Non-participation in choice models: hurdle and latent class models

Michael Burton and Dan Rigby

Burton is Associate Professor, Agricultural and Resource Economics, University of Western Australia, Rigby is Senior Lecturer, Economics, School of Social Sciences, University of Manchester.

Address for Correspondence:

Dr Michael Burton
School of Agricultural and Resource Economics
35 Stirling Highway
University of Western Australia
Perth
WA, Australia
6008

Tel: +61 8 6488 2531
Fax: +61 8 6488 1098

email: Michael.burton@uwa.edu.au

Contributed paper prepared for presentation at the International Association of Agricultural Economists Conference, Gold Coast, Australia, August 12-18, 2006

Non-participation in choice models: hurdle and latent class models

Michael Burton and Dan Rigby(

Abstract

In repeated choice modelling studies, it is often the case that individuals always select the status quo option. Although these choices may reflect considered choices, they may also be the result of alternative decisions about whether to participate in the choice process at all. Alternative methods of dealing with this feature of such data are presented, with the implications for estimates of economic values. In particular we consider the alternatives of excluding such individuals from the data, using hurdle models to explicitly model this group, and consider the possibility of latent class models, that endogenously allow for difference preference structures. The application is to a stated preference choice modelling data set that investigates preferences towards forms of GM foods.

JEL: C8, D6, C23

Ψ Rigby is Senior Lecturer, Economics, School of Social Sciences, University of Manchester. Burton is Associate Professor, Agricultural and Resource Economics, University of Western Australia. This paper draws work commissioned by UK Department of Environment Food and Rural Affairs (DEFRA). The views presented in this paper are those of the authors alone and should not be regarded as those of DEFRA or of individuals within DEFRA.

1. Introduction

It is becoming increasingly recognised that considerable heterogeneity exists in consumer attitudes towards GM foods. As a consequence, new statistical techniques are being developed that allow this heterogeneity to be quantified. The particular concern of this paper is the treatment of those individuals who have an absolute objection to the consumption of GM foods. The means by which this extreme preference will express itself will depend upon the context within which choices are being made. In the case of stated preference studies, where respondents are asked to make hypothetical choices, they may refuse to participate at all, or, if the survey design permits, simply refuse to make any choice, or only make choices that avoid GM foods. Typically the analysis of data generated by such surveys assumes that there are continuous utility functions (typically linear) which are underpinning choices. For those that have the extreme objection, that is not the case: they are following lexicographical choices with respect to GM foods. Any analysis that ignores this possibility will be misrepresenting the nature of preferences, and potentially be drawing inappropriate conclusions.

The possibility that these lexicographic preferences are present in a sub-set of the sample is flagged by particular types of behaviours in repeated choice task designs. Thus, if an individual is presented with a large sequence of choices, and always refuses to participate, or always selects options that exclude GM, then one has *prima facia* evidence that the choices being made are driven by some form of lexicographic preferences. The greater the number of occasions when this occurs the more likely it is the case that this behaviour is occurring.

The purpose of this paper is to apply two alternative approaches to incorporating these forms of preferences within the context of data generated from choice experiments. The following section outlines the standard methods of estimating choice models, and then considers the use of hurdle models, and latent class models for this purpose. Section 3 reports the choice experiment and survey design employed in the study. Section 4 reports the results of the alternative models estimated. Section 5 concludes.

2. Assessing preferences for GM foods from choice experiment data

A number of studies of attitudes in the UK and mainland Europe have found that many consumers do not want to eat GM food and that the majority believe that, if such food is sold, it should be clearly labelled (see Consumer Association, 2003; MORI, 2003; Marris *et al*., 2001; various Eurobarometer surveys; the *GM Nation?*

consultation, for example). Of particular interest here is whether there are benefits attached to extending the labelling regime to include not only those foods containing modified genetic material (which we term ***GM Food***) but also those foods with ingredients derived from GMOs, such as maize and soy oil, despite the absence of modified DNA or protein (which we term ***GM Derived Food***).

The context for this is the EU labelling regime, partly an attempt to defuse the GM trade row with the USA and others, which came into effect from April 2004 (Regulations 1829/2003, 1830/2003). A crucial change to the regulatory framework is the extension of the current labelling provisions to genetically modified food or feed, *regardless of whether it contains detectable modified DNA or protein*. Any food or feed which consist of, contain or are produced from GMOs will require a label. For example, this includes tomato paste and ketchup produced from a GM tomato or starch, as well as oil or flour produced from GM maize. This represents a significant change from the requirement before April 2004 which was based on the detectability of genetically modified DNA or protein in the final food product.

Choice Modelling

The central idea behind choice modelling is that individuals can choose between alternative options that contain a number of attributes with different levels. Respondents are not asked to report how much they prefer alternatives, nor even how much they value individual changes in an attribute; they are merely asked to identify which of a number of options they prefer.

Random utility theory proposes that individual consumers choose alternatives that yield the greatest utility and so the probability of selecting an alternative increases as the utility associated with it increases. A person faces a choice among alternatives in choice set j on each of the occasions they make a choice. The utility that respondent n obtains from alternative j in choice situation t is:

$$U_{njt} = \beta' x_{njt} + \varepsilon_{njt} \qquad (1)$$

where x_{njt} is a vector of observed variables and coefficient vector ß, representing peoples' tastes.

The model is implemented by choosing a particular distribution of disturbances. Typically it is assumed that the disturbances are independently and identically distributed, with a Gumbel distribution. This assumption leads to the variant of the logit model used in discrete choice modelling. Hence the probability of person n choosing option j from N options (π_{nj}) can be expressed as:

$$\pi_{nj} = \frac{\exp(\beta' x_{nj})}{\sum_{j=1}^{N} \exp(\beta' x_{nj})} \qquad (2)$$

In this specification the scale factor has been normalised to one and the t subscript for choice situation has been suppressed (see Louviere *et al* (2000), for further detail). In these models 'partworths' or Willingness to Pay (WTPs) are obtained from the ratio of an attribute's marginal utility to the marginal utility of the payment vehicle, i.e. the ratio of coefficients.

Accounting for Preference Heterogeneity

The standard approach to modelling differences in tastes within choice modelling survey data has been to introduce demographic variables which may explain differences in partworths.

An alternative to this approach of essentially identifying a series of point estimates of WTP is to try to identify the distribution from which WTPs are drawn. Models based on this approach are called "mixed logit" models. The mixed (random parameter) logit model is a form of the random utility model in which it is assumed that the functional form and arguments of utility (equation (1)) are common but that the ß parameters vary across individuals. This approach represents a fundamentally different approach to modelling heterogeneity than that employed in more traditional fixed parameter logit models where the approach is to segment the sample, the attributes or both (Hensher and Greene, 2003). The mixed logit model is becoming more commonly applied within the discrete choice literature: see Train, 1998; Revelt and Train, 1998; Train, 1999 for further details and Bonnet and Simioni, 2001; Rigby and Burton, 2005, for relevant applications.

In mixed logit the coefficient vector in equation (1) is allowed to vary among the population with density $f(\beta|\theta^*)$. This vector of coefficients (β_n) can be expressed as the population mean (b) and the individual specific deviation from that mean η_n. Hence the utility that respondent n obtains from alternative j in choice situation t is re-written as:

$$U_{njt} = b' x_{njt} + \eta_n' x_{njt} + \varepsilon_{njt} \qquad (3)$$

Given that the values of β_n are not known, the probability of choosing option i in choice t is the integral of the conditional probability in (3) over all possible values of ß. This integral takes the form:

$$Q_{nit}(\theta^*) = \int L_{nit}(\beta) f(\beta \mid \theta^*) d\beta \qquad (4)$$

Denoting the alternative that person n chose in period t as i(n,t) and assuming that $\beta_n = \beta$, the probability of person n's observed sequence of choices is given by:

$$S_n(\beta) = \prod_t L_{ni(n,t)t}(\beta) \qquad (5)$$

Given that β_n is unobserved, the probability for the sequence of choices is the integral of (5) evaluated over all possible values of ß which depends on the distribution of the ß:

$$P_n(\theta^*)=\int S_n(\beta) f(\beta \mid \theta^*)\ d\beta \qquad (6)$$

The log-likelihood function is LL(θ)= Σ_n lnP_n (θ) which is maximised via simulation in which P_n (θ) is approximated by summing over values of ß generated by Halton draws (see Train, 1999). For a given value of the parameters θ, a value of ß is drawn from its distribution and on the basis of this draw S_n (ß), the product of standard logits, is calculated. The process is repeated (for 125 draws), and the mean of the values of S_n (ß) is interpreted as the estimate of the choice probability:

$$SP_n(\theta)=(1/R)\sum_{r=1,\ldots,R} S_n(\beta^{r|\theta}) \qquad (7)$$

where R is the number of draws of ß, $ß^{r|\theta}$ is the r^{th} draw, and SP_n (θ) is the simulated probability of person n's sequence of choices. The simulated log-likelihood function is SLL(θ) = Σ_n ln(SP_n (θ)) and the estimated parameters are those that maximize the function. A number of alternatives are feasible for the distribution of ß, including normal, log-normal, triangular and uniform.

All models so far have maintained the assumption that there is a conventional utility function underlying the choices made, even if that utility function is individual specific, through the introduction of individual characteristics or random parameters into the model. Hurdle models change that assumption. Instead it is assumed that the sample comprises a mixture of two (or more) types, with different preferences for the good. The difference expresses itself in an extreme form: for some portion of the sample, the good under consideration does not enter into their relevant choice space: they simply do not consider it to be a valid choice of product and are hence non-participants in the market.

Within the context of choice modelling, one has to be clear as to the definition of non-participation. von Haefen et al (2005) use consistent selection of the *status quo* option within a series of choice estimates as an indication that the respondent is a non-participant. Adamowicz et al (1998) suggest that repeated selection of options that have the lowest (or highest) value for a particular attribute may also be an indication of non-participation (i.e. a simple heuristic is being used).

Assuming there are T choice occasions, the probability that an individual will repeatedly select the status quo (option 1) in all of them is given by:

$$\prod_t^T \pi_{n1}^t \qquad (8)$$

The issue is that the observed number of individuals who achieve this is often high, given the number of choice occasions they are facing.

The alternative is to assume that there is an independent probability that an individual is a non-participant, $\bar{\pi}$. The likelihood that an individual will generate a 'non-participation' outcome is now the sum of the two:

$$\bar{\pi} + \prod_{t}^{T} \pi_{n1}^{t} \qquad (9)$$

Thus, although the double hurdle model introduces a separate probability of non-participation, it assumes that it is still possible that some element of non-participation is generated by the conventional discrete choice framework. This introduces the possibility that the observed serial selection of the *status quo* is due to one of two causes: genuine non-participation in the process or genuine selection of serial *status quo* as a corner solution. Empirically it is not possible to differentiate between these two causes from simple observation of the data, but it can be accommodated within the estimation of the model by reforming the likelihood function for the model (von Haefen et al 2005[1]). It is also possible to make the participation probability a (suitably defined) function of attributes, so that the causes of non-participation can be examined. Here we specify it with a probit function. There is no reason why the use of the hurdle approach is limited to the fixed parameter discrete choice model: it is possible to introduce combined mixture models where there is non-participation as well as random parameters associated with the utility function.

The Latent Class Model

An alternative approach which resembles mixed logit but which relaxes the requirement to make specific assumptions about the distributions of parameters across individual consumers is the latent class (LC) model (Boxall and Adamowicz, 2002, Hu *et al*., 2004). In this model consumers are assumed to belong to different segments or classes, each of which is characterised by unique class-specific utility parameters. The choice probability is defined conditional on class probabilities. That is to say, a consumer can probabilistically be assigned to one of several latent classes depending on his or her characteristics and preferences. In the LC model the probability that person n will choose alternative j is defined as follows:

$$\pi_{nj}^{'} = \sum_{s=1}^{S} \pi_{nj \bullet s} \pi_{ns} \qquad (10)$$

where s denotes the number of segments or classes, $\pi_{nj \bullet s}$ is the probability that individual n chooses alternative j conditional on class s, and π_{ns} is the marginal probability that individual n is in class s. The latter can then be expressed as:

[1] Estimation of the double hurdle model employed the GAUSS code provided by Roger von Haefen: assistance in implementing the code is gratefully acknowledged.

$$\pi_{ns} = \frac{\exp(\gamma'_s C_n)}{\sum_{s=1}^{S} \exp(\gamma'_s C_n)} \qquad (11)$$

where γ_s denote a set of class-specific coefficients on concomitant variables C_n.

It is possible to recast the non-participation problem within the form of a latent class model[2] if one assumes that one of the classes is non-participation, and as a result, there are particular expectations of parameter values within the conditional logit for that class. Thus, if a *status quo* alternative specific constant is employed in the specification, then one would expect a high and limiting value for that parameter, ensuring that all those within the class have a high probability of selecting that option throughout the choice sequence, and insignificant parameters on the other attributes. The advantage of the approach is that it does not require a priori imposition of a definition of non-participation, and potentially will allow one to identify a number of limiting behaviours within the data. For example, it may allow one to simultaneously identify sub-populations who select only the status quo, avoid GM, or only use price as a basis for choice

The following section outlines the CM survey which will be used as the basis for exploring both latent class and hurdle models of participation.

3 The Survey and Study Design

The survey was conducted in England, Wales and Scotland between July and September 2003. The sample was defined as men and women, aged 16 and over who was the main shopper for their household. The choice set attributes and levels were finalised following a series of semi-structured interviews and a number of pilot interviews. The bread attributes and levels designed for the analysis of regulations concerning GM, GM-Derived and Non-GM ingredients are reported in Table 1a. An Example choice set is shown in Table 1b. A sample comprising 608 respondents was achieved. Personal interviews were conducted in the home using computer aided personal interviews. Each respondent was required to answer 4 choice modelling sets, with the first option in each identified as the status quo i.e. their current bread. The use of a status quo option allows participants who are set against GM foods a means to express their concern through serial selection of this option. A considerable proportion of the 608 respondents did this, with 274 (or 45%) always selecting this option in the 4 choice sets presented to them. However, the inclusion of non-gm breads under options 2 and 3 introduces an alternative measure of non-participation: the possibility of dismissing any option that contains GM, but making choices across non-GM breads. 157 respondents (26%) were in this category: they made some non-status quo selections, but never selected GM bread. The remaining portion of the sample made some selection involving GM bread across the 4 choice sets

[2] The authors would like to acknowledge Vic Adamowicz for this suggestion.

presented. It is to the implications and decomposition of these behaviours that we now turn.

4. Results

Rigby et al (2004) report extensive model analysis using this data set. Here we use a restricted specification that captures the most important aspects of the model. The 4 attributes of the choice sets are included, with price, shelf life and fibre content included as continuous variables. The nature of the GM technology used is included as 2 dummy variables, for GM derived and GM, using non-GM as the baseline. Gender and age effects are included as GM interaction terms in the hurdle models. An alternative specific constant is included for the status quo option, and given the potential importance of this variable in explaining non-participation, it is specified as a random variable, following a normal distribution. Non-participation is defined as selection of the status quo in all 4 choice sets.

Table 2 reports estimates of 3 alternative choice models. Model 1 reports results for the subset of the data that excludes those identified as non-participants. This is the method most often used to accommodate non-participation, but it assumes seperability i.e. that there is no loss in efficiency by imposing an *a priori* segregation of the data. As von Haefen et al (2004) note, this will deliver the true parameters only if one can accommodate for truncation of the data i.e. the impossibility of repeated selection of the *status quo*. The model identifies significant price and GM effects, with increasing concern about GM as people get older (although the quadratic term implies this increasing disutility reaches a turning point at approximately age 50). There is no significant impact of gender on preferences for GM derived food, and only at the 10% level for GM food with women's utility reduced more than men's by the presence of GM ingredients in their bread.. The 'status quo' *ASC* is estimated to have a negative mean but there is a substantial variance around this mean.

Model 2 reports an identical model, but estimated over the full data set, including those identified as non-participants. Although direct comparison of parameters across estimated models is strictly inappropriate, because of the confounding impact of the scale parameter, casual inspection of the two results indicates that they are very similar in their estimates, apart from the ASC. As one may expect, the ASC has become positive, reflecting the substantial portion of the sample who always select this option, and there is a highly significant, and large, standard deviation for the distribution of the ASC, suggesting a greater variability in the sample, again, which one would expect. Model 3 reports the estimates from a double hurdle model applied to the full data set. Three variables are used to explain non-participation: age, being a member of social class E ('state pensioners or widows, casual or lowest grade workers'), and their degree of concern about the use of GM in food. This is measured through a composite, normalised attitudinal variable ('GM*trust*'), derived using PCA from Likert scale responses to a series of statements regarding food biotechnology[3].

As one might expect, the latter is significant, and implies that those who are more concerned about GM are more likely to be among the non-participant group. Within the conditional logit component of the DH model, the parameters are very similar to those estimated before. Figure 1 reports a scatter graph of the parameters from the three models against each other. There is a strong linear relationship between them, apart from the *ASC* coefficients, suggesting that the treatment of non-participation in this case is not affecting the estimates of relative weight of the attributes within the utility function.

Table 3 reports the results for the latent class model. Note that within this model no *a priori* assumption is made about participation or non-participation. Because of difficulties in achieving convergence the demographic variables are excluded from the conditional logit model, but are included in the model explaining class membership probabilities.

For Class 1 the parameter on the status quo variable is estimated to be positive and those for the GM and GM Derived terms variables are very large and negative, which for any values of the other attributes, lead to an extremely high probability that members of this class select the status quo option. Class 2 has very strong negative coefficients for the GM variables, implying that within this class, the presence of GM elements is strongly objected to, but the other attributes are significant apart from the status quo variable, which is insignificant. This suggests that this is a class of individuals who are averse to GM, but are prepared to make choices across the other attributes of the choice sets. Class 3 has an insignificant GM derived coefficient, but a significant and negative effect of GM ingredients. The status quo ASC is negative and significant which is difficult to interpret, but the other attributes are as expected *a priori*. This suggests that for this group the presence or otherwise of GM Derived ingredients in the bread is irrelevant, and they are prepared to evaluate the breads on the basis of the more conventional attributes. However, they are averse to the presence of GM ingredients, but not to the extent that its presence is a limiting factor in choices (see discussion of partworths below).

In the "Restricted Model" shown in Table 3 restrictions on coefficients are tested. Specifically, we test whether it is possible to restrict the utility function of Class 1 to

---

[3] The PCA loadings were strongly positive on "GM food is safe to eat" "GM crops will help developing countries", and "I trust the government when it comes to food safety issues", and strongly negative on "Growing GM crops will permanently damage the environment", and "Multinational companies will benefit most from genetic modification".

one comprising just SQ and GM effects, removing the role of Price, Fibre and Shelflife in determining choices. This restriction is accepted[4] (test statistic of 2.4, $\chi^2_{0.05,3}$ = 7.82). The parameter estimates for the utility functions for Classes 2 and 3 and the class membership terms are stable across both model specifications, indicating the restrictions have not been accepted by radically restructuring the nature of the classes nor the determinants of class membership.

The determinants of the class membership have a number of significant coefficients. The PCA composite attitudinal score *GMtrust* was a significant determinant of class membership,, as were 2 additional PCA attitudinal scores '*Green*' and '*Fussy*'. The '*Green*' score is based on Likert responses to the statements: "I try to avoid artificial ingredients", "I try to recycle as much waste as possible", "When I have the choice I always buy organic", "I try to buy environmentally friendly products" and " When I have the choice, I always try and buy ethically responsible products (e.g. Fair Trade)"). The PCA composite 'Fussy' was derived from responses to the following statements: 'Food should be clearly labelled to say if it contains genetically modified (GM) ingredients', 'I read ingredients labels on food items very carefully', 'I consider myself to be knowledgeable about food safety issues', 'I don't mind paying extra for quality food'.

Those who are more trustful of GM food and its regulation are less likely to be members of Classes 1 and 2, compared to Class 3. A positive score on the Green PCA score increased the chance of being in Class 2 and also, but to a lesser extent, Class1 relative to Class 3. The effect of higher Fussy scores was positive and equivalent for both classes 1 and 2.

Age also plays a role in explaining class membership probability. As age increased so did the probability of being in the non-participation Class 1, while there was a marginally significant negative effect on the probability of being in Class 2.

The results from the models are summarised in Table 4 which reports the ***conditional partworths*** from the latent class models, estimated as conditional upon being in a particular class. Although it is more usual to report distributions of unconditional partworths from latent class models, in the current case this is infeasible, as in Model 5 the utility function for Class 1 has a zero coefficient on price, which generates irrelevant estimates of partworths. Note that these partworths are in % terms, where 100% represents the cost of a loaf of bread which in this sample averaged close to 1€. These reveal that those in Class 2 are strongly averse to the GM aspects of the bread, but are relatively unresponsive to changes in the levels of the other attributes. This contracts with member of Class 3, which appears to be

---

[4] We thank William Greene for modifying the NLOGIT code to allow these restrictions to be imposed.

indifferent towards GM Derived inputs, concerned about GM but less so than the other classes, hence the WTP of €0.21 to avoid GM bread. The valuation of changes in the Fibre and Shelflife attributes for Class 3 is also higher than the other 2 segments of the market.

5. Discussion and Conclusion

This paper has explored two alternative methods of accounting for non-participation in choice experiments. In the first, a double hurdle model was used, with non-participation defined as serial selection of the status quo option. Although the double hurdle model was a significant improvement over a standard conditional logit model it appeared to change little in the way of relative weights of the attributes, apart from the alternative specific constant for the status quo. This suggests that there is relatively little lost in terms of understanding behaviour of those making choices from taking the conventional approach to non-participation, which is to exclude them from the data set.

The use of the latent class model to model non-participation is more speculative. It does not require the a priori definition of non-participation, but allows the data to reveal such behaviours. Here, based on an understanding of the data sets, we have imposed 3 classes, as we believed there may be 2 forms of non-participation present in the sample: selection of the status quo, and avoidance of GM. The results supported this split, in so far as the estimated results for each of the classes can be interpreted in this light. The ability to restrict parameters within the conditional logit element of the latent class model is particularly useful, as it allows one to test prior hypotheses about the redundancy of some attributes for some classes. An issue not explored here is the impact of extending the number of classes. However, the approach appears to be of some merit. An associated area that could be explored is extending the double hurdle model to include a number of alternative forms of participation.

References

Adamowicz, W.L., P.C.Boxal, M.Williams and J.J.Louviere (1998) Stated preference approaches for measuring passive use values: choice experiments and contingent valuation American Journal of Agricultural Economics 80: 64-75

Bonnet, C., and M. Simioni (2001), Assessing Consumer Response to Protected Designation of Origin Labeling: A Mixed Multinomial Logit Approach, *European Review of Agricultural Economics*, 28: 433-449

Boxall, P.C. and W.L. Adamowicz (2002) Understanding heterogeneous preferences In random utility models: a latent class approach. *Environment and Resource Economics* 23: 421-446.

Consumers Association (2002) GM Dilemmas - Consumers and Genetically Modified Foods. Policy Report http://www.which.net/campaigns/food/gm/misc/gm_report.pdf

von Haefen, Roger H., D. Matthew Massey, & Wiktor Adamowicz. "Serial Non-Participation in Repeated Discrete Choice Models," *American Journal of Agricultural Economics*. vol. 87(4) 1061-1076

Hensher, D.A. and W.H. Greene (2003) The Mixed Logit: the State of Practice, *Transportation* 30(2) 133-176.

Hu, Wuyang, A. Hunnemeyer, M. Veeman, W. Adamowicz, and L. Srivastava (2004) Trading off Health, Environmental and Genetic Modification Attributes in Food. *European Review of Agricultural Economics*, 31 (3): 389-408.

Louviere, J. J., D. A. Hensher and J. Swait (2000). *Stated Choice Methods: Analysis and Application*, Cambridge University Press

Marris, C, Wynne, B, Simmons, P and S.Weldon (2001) *Public Perceptions of Agricultural Biotechnologies in Europe.* Final Report of the PABE research project. http://www.pabe.net

MORI (2003) "GM Food Opposition Continues" 3 July 2003
http://www.mori.com/polls/2003/gmfood.shtml

Revelt, D. and K. Train (1998), 'Mixed logit with repeated choices', *Review of Economics and Statistics* (80): 647–657.

Rigby, D. and Burton,M.P. (2005) Capturing preference heterogeneity in stated choice models: a random parameter logit model *European Review of Agricultural Economics* 32(2) 269-288.

Rigby, D., T. Young and M. Burton (2004) *Consumer Willingness to Pay to Reduce GMOs in Food and Increase the Robustness of GM Labelling*. Report to the Department of Environment, Food and Rural Affairs (DEFRA).

Train, K (1998) Recreation Demand Models with Taste Differences Over People. *Land Economics*, Vol. 74(2): 230-239.

Train, K (1999) *Halton Sequences for Mixed Logit.* Working paper, University of California, Department of Economics.

Train, K. and G. Sonnier (2003) Mixed Logit with Bounded Distributions of Partworths. Working Paper, Dept. of Economics, University of California, Berkeley.

Table 1a Attributes and attribute levels in the CM design.

Attribute	**Levels**
Price (%)	-67, -50, -33, -17, Usual, +17, +33
GM Type	Non-GM, GM-Derived, GM
Shelflife	Usual, Usual + 1 day, Usual + 2 days, Usual + 3 days
Fibre Content	Usual, Usual + 10%, Usual + 30%, Usual + 50%

Table 1b. An Example Choice Set

	Bread 1	**Bread 2**	**Bread 3**
	Usual brand	**Usual brand - alternative option 2**	**Usual brand - alternative option 3**
Price	Usual	Usual	Usual -50%
GM Type	Non-GM	GM-Derived	GM
Shelflife	Usual shelflife	Usual shelflife	Usual +2 days
Fibre Content	Usual fibre content	Usual +30%	Usual +10%
Which bread do you prefer ? (tick one ✓)	□	□	□

Table 2 Estimates of the Mixed Logit and Hurdle Models

	Model 1 Mixed Logit: restricted data set		Model 2 Mixed Logit: full data set		Model 3 Double hurdle	
	parameter	*s.e.*	parameter	*s.e.*	parameter	*s.e.*
Hurdle						
const					-0.8691	*0.1958*
gmtrust					-0.18	*0.0623*
age					0.0107	*0.0035*
soc class E					0.3201	*0.1898*
Discrete choice						
price	-0.013	*0.0019*	-0.0138	*0.002*	-0.0134	*0.002*
SQ	-0.4339	*0.1142*	1.7646	*0.201*	-0.1281	*0.1768*
GM derived	3.0613	*0.9869*	3.5936	*1.0507*	3.3574	*1.0263*
GM	1.0782	*0.9526*	1.4229	*0.996*	1.1755	*0.9895*
GM derived *fem	-0.3349	*0.2785*	-0.4009	*0.3032*	-0.3667	*0.2939*
GM* fem	-0.4374	*0.261*	-0.4397	*0.2698*	-0.4284	*0.2702*
GM derived* age	-1.6925	*0.4023*	-1.8609	*0.4281*	-1.8068	*0.4192*
GM*age	-0.8135	*0.3628*	-0.9274	*0.3762*	-0.8587	*0.3751*
GM * age^2	0.1619	*0.0398*	0.174	*0.0424*	0.1724	*0.0416*
GM derived* age^2	0.0848	*0.0349*	0.0927	*0.0358*	0.0889	*0.0359*
shelf	0.0481	*0.0414*	0.0526	*0.0421*	0.0508	*0.0419*
fibre	0.0041	*0.0028*	0.0037	*0.0029*	0.0038	*0.0029*
SQ sd	0.9246	*0.0969*	3.3165	*0.2077*	1.5967	0.2246

NB: restricted data set is for only those who are defined as participants. Full data set includes all individuals, both participants and non-participants.

Table 3 Parameter estimates for the latent class model

	Model 4: Unrestricted model		Model 5: Restricted model	
	parameter	*s.e.*	parameter	*s.e.*
	Utility parameters in latent class 1			
SQ	2.213	*0.544*	3.083	*0.220*
Price	0.013	*0.010*		
GM	-2.894	*0.838*	-2.695	*0.828*
GM derived	-2.252	*0.684*	-2.084	*0.619*
Shelf	-0.113	*0.222*		
Fibre	-0.020	*0.012*		
	Utility parameters in latent class 2			
SQ	0.061	*0.284*	0.021	*0.289*
Price	-0.038	*0.006*	-0.039	*0.006*
GM	-5.773	*0.527*	-5.863	*0.542*
GM derived	-5.334	*0.441*	-5.434	*0.455*
Shelf	0.203	*0.098*	0.216	*0.100*
Fibre	0.013	*0.006*	0.012	*0.006*
	Utility parameters in latent class 3			
SQ	-0.614	*0.147*	-0.619	*0.146*
Price	-0.013	*0.002*	-0.012	*0.002*
GM	-0.250	*0.113*	-0.249	*0.113*
GM derived	0.061	*0.113*	0.049	*0.113*
Shelf	0.154	*0.038*	0.152	*0.038*
Fibre	0.010	*0.003*	0.010	*0.003*
	Coefficients on class model 1			
Constant	0.387	*0.338*	0.396	*0.337*
GM concern	-0.824	*0.137*	-0.823	*0.137*
Green	0.187	*0.128*	0.184	*0.128*
Fussy	0.264	*0.122*	0.266	*0.122*
Age	0.012	*0.006*	0.012	*0.006*
	Coefficients on class model 2			
Constant	0.599	*0.439*	0.590	*0.442*
GM concern	-1.132	*0.176*	-1.138	*0.177*
Green	0.373	*0.183*	0.377	*0.185*
Fussy	0.253	*0.171*	0.254	*0.172*
Age	-0.013	*0.009*	-0.014	*0.009*
LL value	-1535.1		-1536.8	

Table 4 Conditional WTP for unit changes in attributes estimated from the latent class models: % change in price of bread (100% = 1€).

	Unrestricted model			Restricted model		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
SQ	-170.2	1.6	-47.2	N/a	0.5	-51.6
GM	222.6	-151.9	-19.2	N/a	-150.3	-20.8
GM derived	173.2	-140.4	4.7	N/a	-139.3	4.1
Shelf	8.7	5.3	11.8	N/a	5.5	12.7
Fibre	1.5	0.3	0.8	N/a	0.3	0.8

Figure 1. Scatter plot of attribute parameters from models in Table 2. Double hurdle parameters on the x axis, restricted and unrestricted CL parameter estimates on y axis. SQ coefficients (mean and SD) marked with triangles.