(organic) | 55 | 28.6 |
| | P(battery), P(barn), P(organic), P(free-range) | 71 | 37.0 |
| | P(battery), P(free-range), P(barn), P(organic) | 24 | 12.5 |
| | P(battery), P(free-range), P(organic), P(barn) | 29 | 15.1 |
| | P(battery), P(organic), P(barn), P(free-range) | 8 | 4.2 |
| | P(battery), P(organic), P(free-range), P(barn) | 5 | 2.6 |
| Fakta (Discount) | P(battery), P(barn), P(free-range), P(organic) | 926 | 49.4 |
| | P(battery), P(barn), P(organic), P(free-range) | 948 | 50.6 |

[4] Price 1 is the price of battery eggs, price 2 is the price of barn eggs, price 3 is the price of free-range eggs and price 4 is the price of organic eggs.

| **Aggregated store** | **Order of prices**[4] | **Count** | **Percent** |
|---|---|---|---|
| Føtex | P(battery), P(barn), P(free-range), P(organic) | 455 | 28.7 |
| | P(battery), P(barn), P(organic), P(free-range) | 363 | 22.9 |
| | P(battery), P(free-range), P(barn), P(organic) | 404 | 25.5 |
| | P(battery), P(free-range), P(organic), P(barn) | 78 | 4.9 |
| | P(battery), P(organic), P(barn), P(free-range) | 122 | 7.7 |
| | P(battery), P(organic), P(free-range), P(barn) | 33 | 2.1 |
| | P(barn), P(battery), P(free-range), P(organic) | 91 | 5.7 |
| | P(barn), P(battery), P(organic), P(free-range) | 37 | 2.3 |
| Netto (Discount) | Two or more prices are equal | 85 | |
| | P(battery), P(barn), P(free-range), P(organic) | 1814 | 40.0 |
| | P(battery), P(barn), P(organic), P(free-range) | 1429 | 31.5 |
| | P(battery), P(organic), P(barn), P(free-range) | 148 | 3.3 |
| | P(battery), P(organic), P(free-range), P(barn) | 86 | 1.9 |
| | P(barn), P(battery), P(free-range), P(organic) | 530 | 11.7 |
| | P(barn), P(battery), P(organic), P(free-range) | 533 | 11.7 |
| Aldi (Discount) | Two or more prices are equal | 67 | |
| | P(battery), P(barn), P(free-range), P(organic) | 716 | 80.8 |
| | P(battery), P(free-range), P(barn), P(organic) | 62 | 7.0 |
| | P(battery), P(free-range), P(organic), P(barn) | 22 | 2.5 |
| | P(barn), P(battery), P(free-range), P(organic) | 19 | 2.1 |
| | P(barn), P(battery), P(organic), P(free-range) | 27 | 3.0 |
| | P(free-range), P(battery), P(barn), P(organic) | 40 | 4.5 |
| Prima | Two or more prices are equal | 22 | |
| | P(battery), P(barn), P(free-range), P(organic) | 340 | 54.7 |
| | P(battery), P(barn), P(organic), P(free-range) | 89 | 14.3 |
| | P(battery), P(free-range), P(barn), P(organic) | 108 | 17.4 |
| | P(battery), P(free-range), P(organic), P(barn) | 9 | 1.4 |
| | P(battery), P(organic), P(barn), P(free-range) | 22 | 3.5 |
| | P(battery), P(organic), P(free-range), P(barn) | 19 | 3.1 |
| | P(barn), P(battery), P(free-range), P(organic) | 14 | 2.3 |
| | P(barn), P(battery), P(organic), P(free-range) | 10 | 1.6 |
| | P(organic), P(barn), P(battery), P(free-range) | 11 | 1.8 |
| Favør | P(battery), P(barn), P(free-range), P(organic) | 254 | 49.6 |
| | P(battery), P(barn), P(organic), P(free-range) | 13 | 2.5 |
| | P(battery), P(free-range), P(barn), P(organic) | 107 | 20.9 |
| | P(battery), P(organic), P(barn), P(free-range) | 10 | 2.0 |
| | P(battery), P(organic), P(free-range), P(barn) | 20 | 3.9 |
| | P(barn), P(battery), P(free-range), P(organic) | 46 | 9.0 |
| | P(free-range), P(battery), P(barn), P(organic) | 62 | 12.1 |
| Various grocers (eg Spar, Superbest, Iso osv) | P(battery), P(barn), P(free-range), P(organic) | 2204 | 61.2 |
| | P(battery), P(barn), P(organic), P(free-range) | 392 | 10.9 |
| | P(battery), P(free-range), P(barn), P(organic) | 939 | 26.1 |
| | P(barn), P(battery), P(free-range), P(organic) | 65 | 1.8 |

| Aggregated store | Order of prices[4] | Count | Percent |
|---|---|---|---|
| Various discount stores (eg Suma, Rema, Coma og ABC) | P(battery), P(barn), P(free-range), P(organic) | 671 | 55.9 |
| | P(battery), P(barn), P(organic), P(free-range) | 356 | 29.7 |
| | P(battery), P(free-range), P(barn), P(organic) | 44 | 3.7 |
| | P(barn), P(battery), P(free-range), P(organic) | 60 | 5.0 |
| | P(barn), P(battery), P(organic), P(free-range) | 24 | 2.0 |
| | P(free-range), P(battery), P(barn), P(organic) | 45 | 3.8 |
| Greengrocers etc. | Two or more prices are equal | 4 | |
| | P(battery), P(barn), P(free-range), P(organic) | 21 | 10.0 |
| | P(battery), P(free-range), P(barn), P(organic) | 75 | 35.9 |
| | P(battery), P(free-range), P(organic), P(barn) | 6 | 2.9 |
| | P(barn), P(battery), P(free-range), P(organic) | 7 | 3.3 |
| | P(free-range), P(battery), P(barn), P(organic) | 59 | 28.2 |
| | P(free-range), P(battery), P(organic), P(barn) | 12 | 5.7 |
| | P(free-range), P(barn), P(battery), P(organic) | 13 | 6.2 |
| | P(free-range), P(barn), P(organic), P(battery) | 4 | 1.9 |
| | P(organic), P(battery), P(barn), P(free-range) | 8 | 3.8 |
| | P(organic), P(free-range), P(battery), P(barn) | 4 | 1.9 |
| Directly from farms | Two or more prices are equal | 49 | |
| | P(battery), P(barn), P(free-range), P(organic) | 209 | 8.0 |
| | P(battery), P(barn), P(organic), P(free-range) | 58 | 2.2 |
| | P(battery), P(free-range), P(barn), P(organic) | 146 | 5.6 |
| | P(battery), P(free-range), P(organic), P(barn) | 219 | 8.4 |
| | P(battery), P(organic), P(barn), P(free-range) | 56 | 2.1 |
| | P(battery), P(organic), P(free-range), P(barn) | 57 | 2.2 |
| | P(barn), P(battery), P(free-range), P(organic) | 301 | 11.5 |
| | P(barn), P(free-range), P(battery), P(organic) | 232 | 8.9 |
| | P(barn), P(free-range), P(organic), P(battery) | 431 | 16.4 |
| | P(barn), P(organic), P(battery), P(free-range) | 99 | 3.8 |
| | P(barn), P(organic), P(free-range), P(battery) | 101 | 3.9 |
| | P(free-range), P(battery), P(barn), P(organic) | 121 | 4.6 |
| | P(free-range), P(barn), P(battery), P(organic) | 345 | 13.2 |
| | P(free-range), P(barn), P(organic), P(battery) | 29 | 1.1 |
| | P(free-range), P(organic), P(battery), P(barn) | 95 | 3.6 |
| | P(free-range), P(organic), P(barn), P(battery) | 122 | 4.7 |
| Bilka | Two or more prices are equal | 17 | |
| | P(battery), P(barn), P(free-range), P(organic) | 203 | 24.4 |
| | P(battery), P(barn), P(organic), P(free-range) | 256 | 30.8 |
| | P(battery), P(free-range), P(barn), P(organic) | 36 | 4.3 |
| | P(battery), P(free-range), P(organic), P(barn) | 90 | 10.8 |
| | P(battery), P(organic), P(barn), P(free-range) | 157 | 18.9 |
| | P(battery), P(organic), P(free-range), P(barn) | 43 | 5.2 |
| | P(barn), P(battery), P(free-range), P(organic) | 17 | 2.0 |
| | P(barn), P(free-range), P(battery), P(organic) | 22 | 2.6 |
| | P(organic), P(battery), P(barn), P(free-range) | 8 | 1.0 |

Calculations based on GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z.

## E.4 Tray size

**Table E.4.1 Distribution of tray size for different egg types**

| Egg type | Number of eggs in the egg tray, defined by GfK | Number of egg trays sold to the panel | PERCENT |
|---|---|---|---|
| Battery eggs | 1-5 eggs | 2 | 0.0 |
| | 6-9 eggs | 1493 | 15.2 |
| | 10-14 eggs | 4554 | 46.3 |
| | 15-19 eggs | 3461 | 35.2 |
| | 20-29 eggs | 235 | 2.4 |
| | 30-39 eggs | 93 | 0.9 |
| | 40 + eggs | 3 | 0.0 |
| | Nbr eggs not stated | 1 | 0.0 |
| Barn eggs | 6-9 eggs | 961 | 24.5 |
| | 10-14 eggs | 2393 | 61.1 |
| | 15-19 eggs | 411 | 10.5 |
| | 20-29 eggs | 34 | 0.9 |
| | 30-39 eggs | 117 | 3.0 |
| | Nbr eggs not stated | 1 | 0.0 |
| Free-range eggs | 1-5 eggs | 8 | 0.2 |
| | 6-9 eggs | 1086 | 28.0 |
| | 10-14 eggs | 1354 | 34.9 |
| | 15-19 eggs | 803 | 20.7 |
| | 20-29 eggs | 73 | 1.9 |
| | 30-39 eggs | 534 | 13.8[5] |
| | 40 + eggs | 16 | 0.4 |
| | Nbr eggs not stated | 1 | 0.0 |
| Organic eggs | 1-5 eggs | 5 | 0.1 |
| | 6-9 eggs | 4643 | 72.9 |
| | 10-14 eggs | 1240 | 19.5 |
| | 15-19 eggs | 212 | 3.3 |
| | 20-29 eggs | 34 | 0.5 |
| | 30-39 eggs | 227 | 3.6 |
| | 40 + eggs | 7 | 0.1 |
| | Nbr eggs not stated | 1 | 0.0 |

Source: GfK purchase data on battery eggs from 26 June 1999 to 30 June 2000. All stores except non-food, Canteens and A-Z.

[5] 96% of these eggs are purchased directly from farms, market places, vans, market garden. This might help explain the low price of free-range eggs.

# Appendix F Definition of geographical categories



## F.1 By geographical category

**Table F.1.1 Capital**

| Code | Municipality | Code | Municipality | Code | Municipality |
|---|---|---|---|---|---|
| 101 | Københavns Kommune | 161 | Glostrup | 175 | Rødovre |
| 147 | Frederiksberg Kommune | 163 | Herlev | 181 | Søllerød |
| 151 | Ballerup | 165 | Albertslund | 183 | Ishøj |
| 153 | Brøndby | 167 | Hvidovre | 185 | Tårnby |
| 155 | Dragør | 169 | Høje Taastrup | 187 | Vallensbæk |
| 157 | Gentofte | 171 | Ledøje-Smørum | 189 | Værløse |
| 159 | Gladsaxe | 173 | Lyngby-Taarbæk | | |

Source: GfK background data 1999

**Table F.1.2 Island *city*-municipality**

| Code | Municipality | Code | Municipality | Code | Municipality |
|---|---|---|---|---|---|
| 201 | Allerød | 235 | Stenløse | 367 | Nakskov |
| 205 | Birkerød | 253 | Greve | 369 | Nykøbing F. |
| 207 | Farum | 259 | Køge | 373 | Næstved |
| 208 | Fredensborg-Humlebæk | 265 | Roskilde | 407 | Rønne |
| 209 | Frederikssund | 269 | Solrød | 445 | Middelfart |
| 211 | Frederiksværk | 315 | Holbæk | 449 | Nyborg |
| 217 | Helsingør | 323 | Kalundborg | 461 | Odense |
| 219 | Hillerød | 325 | Korsør | 479 | Svendborg |
| 223 | Hørsholm | 329 | Ringsted | | |
| 227 | Karlebo | 333 | Slagelse | | |

Source: GfK background data 1999

**Table F.1.3 Other Island municipality**

| Code | Municipality | Code | Municipality | Code | Municipality |
|---|---|---|---|---|---|
| 213 | Græsted-Gilleleje | 339 | Svinninge | 421 | Assens |
| 215 | Helsinge | 341 | Tornved | 423 | Bogense |
| 221 | Hundested | 343 | Trundholm | 425 | Broby |
| 225 | Jægerspris | 345 | Tølløse | 427 | Egebjerg |
| 229 | Skibby | 351 | Fakse | 429 | Ejby |
| 231 | Skævinge | 353 | Fladså | 431 | Faaborg |
| 233 | Slangerup | 355 | Holeby | 433 | Glamsbjerg |
| 237 | Ølstykke | 357 | Holmegaard | 435 | Gudme |
| 251 | Bramsnæs | 359 | Højreby | 437 | Haarby |
| 255 | Gundsø | 361 | Langebæk | 439 | Kerteminde |
| 257 | Hvalsø | 363 | Maribo | 441 | Langeskov |
| 261 | Lejre | 365 | Møn | 443 | Marstal |
| 263 | Ramsø | 371 | Nysted | 447 | Munkebo |
| 267 | Skovbo | 375 | Nr. Alslev | 451 | Nørre Aaby |
| 271 | Vallø | 377 | Præstø | 471 | Otterup |
| 301 | Bjergsted | 379 | Ravnsborg | 473 | Ringe |
| 303 | Dianalund | 381 | Rudbjerg | 475 | Rudkøbing |
| 305 | Dragsholm | 383 | Rødby | 477 | Ryslinge |
| 307 | Fuglebjerg | 385 | Rønnede | 481 | Sydlangeland |
| 309 | Gørlev | 387 | Sakskøbing | 483 | Søndersø |
| 311 | Hashøj | 389 | Stevns | 485 | Tommerup |
| 313 | Haslev | 391 | Stubbekøbing | 487 | Tranekær |
| 317 | Hvidebæk | 393 | Suså | 489 | Ullerslev |
| 319 | Høng | 395 | Sydfalster | 491 | Vissenbjerg |
| 321 | Jernløse | 397 | Vordingborg | 495 | Ørbæk |
| 327 | Nykøbing-Rørvig | 401 | Allinge-Gudhjem | 497 | Årslev |
| 331 | Skælskør | 403 | Hasle | 499 | Aarup |
| 335 | Sorø | 405 | Nexø | | |
| 337 | Stenlille | 409 | Aakirkeby | | |

Source: GfK background data 1999

**Table F.1.4 Jutland *city* municipality**

| Code | Municipality | Code | Municipality | Code | Municipality |
|---|---|---|---|---|---|
| 515 | Haderslev | 657 | Herning | 779 | Skive |
| 537 | Sønderborg | 661 | Holstebro | 787 | Thisted |
| 545 | Aabenraa | 663 | Ikast | 791 | Viborg |
| 561 | Esbjerg | 671 | Struer | 805 | Brønderslev |
| 573 | Varde | 707 | Grenaa | 813 | Frederikshavn |
| 607 | Fredericia | 731 | Randers | 821 | Hjørring |
| 615 | Horsens | 743 | Silkeborg | 823 | Hobro |
| 621 | Kolding | 745 | Skanderborg | 841 | Skagen |
| 631 | Vejle | 751 | Århus | 851 | Aalborg |

Source: GfK background data 1999

**Table F.1.5 Jutland other municipality**

| Code | Municipality | Code | Municipality | Code | Municipality |
|---|---|---|---|---|---|
| 501 | Augustenborg | 617 | Jelling | 741 | Samsø |
| 503 | Bov | 619 | Juelsminde | 747 | Sønderhald |
| 505 | Bredebro | 623 | Lunderskov | 749 | Them |
| 507 | Broager | 625 | Nr. Snede | 761 | Bjerringbro |
| 509 | Christiansfeld | 627 | Tørring-Uldum | 763 | Fjends |
| 511 | Gram | 629 | Vamdrup | 767 | Hvorslev |
| 513 | Gråsten | 651 | Aulum-Haderup | 769 | Karup |
| 517 | Højer | 653 | Brande | 771 | Kjellerup |
| 519 | Lundtoft | 655 | Egvad | 773 | Morsø |
| 521 | Løgumkloster | 659 | Holmsland | 775 | Møldrup |
| 523 | Nordborg | 665 | Lemvig | 777 | Sallingsund |
| 525 | Nr. Rangstrup | 667 | Ringkøbing | 781 | Spøttrup |
| 527 | Rødding | 669 | Skjern | 783 | Sundsøre |
| 529 | Rødekro | 673 | Thyborøn-Harboøre | 785 | Sydthy |
| 531 | Skærbæk | 675 | Thyholm | 789 | Tjele |
| 533 | Sundeved | 677 | Trehøje | 793 | Aalestrup |
| 535 | Sydals | 679 | Ulfborg-Vemb | 801 | Arden |
| 539 | Tinglev | 681 | Videbæk | 803 | Brovst |
| 541 | Tønder | 683 | Vinderup | 807 | Dronninglund |
| 543 | Vojens | 685 | Åskov | 809 | Farsø |
| 551 | Billund | 701 | Ebeltoft | 811 | Fjerritslev |
| 553 | Blåbjerg | 703 | Galten | 815 | Hadsund |
| 555 | Blåvandshuk | 705 | Gjern | 817 | Hals |
| 557 | Bramming | 709 | Hadsten | 819 | Hirtshals |
| 559 | Brørup | 711 | Hammel | 825 | Læsø |
| 563 | Fanø | 713 | Hinnerup | 827 | Løgstør |
| 565 | Grindsted | 715 | Hørning | 829 | Løkken-Vrå |
| 567 | Helle | 717 | Langå | 831 | Nibe |
| 569 | Holsted | 719 | Mariager | 833 | Nørager |
| 571 | Ribe | 721 | Midtdjurs | 835 | Pandrup |
| 575 | Vejen | 723 | Nørhald | 837 | Sejlflod |
| 577 | Ølgod | 725 | Nørre Djurs | 839 | Sindal |
| 601 | Brædstrup | 727 | Odder | 843 | Skørping |
| 603 | Børkop | 729 | Purhus | 845 | Støvring |
| 605 | Egtved | 733 | Rosenholm | 847 | Sæby |
| 609 | Gedved | 735 | Rougsø | 849 | Aabybro |
| 611 | Give | 737 | Ry | 861 | Aars |
| 613 | Hedensted | 739 | Rønde | | |

Source: GfK background data 1999

**Table F.1.6 Not in data**

| Code | Municipality |
|---|---|
| 411 | Christiansø |
| 493 | Ærøskøbing |
| 765 | Hanstholm |

Source: GfK background data 1999

## F.2 By municipality code

**Table F.2.1 Non-Jutland**

| Code | Municipality | Categ. | Code | Municipality | Categ. | Code | Municipality | Categ. |
|---|---|---|---|---|---|---|---|---|
| 101 | Københavns Kommune | Capital | 263 | Ramsø | Other I. | 385 | Rønnede | Other I. |
| 147 | Frederiksberg Kommune | Capital | 265 | Roskilde | Island C. | 387 | Sakskøbing | Other I. |
| 151 | Ballerup | Capital | 267 | Skovbo | Other I. | 389 | Stevns | Other I. |
| 153 | Brøndby | Capital | 269 | Solrød | Island C. | 391 | Stubbekøbing | Other I. |
| 155 | Dragør | Capital | 271 | Vallø | Other I. | 393 | Suså | Other I. |
| 157 | Gentofte | Capital | 301 | Bjergsted | Other I. | 395 | Sydfalster | Other I. |
| 159 | Gladsaxe | Capital | 303 | Dianalund | Other I. | 397 | Vordingborg | Other I. |
| 161 | Glostrup | Capital | 305 | Dragsholm | Other I. | 401 | Allinge-Gudhj. | Other I. |
| 163 | Herlev | Capital | 307 | Fuglebjerg | Other I. | 403 | Hasle | Other I. |
| 165 | Albertslund | Capital | 309 | Gørlev | Other I. | 405 | Nexø | Other I. |
| 167 | Hvidovre | Capital | 311 | Hashøj | Other I. | 407 | Rønne | Island C. |
| 169 | Høje Taastrup | Capital | 313 | Haslev | Other I. | 409 | Aakirkeby | Other I. |
| 171 | Ledøje-Smørum | Capital | 315 | Holbæk | Island C. | 411 | Christiansø, not in data | |
| 173 | Lyngby-Taarbæk | Capital | 317 | Hvidebæk | Other I. | 421 | Assens | Other I. |
| 175 | Rødovre | Capital | 319 | Høng | Other I. | 423 | Bogense | Other I. |
| 181 | Søllerød | Capital | 321 | Jernløse | Other I. | 425 | Broby | Other I. |
| 183 | Ishøj | Capital | 323 | Kalundborg | Island C. | 427 | Egebjerg | Other I. |
| 185 | Tårnby | Capital | 325 | Korsør | Island C. | 429 | Ejby | Other I. |
| 187 | Vallensbæk | Capital | 327 | Nykøbing-Rørvig | Other I. | 431 | Faaborg | Other I. |
| 189 | Værløse | Capital | 329 | Ringsted | Island C. | 433 | Glamsbjerg | Other I. |
| 201 | Allerød | Island C. | 331 | Skælskør | Other I. | 435 | Gudme | Other I. |
| 205 | Birkerød | Island C. | 333 | Slagelse | Island C. | 437 | Haarby | Other I. |
| 207 | Farum | Island C. | 335 | Sorø | Other I. | 439 | Kerteminde | Other I. |
| 208 | Fredensborg-Humlebæk | Island C. | 337 | Stenlille | Other I. | 441 | Langeskov | Other I. |
| 209 | Frederikssund | Island C. | 339 | Svinninge | Other I. | 443 | Marstal | Other I. |
| 211 | Frederiksværk | Island C. | 341 | Tornved | Other I. | 445 | Middelfart | Island C. |
| 213 | Græsted-Gilleleje | Other I. | 343 | Trundholm | Other I. | 447 | Munkebo | Other I. |
| 215 | Helsinge | Other I. | 345 | Tølløse | Other I. | 449 | Nyborg | Island C. |
| 217 | Helsingør | Island C. | 351 | Fakse | Other I. | 451 | Nørre Aaby | Other I. |
| 219 | Hillerød | Island C. | 353 | Fladså | Other I. | 461 | Odense | Island C. |
| 221 | Hundested | Other I. | 355 | Holeby | Other I. | 471 | Otterup | Other I. |
| 223 | Hørsholm | Island C. | 357 | Holmegaard | Other I. | 473 | Ringe | Other I. |
| 225 | Jægerspris | Other I. | 359 | Højreby | Other I. | 475 | Rudkøbing | Other I. |
| 227 | Karlebo | Island C. | 361 | Langebæk | Other I. | 477 | Ryslinge | Other I. |
| 229 | Skibby | Other I. | 363 | Maribo | Other I. | 479 | Svendborg | Island C. |
| 231 | Skævinge | Other I. | 365 | Møn | Other I. | 481 | Sydlangeland | Other I. |
| 233 | Slangerup | Other I. | 367 | Nakskov | Island C. | 483 | Søndersø | Other I. |
| 235 | Stenløse | Island C. | 369 | Nykøbing F. | Island C. | 485 | Tommerup | Other I. |
| 237 | Ølstykke | Other I. | 371 | Nysted | Other I. | 487 | Tranekær | Other I. |
| 251 | Bramsnæs | Other I. | 373 | Næstved | Island C. | 489 | Ullerslev | Other I. |
| 253 | Greve | Island C. | 375 | Nr. Alslev | Other I. | 491 | Vissenbjerg | Other I. |
| 255 | Gundsø | Other I. | 377 | Præstø | Other I. | 493 | Ærøskøbing, not in data | |
| 257 | Hvalsø | Other I. | 379 | Ravnsborg | Other I. | 495 | Ørbæk | Other I. |
| 259 | Køge | Island C. | 381 | Rudbjerg | Other I. | 497 | Årslev | Other I. |
| 261 | Lejre | Other I. | 383 | Rødby | Other I. | 499 | Aarup | Other I. |

Source: GfK background data 1999

**Table F.2.2 Jutland**

| Code | Municipality | Categ. | Code | Municipality | Categ. | Code | Municipality | Categ. |
|---|---|---|---|---|---|---|---|---|
| 501 | Augustenborg | Other | 621 | Kolding | *City* | 747 | Sønderhald | Other |
| 503 | Bov | Other | 623 | Lunderskov | Other | 749 | Them | Other |
| 505 | Bredebro | Other | 625 | Nr. Snede | Other | 751 | Århus | *City* |
| 507 | Broager | Other | 627 | Tørring-Uldum | Other | 761 | Bjerringbro | Other |
| 509 | Christiansfeld | Other | 629 | Vamdrup | Other | 763 | Fjends | Other |
| 511 | Gram | Other | 631 | Vejle | *City* | 765 | Hanstholm, not in data | |
| 513 | Gråsten | Other | 651 | Aulum-Haderup | Other | 767 | Hvorslev | Other |
| 515 | Haderslev | *City* | 653 | Brande | Other | 769 | Karup | Other |
| 517 | Højer | Other | 655 | Egvad | Other | 771 | Kjellerup | Other |
| 519 | Lundtoft | Other | 657 | Herning | *City* | 773 | Morsø | Other |
| 521 | Løgumkloster | Other | 659 | Holmsland | Other | 775 | Møldrup | Other |
| 523 | Nordborg | Other | 661 | Holstebro | *City* | 777 | Sallingsund | Other |
| 525 | Nr. Rangstrup | Other | 663 | Ikast | *City* | 779 | Skive | *City* |
| 527 | Rødding | Other | 665 | Lemvig | Other | 781 | Spøttrup | Other |
| 529 | Rødekro | Other | 667 | Ringkøbing | Other | 783 | Sundsøre | Other |
| 531 | Skærbæk | Other | 669 | Skjern | Other | 785 | Sydthy | Other |
| 533 | Sundeved | Other | 671 | Struer | *City* | 787 | Thisted | *City* |
| 535 | Sydals | Other | 673 | Thyborøn-Harboøre | Other | 789 | Tjele | Other |
| 537 | Sønderborg | *City* | 675 | Thyholm | Other | 791 | Viborg | *City* |
| 539 | Tinglev | Other | 677 | Trehøje | Other | 793 | Aalestrup | Other |
| 541 | Tønder | Other | 679 | Ulfborg-Vemb | Other | 801 | Arden | Other |
| 543 | Vojens | Other | 681 | Videbæk | Other | 803 | Brovst | Other |
| 545 | Aabenraa | *City* | 683 | Vinderup | Other | 805 | Brønderslev | *City* |
| 551 | Billund | Other | 685 | Åskov | Other | 807 | Dronninglund | Other |
| 553 | Blåbjerg | Other | 701 | Ebeltoft | Other | 809 | Farsø | Other |
| 555 | Blåvandshuk | Other | 703 | Galten | Other | 811 | Fjerritslev | Other |
| 557 | Bramming | Other | 705 | Gjern | Other | 813 | Frederikshavn | *City* |
| 559 | Brørup | Other | 707 | Grenaa | *City* | 815 | Hadsund | Other |
| 561 | Esbjerg | *City* | 709 | Hadsten | Other | 817 | Hals | Other |
| 563 | Fanø | Other | 711 | Hammel | Other | 819 | Hirtshals | Other |
| 565 | Grindsted | Other | 713 | Hinnerup | Other | 821 | Hjørring | *City* |
| 567 | Helle | Other | 715 | Hørning | Other | 823 | Hobro | *City* |
| 569 | Holsted | Other | 717 | Langå | Other | 825 | Læsø | Other |
| 571 | Ribe | Other | 719 | Mariager | Other | 827 | Løgstør | Other |
| 573 | Varde | *City* | 721 | Midtdjurs | Other | 829 | Løkken-Vrå | Other |
| 575 | Vejen | Other | 723 | Nørhald | Other | 831 | Nibe | Other |
| 577 | Ølgod | Other | 725 | Nørre Djurs | Other | 833 | Nørager | Other |
| 601 | Brædstrup | Other | 727 | Odder | Other | 835 | Pandrup | Other |
| 603 | Børkop | Other | 729 | Purhus | Other | 837 | Sejlflod | Other |
| 605 | Egtved | Other | 731 | Randers | *City* | 839 | Sindal | Other |
| 607 | Fredericia | *City* | 733 | Rosenholm | Other | 841 | Skagen | *City* |
| 609 | Gedved | Other | 735 | Rougsø | Other | 843 | Skørping | Other |
| 611 | Give | Other | 737 | Ry | Other | 845 | Støvring | Other |
| 613 | Hedensted | Other | 739 | Rønde | Other | 847 | Sæby | Other |
| 615 | Horsens | *City* | 741 | Samsø | Other | 849 | Aabybro | Other |
| 617 | Jelling | Other | 743 | Silkeborg | *City* | 851 | Aalborg | *City* |
| 619 | Juelsminde | Other | 745 | Skanderborg | *City* | 861 | Aars | Other |

Source: GfK background data 1999

# Appendix G Estimation results



**Table G.1 Comparing result of GAUSS and Limdep estimation, using relative prices**

| | GAUSS results | | | | | | Limdep results | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Subsample** | **A** | | **B** | | **C** | | **A** | | **B** | | **C** | |
| **Parameters** | | | | | | | | | | | | |
| Price | -1.06 | ** | -0.31 | ** | -0.32 | ** | -1.06 | ** | -0.31 | ** | -0.32 | ** |
| Barn eggs | -0.70 | ** | -0.92 | ** | -0.95 | ** | -0.70 | ** | -0.92 | ** | -0.95 | ** |
| Free-range eggs | -0.49 | ** | -1.34 | ** | -1.36 | ** | -0.49 | ** | -1.34 | ** | -1.36 | ** |
| Organic eggs | 0.16 | * | -0.34 | ** | -0.21 | * | 0.16 | ** | -0.34 | ** | -0.21 | ** |
| **WTP in percent/100 of battery egg price** | | | | | | | | | | | | |
| Barn eggs | -0.66 | ** | -2.94 | ** | -2.98 | ** | -0.66 | ** | -2.94 | ** | -2.98 | ** |
| Free-range eggs | -0.46 | ** | -4.27 | ** | -4.28 | ** | -0.46 | ** | -4.27 | ** | -4.28 | ** |
| Organic eggs | 0.15 | * | -1.09 | ** | -0.67 | * | 0.15 | ** | -1.09 | ** | -0.67 | ** |
| **Log likelihood value** | | | | | | | | | | | | |
| | -30,664 | | -26,204 | | -19,612 | | -30,664 | | -26,204 | | -19,612 | |

Source: Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Separate estimations using data from one subsample at a time.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.

**Table G.2 Standard errors using GAUSS and Limdep. Relative prices**

| | GAUSS, Robust standard errors | | | Limdep, non-Robust standard errors | | |
|---|---|---|---|---|---|---|
| **Sub sample** | **A** | **B** | **C** | **A** | **B** | **C** |
| Price | 0.096 | 0.099 | 0.097 | 0.041 | 0.046 | 0.052 |
| Barn eggs | 0.053 | 0.057 | 0.065 | 0.021 | 0.022 | 0.026 |
| Free-range eggs | 0.071 | 0.081 | 0.081 | 0.026 | 0.032 | 0.038 |
| Organic eggs | 0.071 | 0.088 | 0.091 | 0.028 | 0.034 | 0.037 |

Source: Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Separate estimations using data from one subsample at a time.

**Table G.3 Different base brands, estimating on Superbrugsen, with relative prices**

| Superbrugsen | Wtp when base brand is: | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Wtp for:** | **Battery eggs** | | **Barn eggs** | | **Free-range eggs** | | **Organic eggs** | |
| Battery eggs | 0.00 | | -0.57 | | -0.06 | | -1.11 | |
| Barn eggs | 0.57 | ** | 0.00 | | 0.50 | ** | -0.54 | ** |
| Free range eggs | 0.06 | | -0.51 | ** | 0.00 | | -1.04 | ** |
| Organic eggs | 1.11 | ** | 0.54 | ** | 1.04 | ** | 0.00 | |
| | **Recalculated wtp relative to battery eggs** | | | | | | | |
| Battery eggs | 0.000 | | 0.000 | | 0.000 | | 0.000 | |
| Barn eggs | 0.568 | | 0.567 | | 0.567 | | 0.568 | |
| Free range eggs | 0.062 | | 0.062 | | 0.063 | | 0.062 | |
| Organic eggs | 1.106 | | 1.106 | | 1.105 | | 1.106 | |

Source: Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level. Relative prices mean that prices are divided by the price of the battery egg.

**Table G.4 Estimation results using different number of repetitions**

| Number of repetitions: | 250 | 500 | 1000 | 1500 | 2000 |
|---|---|---|---|---|---|
| **Parameters** | | | | | |
| Price | -0.27 | -0.28 | -0.28 | -0.28 | -0.28 |
| E(Barn eggs) | -1.23 | -1.24 | -1.25 | -1.25 | -1.26 |
| STD(Barn eggs) | 4.33 | 4.45 | 4.48 | 4.42 | 4.42 |
| E(Free-range eggs) | -1.93 | -1.93 | -1.98 | -1.98 | -1.98 |
| STD(Free-range eggs) | 4.82 | 5.03 | 5.06 | 4.98 | 5.03 |
| E(Organic eggs) | -1.53 | -1.74 | -1.70 | -1.65 | -1.63 |
| STD(Organic eggs) | -8.98 | -8.68 | -8.82 | -8.82 | -8.88 |
| **Mean wtp** | | | | | |
| Barn eggs | -4.59 | -4.40 | -4.46 | -4.50 | -4.50 |
| Free-range eggs | -7.21 | -6.82 | -7.07 | -7.13 | -7.08 |
| Organic eggs | -5.71 | -6.17 | -6.09 | -5.95 | -5.83 |
| **Minimum wtp** | | | | | |
| Barn eggs | -20.79 | -20.16 | -20.50 | -20.44 | -20.30 |
| Free-range eggs | -14.43 | -13.65 | -14.15 | -14.27 | -14.16 |
| Organic eggs | -39.32 | -36.92 | -37.67 | -37.77 | -37.59 |
| **Maximum wtp** | | | | | |
| barn | 11.62 | 11.36 | 11.57 | 11.45 | 11.31 |
| Free-range | 10.84 | 11.01 | 11.05 | 10.85 | 10.90 |
| Organic | 27.90 | 24.59 | 25.49 | 25.88 | 25.94 |

Source: Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Using data from subsample C.
Note: The parameter for price is fixed and the parameters for each egg type is mixed with the *triangular* distribution.

**Table G.5 Superbrugsen by geography**

| Superbrugsen | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Std. dev. | | Mean | | Std. dev. | | Mean | | Std. dev. | |
| | **Parameters** | | | | | | | | | | | |
| Price | | | | | -2.25 | ** | | | | | | |
| Capital | 1.98 | ** | 1.37 | ** | 1.64 | ** | 1.42 | ** | 3.32 | ** | 3.44 | ** |
| Island City-municip. | 0.88 | ** | 0.61 | | 0.91 | | -2.15 | ** | 1.32 | | 4.14 | ** |
| Oth. Island municip. | 0.73 | * | 2.32 | ** | -2.08 | * | -4.19 | ** | -0.35 | | -6.16 | ** |
| Jutland city municip. | 0.72 | | 2.28 | ** | -0.29 | | 1.98 | ** | 1.38 | * | 4.39 | ** |
| Jutland other municip. | 0.59 | * | 2.14 | ** | -1.23 | * | 2.28 | ** | -0.25 | | 5.29 | ** |
| | **Willingness to pay (wtp)** | | | | | | | | | | | |
| Capital | 0.88 | | 0.61 | | 0.73 | | 0.63 | | 1.48 | | 1.53 | |
| Island City-municip. | 0.39 | | 0.27 | | 0.40 | | 0.96 | | 0.59 | | 1.84 | |
| Oth. Island municip. | 0.33 | | 1.03 | | -0.93 | | 1.86 | | -0.16 | | 2.74 | |
| Jutland city municip. | 0.32 | | 1.01 | | -0.13 | | 0.88 | | 0.61 | | 1.95 | |
| Jutland other municip. | 0.26 | | 0.95 | | -0.55 | | 1.02 | | -0.11 | | 2.35 | |
| | **Percentage of households that have positive wtp** | | | | | | | | | | | |
| Capital | 92 | | | | 88 | | | | 83 | | | |
| Island City-municip. | 92 | | | | 66 | | | | 63 | | | |
| Oth. Island municip. | 62 | | | | 31 | | | | 48 | | | |
| Jutland city municip. | 62 | | | | 44 | | | | 62 | | | |
| Jutland other municip. | 61 | | | | 30 | | | | 48 | | | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. Parameters for egg types are mixed with the normal distribution.
The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
In this study the standard errors of the estimated parameters are ignored when calculating willingness to pay. The estimated distribution of willingness to pay therefore has no standard error.

**Table G.6 Absolute prices, two different subsamples**

| | Subsample B | | Subsample C | |
|---|---|---|---|---|
| **About the subsample:** | | | | |
| Number of observations | 21,050 | | 15,816 | |
| Number of families | 1,846 | | 1,693 | |
| **Parameters:** | | | | |
| Price | -0.16 *(0.100)*† | | -0.06 *(0.098)* | |
| Barn eggs | -0.95 *(0.061)* | ** | -1.01 *(0.069)* | ** |
| Free-range eggs | -1.39 *(0.091)* | ** | -1.48 *(0.089)* | ** |
| Organic eggs | -0.41 *(0.099)* | ** | -0.36 *(0.100)* | ** |
| **Log-likelihood value**[1] | -26,220 | | -19,630 | |

Source: Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Separate estimations using data from one subsample at a time.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
†: Numbers in parenthesis are standard errors.

[1] The likelihood value is the sum of the individual likelihood values. The number of families therefore influences the likelihood value.

**Table G.7 Estimation results using the conventional multinomial logit**

| Aggregated store | Price | | Barn eggs | | Free-range eggs | | Organic eggs | | Log likelihood val. |
|---|---|---|---|---|---|---|---|---|---|
| Subsample B | -0.40 | ** | -0.80 | ** | -1.23 | ** | -0.31 | ** | -25,148 |
| | *(0.089)* † | | *(0.056)* | | *(0.073)* | | *(0.084)* | | |
| Subsample C | -0.31 | ** | -0.79 | ** | -1.22 | ** | -0.20 | * | -18,975 |
| | *(0.094)* | | *(0.066)* | | *(0.082)* | | *(0.090)* | | |
| SuperBrugsen: | -1.28 | ** | 0.73 | ** | 0.08 | | 1.41 | ** | -3,072 |
| | *(0.280)* | | *(0.170)* | | *(0.238)* | | *(0.250)* | | |
| DagligBrugsen | -0.14 | | -0.56 | * | -0.80 | * | -0.64 | | -787 |
| | *(0.233)* | | *(0.251)* | | *(0.345)* | | *(0.366)* | | |
| Kvickly and OBS | -1.21 | ** | 0.75 | ** | 0.11 | | 1.38 | ** | -2,295 |
| | *(0.254)* | | *(0.163)* | | *(0.237)* | | *(0.257)* | | |
| Irma | -0.27 | | 1.15 | | 1.34 | | 1.79 | | -174 |
| | *(0.418)* | | *(1.157)* | | *(1.164)* | | *(1.182)* | | |
| Fakta | 0.32 | | -0.98 | ** | -1.60 | ** | -0.67 | * | -2,344 |
| | *(0.475)* | | *(0.138)* | | *(0.294)* | | *(0.281)* | | |
| Føtex | -1.12 | ** | 0.07 | | 0.05 | | 0.79 | ** | -2,114 |
| | *(0.170)* | | *(0.134)* | | *(0.152)* | | *(0.186)* | | |
| Netto | -0.10 | | -3.19 | ** | -2.29 | ** | -0.75 | ** | -4,137 |
| | *(0.234)* | | *(0.129)* | | *(0.128)* | | *(0.138)* | | |
| Aldi | -0.53 | | -2.57 | ** | -2.39 | ** | -0.49 | | -620 |
| | *(0.607)* | | *(0.406)* | | *(0.309)* | | *(0.539)* | | |
| Prima | -1.01 | ** | -0.46 | * | -0.62 | * | 0.02 | | -675 |
| | *(0.263)* | | *(0.212)* | | *(0.293)* | | *(0.305)* | | |
| Favør | -0.53 | ** | -0.78 | * | -1.29 | ** | -0.65 | | -481 |
| | *(0.180)* | | *(0.330)* | | *(0.286)* | | *(0.344)* | | |
| Various grocers | -0.47 | * | -0.86 | ** | -1.49 | ** | -0.82 | ** | -4,283 |
| | *(0.198)* | | *(0.092)* | | *(0.143)* | | *(0.167)* | | |
| Various discount | -0.28 | | -1.04 | ** | -1.62 | ** | -1.20 | ** | -1,325 |
| | *(0.345)* | | *(0.216)* | | *(0.262)* | | *(0.290)* | | |
| Greengrocers etc | 0.08 | | 0.26 | | 0.72 | | 0.29 | | -166 |
| | *(0.527)* | | *(0.424)* | | *(0.507)* | | *(0.468)* | | |
| Direct sales | 0.14 | | 1.20 | ** | 2.85 | ** | 1.90 | ** | -2,608 |
| | *(0.310)* | | *(0.310)* | | *(0.265)* | | *(0.291)* | | |
| Bilka | -0.94 | ** | -0.49 | ** | -0.35 | | 0.22 | | -1,016 |
| | *(0.207)* | | *(0.161)* | | *(0.187)* | | *(0.207)* | | |

Source: Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Separate estimations using data from one subsample at a time.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
†: Numbers in parenthesis are standard errors.

**Table G.8 Marginal relative willingness to pay with and without rationing. Only store aggregates with price parameter significantly different from zero**

| Relative WTP in | Without rationing | | | With rationing | | |
|---|---|---|---|---|---|---|
| | Barn eggs | Free-range eggs | Organic eggs | Barn eggs | Free-range eggs | Organic eggs |
| Subsample B | -2.94 | -4.27 | -1.09 | -1.98 | -3.03 | -0.76 |
| Subsample C | -2.98 | -4.28 | -0.67 | -2.50 | -3.87 | -0.64 |
| SuperBrugsen | 0.57 | † | 1.10 | 0.57 | † | 1.11 |
| Kvickly and OBS | 0.63 | † | 1.14 | 0.62 | † | 1.13 |
| Føtex | † | † | 0.70 | † | † | 0.70 |
| Prima | -0.82 | -1.09 | † | -0.45 | -0.61 | † |
| Favør | -1.56 | -3.01 | † | -1.46 | -2.42 | † |
| Various grocers | -1.82 | -3.18 | -1.75 | -1.82 | -3.18 | -1.75 |
| Bilka | -0.78 | -0.75 | † | -0.52 | † | † |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Separate estimations for each store aggregate.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed.
Note: '†' means that the reaction to *egg type* was not significantly different from zero at the five percent level. The marginal willingness to pay is therefore not significantly different from zero either. The relative marginal willingness to pay is measured as a percentage of the battery egg price divided by one hundred. In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated marginal willingness to pay therefore has no standard error.

**Table G.9 Price parameters and standard errors of the parameters in mixed multinomial logit estimations using different subsamples**

| | Parameter | Standard *error* | Significance level |
|---|---|---|---|
| Subsample B | -0.31 | *(0.092)* | ** |
| Subsample C | -0.27 | *(0.103)* | ** |
| SuperBrugsen | -2.38 | *(0.534)* | ** |
| DagligBrugsen | -0.91 | *(0.395)* | * |
| Kvickly and OBS | -2.26 | *(0.415)* | ** |
| Irma | -0.16 | *(0.605)* | |
| Fakta | 0.22 | *(0.771)* | |
| Føtex | -1.60 | *(0.230)* | ** |
| Netto | 0.30 | *(0.351)* | |
| Aldi | 0.16 | *(0.614)* | |
| Prima | -1.35 | *(0.394)* | ** |
| Favør | -0.64 | *(0.402)* | |
| Various grocers | -1.06 | *(0.280)* | ** |
| Various discount stores | -0.79 | *(0.571)* | |
| Greengrocers etc | 0.53 | *(0.890)* | |
| Directly from farms | 0.09 | *(0.492)* | |
| Bilka | -1.10 | *(0.264)* | ** |

Source: Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsample B and C are defined in chapter 4, Table 4.3. Separate estimations on each subsample.
Prices are relative to the price of battery eggs. Rationing is allowed. The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
'**' means that the parameter is significantly different from zero at the one percent level and '*' at the five percent level.

**Table G.10 Mixing the parameters for egg types with the normal distribution**

| Mean and standard deviation of *parameters* for: | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Standard *deviation* | | Mean | | Standard *deviation* | | Mean | | Standard *deviation* | |
| Subsample B | -1.23 *(0.064)* # | ** | 1.74 *(0.061)* | ** | -1.92 *(0.086)* | ** | -2.02 *(0.103)* | ** | -1.82 *(0.175)* | ** | 3.94 *(0.202)* | ** |
| Subsample C | -1.22 *(0.074)* | ** | 1.83 *(0.074)* | ** | -1.92 *(0.102)* | ** | -2.03 *(0.100)* | ** | -1.64 *(0.145)* | ** | 3.76 *(0.143)* | ** |
| SuperBrugsen | 1.03 *(0.185)* | ** | -2.12 *(0.174)* | ** | -0.30 *(0.366)* | | -2.70 *(0.266)* | ** | 1.18 *(0.452)* | ** | 4.82 *(0.410)* | ** |
| Dagligbrugsen | -0.74 *(0.376)* | * | 2.64 *(0.513)* | ** | -2.69 *(0.749)* | ** | 4.20 *(0.582)* | ** | -3.26 *(1.017)* | ** | 6.47 *(0.884)* | ** |
| Kvickly and OBS | 1.26 *(0.183)* | ** | 1.96 *(0.228)* | ** | -0.06 *(0.333)* | | 2.31 *(0.260)* | ** | 0.26 *(0.478)* | | 4.44 *(0.473)* | ** |
| Føtex | -0.02 *(0.161)* | | 1.47 *(0.189)* | ** | -0.23 *(0.197)* | | 1.78 *(0.181)* | ** | 0.10 *(0.275)* | | 3.24 *(0.296)* | ** |
| Prima | -0.39 *(0.280)* | | 1.53 *(0.297)* | ** | -1.57 *(0.594)* | ** | 2.47 *(0.518)* | ** | -0.70 *(0.570)* | | 4.30 *(0.816)* | ** |
| Various grocers | -1.15 *(0.122)* | ** | 1.63 *(0.108)* | ** | -2.42 *(0.253)* | ** | 2.44 *(0.328)* | ** | -3.58 *(0.407)* | ** | 5.04 *(0.334)* | ** |
| Bilka | -0.77 *(0.221)* | ** | 1.26 *(0.193)* | ** | -0.81 *(0.279)* | ** | 1.62 *(0.328)* | ** | -1.17 *(0.367)* | ** | 3.23 *(0.406)* | ** |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsamples B and C are defined in chapter 4, Table 4.3. Separate estimations on each subsample.
Prices are relative to the price of battery eggs. Rationing is allowed. The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
'**' means that the parameter is significantly different from zero at the one percent level and '*' at the five percent level.
#: Numbers in parenthesis are standard errors of the estimated parameters.
Note that the number of observations and families varies between subsamples, leading to very different values of the likelihood function.

**Table G.11 Relative marginal willingness to pay for different egg types in store aggregates with price parameters significantly different from zero**

| Relative marginal willingness to pay: | Barn eggs | | Free-range eggs | | Organic eggs | |
|---|---|---|---|---|---|---|
| | Mean | Standard *deviation* | Mean | Standard *deviation* | Mean | Standard *Deviation* |
| Subsample B | -3.99 | 5.68 | -6.26 | 6.59 | -5.92 | 12.83 |
| Subsample C | -4.57 | 6.84 | -7.17 | 7.60 | -6.14 | 14.08 |
| SuperBrugsen | 0.43 | 0.89 | -0.13 | 1.14 | 0.50 | 2.02 |
| DagligBrugsen | -0.82 | 2.91 | -2.96 | 4.63 | -3.60 | 7.14 |
| Kvickly and OBS | 0.56 | 0.87 | -0.02 | 1.02 | 0.11 | 1.96 |
| Føtex | -0.01 | 0.92 | -0.15 | 1.12 | 0.07 | 2.03 |
| Prima | -0.29 | 1.14 | -1.17 | 1.83 | -0.52 | 3.19 |
| Various grocers | -1.08 | 1.53 | -2.27 | 2.29 | -3.36 | 4.73 |
| Bilka | -0.70 | 1.14 | -0.74 | 1.47 | -1.06 | 2.94 |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsamples B and C are defined in chapter 4, Table 4.3. Separate estimations on each subsample. Prices are relative which means that the relative marginal willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit. The starting values for the standard deviations are set to 0.1.
In this study the standard errors of the estimated parameters are ignored when calculating marginal willingness to pay. The estimated distribution of marginal willingness to pay therefore has no standard error.

**Table G.12 Likelihood ratio test for the need of mixing**

| Aggregated store | Mixed ($lnL_1$) | Not mixed ($lnL_0$) | $-2(lnL_0-lnL_1)$ | Test-Probability $\left(\chi_3^2\right)$ |
|---|---|---|---|---|
| Subsample B | -13,164 | -19,612 | 12,896 | 0.000 |
| Subsample C | -17,420 | -26,204 | 17,568 | 0.000 |
| SuperBrugsen | -1,972 | -3,072 | 2,201 | 0.000 |
| DagligBrugsen | -519 | -787 | 536 | 0.000 |
| Kvickly and OBS | -1,618 | -2,295 | 1,354 | 0.000 |
| Føtex | -1,718 | -2,114 | 792 | 0.000 |
| Prima | -545 | -675 | 261 | 0.000 |
| Various grocers | -3,224 | -4,283 | 2,117 | 0.000 |
| Bilka | -852 | -1,016 | 329 | 0.000 |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000. Subsamples B and C are defined in chapter 4, Table 4.3. Separate estimations on each subsample. Prices are relative to the price of battery eggs. Rationing is allowed. The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation. The test probability is the probability that the conventional multinomial logit is as good as the mixed multinomial logit.

**Table G.13 Subsample C by geography**

| Sub-sample C | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Std. dev. | | Mean | | Std. dev. | | Mean | | Std. dev. | |
| | **Parameters** | | | | | | | | | | | |
| Price | | | | | -0.26 | * | | | | | | |
| Capital | -1.42 | ** | -1.67 | ** | -1.47 | ** | 1.63 | ** | -0.48 | * | 3.34 | ** |
| Island City-municip. | -1.23 | ** | 1.46 | ** | -1.88 | ** | -1.99 | ** | -1.66 | ** | 4.18 | ** |
| Oth. Island municip. | -1.36 | ** | 2.27 | ** | -2.93 | ** | 2.72 | ** | -3.24 | ** | 5.13 | ** |
| Jutland city municip. | -0.91 | ** | 1.78 | ** | -1.73 | ** | 2.00 | ** | -1.21 | ** | 3.54 | ** |
| Jutland other municip. | -0.89 | ** | 1.90 | ** | -2.93 | ** | 2.15 | ** | -3.08 | ** | 3.36 | ** |
| | **Willingness to pay (wtp)** | | | | | | | | | | | |
| Capital | -5.43 | | 6.41 | | -5.65 | | 6.26 | | -1.85 | | 12.81 | |
| Island City-municip. | -4.72 | | 5.60 | | -7.22 | | 7.64 | | -6.37 | | 16.02 | |
| Oth. Island municip. | -5.20 | | 8.69 | | -11.22 | | 10.44 | | -12.41 | | 19.67 | |
| Jutland city municip. | -3.50 | | 6.83 | | -6.64 | | 7.67 | | -4.62 | | 13.58 | |
| Jutland other municip. | -3.41 | | 7.27 | | -11.23 | | 8.24 | | -11.82 | | 12.89 | |
| | **Percentage of households that have positive wtp** | | | | | | | | | | | |
| Capital | 20 | | | | 18 | | | | 44 | | | |
| Island City-municip. | 20 | | | | 17 | | | | 35 | | | |
| Oth. Island municip. | 27 | | | | 14 | | | | 26 | | | |
| Jutland city municip. | 30 | | | | 19 | | | | 37 | | | |
| Jutland other municip. | 32 | | | | 9 | | | | 18 | | | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. Parameters for egg types are mixed with the normal distribution.
The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
In this study the standard errors of the estimated parameters are ignored when calculating willingness to pay. The estimated distribution of willingness to pay therefore has no standard error.

**Table G.14 Subsample C by age of main buyer**

| Sub-sample C | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Standard deviation | | Mean | | Standard deviation | | Mean | | Standard deviation | |
| | Parameters | | | | | | | | | | | |
| Price | -0.27 | ** | | | | | | | | | | |
| < 30 years | -1.16 | ** | -1.41 | ** | -1.59 | ** | -1.56 | ** | -1.29 | ** | -3.78 | ** |
| 30 - 44 years | -1.63 | ** | 2.10 | ** | -2.15 | ** | 2.08 | ** | -1.87 | ** | 4.07 | ** |
| 45 - 59 years | -1.16 | ** | 1.84 | ** | -2.17 | ** | -2.00 | ** | -2.22 | ** | 4.02 | ** |
| > 60 years | -0.85 | ** | 1.76 | ** | -1.56 | ** | 1.79 | ** | -0.99 | ** | 3.28 | ** |
| | Willingness to pay (wtp) | | | | | | | | | | | |
| < 30 years | -4.29 | | 5.20 | | -5.85 | | 5.75 | | -4.76 | | 13.89 | |
| 30 - 44 years | -5.99 | | 7.72 | | -7.91 | | 7.66 | | -6.89 | | 14.97 | |
| 45 - 59 years | -4.27 | | 6.78 | | -7.97 | | 7.37 | | -8.15 | | 14.79 | |
| > 60 years | -3.11 | | 6.48 | | -5.73 | | 6.57 | | -3.64 | | 12.05 | |
| | Percentage of households that have positive wtp | | | | | | | | | | | |
| < 30 years | 21 | | | | 15 | | | | 37 | | | |
| 30 - 44 years | 22 | | | | 15 | | | | 32 | | | |
| 45 - 59 years | 26 | | | | 14 | | | | 29 | | | |
| > 60 years | 32 | | | | 19 | | | | 38 | | | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. Parameters for egg types are mixed with the normal distribution.
The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
In this study the standard errors of the estimated parameters are ignored when calculating willingness to pay. The estimated distribution of willingness to pay therefore has no standard error.

**Table G.15 Subsample C by attitudes to branded goods**

| Sub-sample C | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Standard deviation | | Mean | | Standard deviation | | Mean | | Standard deviation | |
| | Parameters | | | | | | | | | | | |
| Price | -0.32 | ** | | | | | | | | | | |
| Brand > low price | -0.64 | ** | -1.74 | ** | -1.36 | ** | -2.19 | ** | -0.31 | | 3.88 | ** |
| Low price > brand | -1.44 | ** | 1.89 | ** | -2.06 | ** | 1.87 | ** | -2.02 | ** | 3.64 | ** |
| | Willingness to pay (wtp) | | | | | | | | | | | |
| Brand > low price | -1.99 | | 5.44 | | -4.27 | | 6.87 | | -0.97 | | 12.13 | |
| Low price > brand | -4.50 | | 5.91 | | -6.44 | | 5.87 | | -6.31 | | 11.38 | |
| | Percentage of households that have positive wtp | | | | | | | | | | | |
| Brand > low price | 36 | | | | 27 | | | | 47 | | | |
| Low price > brand | 22 | | | | 14 | | | | 29 | | | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. Parameters for egg types are mixed with the normal distribution.
The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.

In this study the standard errors of the estimated parameters are ignored when calculating willingness to pay. The estimated distribution of willingness to pay therefore has no standard error.

**Table G.16 Superbrugsen by attitudes to branded goods**

| Superbrugsen | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Standard deviation | | Mean | | Standard deviation | | Mean | | Standard deviation | |
| | Parameters | | | | | | | | | | | |
| Price | -2.36 | ** | | | | | | | | | | |
| Brand > low price | 1.31 | ** | 2.44 | ** | -0.18 | | 3.05 | ** | 2.01 | ** | 4.91 | ** |
| Low price > brand | 0.92 | ** | 1.98 | ** | -0.34 | | 2.43 | ** | 0.63 | | 4.85 | ** |
| | Willingness to pay (wtp) | | | | | | | | | | | |
| Brand > low price | 0.56 | | 1.03 | | -0.08 | | 1.29 | | 0.85 | | 2.08 | |
| Low price > brand | 0.39 | | 0.84 | | -0.14 | | 1.03 | | 0.27 | | 2.06 | |
| | Percentage of households that have positive wtp | | | | | | | | | | | |
| Brand > low price | 70 | | | | 48 | | | | 66 | | | |
| Low price > brand | 68 | | | | 44 | | | | 55 | | | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. Parameters for egg types are mixed with the normal distribution.
The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
In this study the standard errors of the estimated parameters are ignored when calculating willingness to pay. The estimated distribution of willingness to pay therefore has no standard error.

**Table G.17 Føtex by attitudes to branded goods**

| Føtex | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Standard deviation | | Mean | | Standard deviation | | Mean | | Standard deviation | |
| | Parameters | | | | | | | | | | | |
| Price | -1.60 | ** | | | | | | | | | | |
| Brand > low price | -0.04 | | 1.54 | ** | -0.01 | | 1.71 | ** | 1.47 | ** | 3.96 | ** |
| Low price > brand | 0.02 | | 1.34 | ** | -0.29 | | 1.59 | ** | -0.26 | | 2.75 | ** |
| | Willingness to pay (wtp) | | | | | | | | | | | |
| Brand > low price | -0.02 | | 0.96 | | 0.00 | | 1.07 | | 0.92 | | 2.47 | |
| Low price > brand | 0.01 | | 0.84 | | -0.18 | | 0.99 | | -0.16 | | 1.72 | |
| | Percentage of households that have positive wtp | | | | | | | | | | | |
| Brand > low price | 49 | | | | 50 | | | | 64 | | | |
| Low price > brand | 50 | | | | 43 | | | | 46 | | | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. Parameters for egg types are mixed with the normal distribution.
The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
In this study the standard errors of the estimated parameters are ignored when calculating willingness to pay. The estimated distribution of willingness to pay therefore has no standard error.

**Table G.18 Bilka by attitudes to branded goods**

| Bilka | Barn eggs | | | | Free-range eggs | | | | Organic eggs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | | Standard deviation | | Mean | | Standard deviation | | Mean | | Standard deviation | |
| | Parameters | | | | | | | | | | | |
| Price | | | | | -1.06 | ** | | | | | | |
| Brand > low price | -0.17 | | 0.68 | | 0.16 | | -0.88 | * | 0.06 | | 2.48 | ** |
| Low price > brand | -1.02 | ** | 1.45 | ** | -1.03 | ** | 1.58 | ** | -1.96 | ** | 3.62 | ** |
| | Willingness to pay (wtp) | | | | | | | | | | | |
| Brand > low price | -0.16 | | 0.64 | | 0.15 | | 0.83 | | 0.06 | | 2.33 | |
| Low price > brand | -0.96 | | 1.36 | | -0.97 | | 1.49 | | -1.84 | | 3.40 | |
| | Percentage of households that have positive wtp | | | | | | | | | | | |
| Brand > low price | 40 | | | | 57 | | | | 51 | | | |
| Low price > brand | 24 | | | | 26 | | | | 29 | | | |

Estimations using GfK purchase data on eggs from 26 June 1999 to 30 June 2000.
Prices are relative which means that the relative willingness to pay is measured in percent of the battery egg price divided by one hundred. Rationing is allowed. Parameters for egg types are mixed with the normal distribution.
The number of repetitions is 500. The starting values for means are taken from the conventional multinomial logit and are set to 0.1 for the standard deviation.
Note: '**' means that the parameter is significantly different from zero at the 1% level, '*' at the 5% level.
In this study the standard errors of the estimated parameters are ignored when calculating willingness to pay. The estimated distribution of willingness to pay therefore has no standard error.