MODEL SELECTION IN THE BAYESIAN MIXED LOGIT: MISREPORTING OR HETEROGENEOUS PREFERENCES?

Kelvin Balcombe(1), Ali Chalak(2), and Iain Fraser(2)

(1) - Agricultural and Food Economics, University of Reading

and

(2) - Kent Business School, University of Kent

August 2007

Address for Correspondence:

Iain Fraser

Kent Business School

University of Kent

Wye Campus

Wye, Kent, UK

TN25 5AH

i.m.fraser@kent.ac.uk

MODEL SELECTION IN THE BAYESIAN MIXED LOGIT: MISREPORTING OR HETEROGENEOUS PREFERENCES?

Abstract

The Bayesian Mixed Logit model is estimated in both 'preference space' and 'willingness-to-pay space' incorporating a number of distributions for the random parameters, models that contain constant and random parameters, and misreporting. We calculate the marginal likelihood for the Mixed Logit, which is required for Bayesian model comparison and testing, but which has so far received little consideration within the Mixed Logit literature. We use this model to estimate willingness-to-pay (WTP) to consume bread which has been produced with reduced levels of various pesticides so as to protect biodiversity. We find some support for estimation in preference space, and our results indicate strong support for the Mixed Logit compared to the fixed parameter Logit. Furthermore, although the Logit model identifies misreporting of preferences, misreporting disappears once heterogeneity is incorporated into the model. As a result we conclude that with this data set we find support for preference heterogeneity as opposed to misreporting.

Key Words: Mixed Logit, Willingness-to-Pay, Model Comparison

JEL: C11, C25, C52, L92, Q51

1.Introduction

There has been a rapid adoption and implementation of the Mixed Logit (ML) model in discrete choice analysis (e.g., Revelt and Train, 1998, Train, 1998, McFadden and Train, 2000, Layton and Levine, 2003, 2005, Train, 2003, Scarpa and Alberini, 2005, and Smith, 2005). The attraction of the ML model stems from the flexibility it provides in terms of approximating any random utility choice problem (Train and Sonnier, 2005). In this paper we examine several important generalisations of the ML model using Bayesian methods.

The attractiveness of employing Bayesian methods to estimate the ML model has been noted by Huber and Train (2001) and Hensher and Greene (2003). First, the likelihood function of the ML may be multi-modal and inference based on analysis at one mode, and curvature of the likelihood at that mode may be misleading. Second, a Bayesian approach can incorporate prior knowledge about the experimental design in a way that cannot be achieved by using Classical methods (Ferrini and Scarpa, 2007). Ruud (1996) and Train and Weeks (2005) note that sometimes the ML may be near non-identified or weakly identified in certain regions of the parameter space. The existence of such regions may stall a Classical optimisation algorithm and impair inference that is solely based on the curvature of the likelihood at a given point. However, the fact that such regions exist, need not imply that the data is non-informative about the parameters. A prior that does not dominate the data, can never the less prove important since it can prohibit, or limit the propensity of, parameters 'wandering' into near non-identified regions. By contrast, Classical estimation of the ML can prove difficult and infeasible where the likelihood is not well behaved in this respect. Third, the calculation of the marginal likelihood enables model comparison and testing in a way that is more general than can be achieved using

Classical methods. Within the ML literature, much attention has been paid to the relative performance of non-nested alternatives. Whether models are nested or non-nested has no implications for the use of the marginal likelihood in Bayesian inference.

In spite of these potential advantages, Bayesian estimation of the ML has not been extensively exploited. By contrast, Classical estimation of the ML is common (i.e., Hensher and Greene, 2003 and Train, 2003). There are several reasons for this situation. First, in contrast with classical software, publically available Bayesian software has not easily allowed the utility coefficients to be conditioned on variables that explain individuals' preferences. This is in spite of the fact that this approach has been developed and implemented in the work of Rossi et al. (1996) and Allenby and Rossi (1999). Second, whereas Classical methods have allowed for the mixture of fixed and random coefficients, this has not been facilitated in Bayesian procedures. Moreover, proper model comparison within a Bayesian framework requires the computation of the marginal likelihood. Although relatively simple in principle, practical implementation is difficult. The research presented in this paper addresses these, and other, issues.

Specifically, our work integrates four key components. First, unlike most previous Bayesian applications using the ML, we allow parameters (some or all) to be fixed (i.e., have no unobserved heterogeneity). At first glance this may seem trivial as, in a sense, it is just a special case of the ML. However, implementing this procedure requires additional Metropolis-Hastings (MH) steps so as to estimate the fixed parameters. As is well recognised in the literature (e.g., Louviere, 2006), the division of utility coefficients by a numeraire coefficient (usually the payment attribute coefficient) is the primary source of instability for the WTP estimates derived from choice models estimated in preference space. When the payment coefficients are random (have unobserved heterogeneity), the moments of WTP ratio do not exist. Fixing the parameter offers one potential remedy to this problem, and further justifications are discussed in Train (2003, p.311). Therefore, the payment coefficient in choice models has often been fixed in applied Classical studies and it is worth having an equivalent facility when using a Bayesian approach. We would argue, however, that this restriction is somewhat adhoc. Its imposition has been justified on the grounds that it stablises WTP estimates, rather than through sound theoretical arguments. For this reason, we regard it as important to test this restriction, rather than impose it *a priori*. In addition, by having general algorithms which allow any or all of the parameters to be fixed or random, this also allows us to assess the support for the fixed parameter logit relative to the ML providing the marginal likelihoods for the models are calculable (see below).

Second, as suggested by the Classical contributions of Cameron and James (1987), Cameron (1988) and Train and Weeks (2005), we estimate the ML in WTP space as well as the conventional preference space. In essence, WTP space estimation specifies different distributions for the marginal rates of substitution compared to those estimated in preference space. As already noted, the instability of the WTP estimates is especially acute where the parameter of the payment attribute is variable and is not bounded above zero. Train and Sonnier (2005) explore transformations of normals, partly because this then presents the opportunity to bound the resulting WTPs. However, transformations such as the exponential (so that the distribution of the utility coefficient is log-normal) are problematic in theory, not least because the model will depend on how the attribute variable is scaled, as well as in practice. For this reason, WTP space estimation is attractive on a practical level, since it avoids this problematic *ex post* estimation problem.

Third, the estimation of the marginal likelihood is performed using Halton sequences and the support for different specifications can, therefore, be measured. In particular, we compare models with alternative distributions for the utility parameters, including the case where each or all are constant. This enables a test for the dominance of the ML over its fixed parameter version, a test that has not been previously conducted. In addition, we assess the support for the (special) case where only the price coefficient is held constant, since this may greatly decrease the variance of WTP estimates, without being overly restrictive. The support for models estimated in WTP space is also assessed. To date, no attempt has been made within the Bayesian literature to test whether this approach is more compatible with data, relative to the preference space approach. By employing Halton sequences, the marginal likelihood is simulated and a number of models, both with and without misreporting, are compared. The numerical accuracy of the marginal likelihood is measured by employing a stationary bootstrap (described in Li and Maddala, 1997) once the sequence of likelihoods has been generated.

Fourth, we generalise the ML model to allow for the possibility of misreporting as defined and analysed Balcombe *et al.* (2007). They examined issues of misreporting in discrete choice models (i.e., Conditional Logit) in a way that is similar to the 'misclassification' approach outlined in Hausman *et al.* (1998) and Caudhill and Mixon (2005). Implicitly, the ML model assumes that all respondents reply in accordance with how they would behave in a revealed preference study. However, it is possible that in many stated preference surveys, respondents express preferences that are not in accordance with their 'real' preferences. Potential reasons include be strategic responses, 'yeh saying' or cognitive limitations on behalf of respondents (i.e. either they do not understand, or have not got the time to understand the choices) and associated effects induced by the way that questions are framed. The misreporting approach attributes each respondent with a 'true' utility function, but posits that respondents may (with some probability) answer in a way that is completely independent of these preferences. Whereas a Gumbel error in a utility function allows for divergent choices between individuals with similar utility functions, it does not actually allow for the possibility that agents have very strong preferences, yet ignore these preferences when responding to a stated preference survey. The description 'misreporting' accurately portrays the underlying idea that the utility function exists and the reporting is not always in accordance with these utilities. Balcombe et al. (2007) found strong evidence of misreporting using a fixed parameter model. However, is the finding of misreporting in a Logit, simply a consequence of ignoring heterogeneity in preferences? Or, alternatively, how much preference heterogeneity identified by a ML model is attributable to preferences and how much might be a function of misreporting?

Taking these various components together, the analysis in this paper extends the Bayesian ML models employed in the literature to date. , We also note that as in Rossi et al. (1996), the framework and models we develop in this paper enable the parameters within an agents' utility function to be dependent on the agents' characteristics. However, in the empirical example we present, we do not condition the parameters of individuals' characteristics. Nevertheless, it is worth noting that the framework is general in this regard.

To demonstrate the utility of the methods developed in this paper we employ a new data set derived from a Choice Experiment (CE) undertaken to examine and estimate consumer WTP to consume food (i.e., bread) produced using wheat grown using reduced levels of pesticides (Chalak *et al*, 2006). The motivation for this CE stems from the impact on the landscape and its associated biodiversity from the use of pesticides to allow intensification of agriculture. For example, bird populations have been affected by pesticide use, especially insecticides and herbicides, because they kill invertebrate prey and insect host plants and as a result impact the quantity and quality of feed available. This research and the results we present add to a small number of stated preference studies undertaken with respect to pesticide use (e.g., Foster and Mourato, 2000, 2002, Hamilton et al., 2005, Canavari et al, 2005, Florax et al., 2005, and Balcombe et al., 2007).

2. The Model

2.1. Notation and Model Specification

$x_{j,s,n}$ denotes the $k \times 1$ vector of attributes presented to the *jth* individual $(j = 1, ..., J)$ in the sth option $(s = 1, ..., S)$ of the nth choice set $(n = 1, ..., N)$. $U_{j,s,n}$ is the utility that the jth individual derives from $x_{j,s,n}$ and in accordance with notational conventions in the literature $y_{j,s,n}$ denotes an indicator variable that is 1 if the jth individual indicates that they would choose the *sth* option within the *nth* choice set, and 0 if they would not. It is stressed that $y_{j,s,n}$ denotes only whether an individual indicates that they would choose a particular option, not that they necessarily prefer that option. Whereas in the standard framework choices are dependent only on the relative utility derived from choice attributes, with potential misreporting there is the chance that utilities and choices will diverge.

The notation $f(x)$ and $f(x|.)$ are used to denote density and conditional density functions that take an unspecified form, and $F(x)$ and $F(x|.)$ the associated cumulative distributions. The notation $f_N(x|\mu, \Omega)$ denotes that a random vector x has a normal distribution with mean μ and variance Ω, and $f_{IW}(x|T_0, v_0)$ denotes that x has an inverse Wishart distribution with the parameters T_0, and v_0 and $f_U(a, b)$ will denote the uniform distribution over the interval (a,b).

The utility that the jth individual receives from the sth choice in the nth choice set is assumed to be linear, except that the parameters may be transformed. Consequently, the utility function is of the form

$$U_{j,s,n} = x'_{j,s,n} g(\beta_j) + e_{s,j,n} \tag{1}$$

where β_j is a $(k \times 1)$ vector describing the preferences of the jth individual and $g(.)$ is some transformation of the parameters, from and to the space of k vectors.

Without loss of generality, we will assume that the parameters β_j are ordered so that they may contain fixed parameters c_j in the first block, and random parameters b'_j in the second.

$$\beta'_j = (c'_j, b'_j) \tag{2}$$

Both sets can be conditioned on variables describing the characteristics of the *jth* individual. Preferences may therefore be determined by a vector z_j, a $(h \times 1)$ column vector of variables describing the characteristics of the jth individual (h being *1* and z'_j being *1,* for all j, if there are no characteristics). More specifically, defining $Z_j = I_k \otimes z_j$ the components of β'_j are defined as:

$$\begin{aligned} c_j &= Z'_j \alpha_c \\ b_j &= Z'_j \alpha_b + u_j \end{aligned} \tag{3}$$

and u_j is a independently and identitically normally distributed vector with variance covariance matrix Ω. The errors $\{u_j\}$ are assumed to be uncorrelated across individuals.

The function $g(.)$ may take any of the transformation of the normal distribution discussed in Train and Sonnier (2005). In considering estimation in WTP space, we also use reparameterisations of the form

$$g(\beta_j) = g_1\left(\beta_{1j}\right)\left(1, g_2\left(\beta_{2j}\right), ..., g_k\left(\beta_{kj}\right)\right)' \tag{4}$$

in which case the quantities $g_2\left(\beta_{2j}\right), ..., g_k\left(\beta_{kj}\right)$ are marginal rates of substitution with the numeraire element of attribute vector (the first element in the case above). If such a transformation is used then we say that estimation is taking place in WTP space. Otherwise, estimation is being performed in 'preference space'. The error $e_{s,j,n}$ is 'extreme value' (Gumbel) distributed, is independent of $x'_{s,j,n}$ and is also uncorrelated across individuals or across choices.

The set of all stated choices by respondents is $\mathbf{Y} = \{y_{j,s,n}\}_{j,s,n}$. The set of characteristics describing all respondents is $\mathbf{Z} = \{z_j\}_j$. The set of options given to the jth individual is $\mathbf{X}_j = \{x_{j,s,n}\}_{s,n}$ and the set of all option sets given to all respondents is $\mathbf{X} = \{\mathbf{X}_j\}_j$. The data $\mathbf{D}$ is, therefore, the collection $\mathbf{D} = \{\mathbf{Y}, \mathbf{Z}, \mathbf{X}\}$.

Faced with a set of choices, the jth individual will prefer $x_{s_k,n}$ providing that $U_{j,s_k,n} > U_{j,s_q,n}$ for all k not equal to q. The model in this paper extends existing specifications so that each respondent has a probability (π) of misreporting, along with a probability (λ_s) that misreporting (should it occur) will be in favour of option s (where $\sum \lambda_s = 1$).The parameters related to misreporting (described in more detail below) will be denoted as $\Lambda = (\pi, \lambda_1, ...\lambda_{S-1})$.Therefore, the collection of all parameters describing the model will be denoted as $\Theta = (\alpha, \Omega, \Lambda)$. In what follows, the set $\{b_j\}_j$ will be denoted as $\mathbf{B}$ and we will refer to $\mathbf{B}$ as 'latent data'. Finally, for notational convenience the multiple integral $\int_{\beta_n} ... \int_{\beta_1} db_1.....db_n$ is expressed as $\int_{\mathbf{B}} d\mathbf{B}$. This integral is a definite integral, and, therefore, it is implicitly assumed that this integral is over a specified set for $\mathbf{B}$.

2.2. Misreporting.

Under the assumptions above, conditionally on the parameters β_j, the probability of the jth respondent preferring the sth option is logistic:

$$\Pr\left(y_{j,s,n} = 1|\beta_j, \mathbf{X}_j\right) = \dot{p}_{s,j,n} = \frac{e^{x'_{j,s,n}g(\beta_j)}}{\sum_{s=1}^{S} e^{x'_{j,s,n}g(\beta_j)}}. \tag{5}$$

The misreporting approach developed in Balcombe *et al.* (2007) attributes a probability to a respondent correctly responding to a given choice as π. Extending this approach to the case where respondents are presented with multiple choice sets, denote the indicator variable $v_{j,n} = 1$ if individual j correctly reports their preferred choice within the nth choice set and zero otherwise. Defining the probability of correct reporting as $\Pr(v_{j,n} = 1) = \pi$, the probability of the jth individual choosing the sth option in the nth choice set is:

$$\Pr\left(y_{j,s,n} = 1\right) = \Pr\left(y_{j,s,n} = 1|v_{j,n} = 1\right)\pi + \Pr\left(y_{j,s,n} = 1|v_{j,n} = 0\right)(1-\pi). \tag{6}$$

By assigning a probability to the event that the jth individual will mis-report in favour of the sth option (given that they mis-report), $\Pr(y_{j,s,n} = 1|v_{j,n} = 0) = \lambda_s$ where $\sum_{s=1}^{S}\lambda_s = 1$. The probability that $y_{s,j,n} = 1$ becomes

$$p_{j,s,n} = \pi\dot{p}_{j,s,n} + (1-\pi)\lambda_s. \tag{7}$$

This model could be extended by allowing for mis-reporting probabilities to vary over choice sets. However, this rapidly leads to over-parameterised models where individuals are given a high number of choice sets.

2.3. Priors

Bayesian estimation requires priors for the parameters α and Ω and Λ. These are specified as:

$$\left(\alpha'_c, \alpha'_b\right)' = \alpha \sim f_N\left(\alpha|\mu, A_0\right) \tag{8}$$

where A_0 is a diagonal matrix . If there are fixed and random elements, then the associated means for these parameters will be denoted as α_c and α_b respectively, with corresponding means μ_c and μ_b. Likewise, A_0 contains the diagonal blocks $A_{0,b}$ and $A_{0,c}$. The prior for the covariance matrix of the random parameters is:

$$\Omega \sim f_{IW}\left(\Omega|T_0, v_0\right) \tag{9}$$

The 'hyper parameters' μ, A_0, T_0, v_0 are set *a priori.* The misreporting parameters are assumed to have a uniform prior, subject to inequality constraints:

$$\begin{aligned} \pi &\sim f_U\left(0,1\right) \qquad (10) \\ \left(\lambda_1,...,\lambda_{S-1}\right) &\sim f_U\left(0,1\right)^{S-1} \times I\left(\sum_{s=1}^{S-1} \lambda_s \leq 1\right). \end{aligned}$$

where $I\left(\sum_{s=1}^{S-1} \lambda_s \leq 1\right)$ is equal 1 if the constraint is obeyed and zero otherwise. The integrating constant of this distribution is $1/\left(S-1\right)!$. Together the set of the priors above is denoted as $P\left(\Theta\right)$, and the priors on Λ only, as $P\left(\Lambda\right)$.

2.4. Full data Likelihood, the Likelihood and Marginal Likelihood

The full-data (or complete) likelihood function is the likelihood expressed in terms of the parameters and latent data (in this case $\mathbf{B}$). So in the case of the ML with misreporting, the full-data likelihood function is:

$$L_f\left(\mathbf{B},\Theta,\mathbf{D}\right) = \prod_j \left(\prod_s \prod_n p_{j,s,n}^{y_{j,s,n}}\right) f\left(\mathbf{B}|\Omega, \alpha_b, \mathbf{Z}\right) \tag{11}$$

where $p_{j,s,n}$ is defined in [7]. Integrating out the latent data gives the likelihood:

$$L\left(\Theta,\mathbf{D}\right) = \int_{\mathbf{B}} \prod_j \left(\prod_s \prod_n p_{j,s,n}^{y_{j,s,n}}\right) dF\left(\mathbf{B}|\Omega, \alpha_b, \mathbf{Z}\right) \tag{12}$$

In the absence of latent data we could simply write $L_f\left(\{\},\Theta,\mathbf{D}\right) = L\left(\Theta,\mathbf{D}\right)$. $L\left(\Theta,\mathbf{D}\right)$ is the likelihood of the model. It is this quantity that is maximised in classical estimation, and usually calculated by simulation when there are latent variables. The marginal likelihood, given priors on the parameters $P(\Theta)$, is:

$$\mathcal{M}(\mathbf{D}) = \int_{\Theta} L(\Theta, \mathbf{D})\ P(\boldsymbol{\Theta})\ d\Theta. \tag{13}$$

The marginal likelihood can be used to calculate a 'Bayes factor'. In effect, a larger marginal likelihood indicates greater support for a particular model (Koop, 2003).

2.5. The Form of Conditional Posterior Distributions

Since b_j are normally distributed around α_b and Ω , the conditional posteriors for α_b and Ω (given $\mathbf{B}$) are the same as for the case of the normal linear regression with independent Normal and Wishart priors. The two partitioned sets of parameters have independent priors, and are assumed to be independent. Therefore, for the parameters with unknown sources of variation the conditional posterior is:

$$f(\alpha_b|\mathbf{B},\Omega, Z) = f_N(\hat{\alpha}_b, \Gamma) \tag{14}$$

where

$$\begin{aligned} \hat{\alpha}_b &= \Gamma\left(A_{0,b}^{-1}\mu_b + \sum_j Z_j \Omega^{-1} b_j\right) \\ \Gamma &= \left(A_{0,b}^{-1} + \Omega^{-1} \otimes \sum_j Z_j Z_j'\right)^{-1} \end{aligned} \tag{15}$$

The covariance of matrix Ω has the conditional posterior distribution

$$f(\Omega|\mathbf{B}, \alpha_b, \mathbf{Z}) = f_{IW}\left(\Omega|\, T_0 + \sum_j \left(b_j - Z_j'\alpha_b\right)\left(b_j - Z_j'\alpha_b\right)', v_0 + J\right). \tag{16}$$

The posterior distribution for each b_j cannot be given an analytical expression, but they observe the proportionality:

$$f(b_j|\alpha, \Lambda, \mathbf{D}) \propto \left(\prod_n \prod_s p_{j,s,n}^{y_{j,s,n}}\right) f_N\left(b_j|Z_j'\alpha_b, \Omega\right) \tag{17}$$

with

$$f\left(\mathbf{B}|\alpha,\Lambda,\mathbf{D}\right)=\prod_{j}f\left(b_j|\alpha,\Lambda,\mathbf{D}\right) \tag{18}$$

and the posterior for α_c obeys:

$$f\left(\alpha_c|\mathbf{B},\Lambda,\mathbf{D}\right)\propto\prod_{j}\left(\prod_{n}\prod_{s}p_{j,s,n}^{y_{j,s,n}}\right)f_N\left(c_j|Z_j'\mu_c,A_{0,c}\right). \tag{19}$$

With regard to the misreporting parameters Λ, the priors are flat, therefore:

$$f\left(\Lambda|\mathbf{B},\alpha_c,\mathbf{D}\right)\propto\prod_{j}\prod_{s}\prod_{n}p_{j,s,n}^{y_{j,s,n}}P\left(\Lambda\right) \tag{20}$$

3. Model Estimation

Using the posteriors in Section 2.5, fairly straightforward algorithms can be employed to map the posteriors of α_b, and Ω, conditionally on values of α_c and Λ. In summary, equations [14] through [16] can be used for the 'Gibbs steps' and [17] provides the basis for a M-H step for the latent data. However, should some of the parameters be fixed or contain misreporting probabilities, then additional M-H steps are required to map the posterior distributions of α_c and Λ, based on the proportionalities in [19] and [20] respectively. Alternatively, should all parameters be fixed, then a M-H algorithm using only [19] and [20] can be employed to estimate the model.

When estimating the models without mis-reporting or fixed parameters, our initial estimation procedures and proposal densities were of a similar form to that described in Train and Sonnier (2005). The performance of these unmodified algorithms were investigated using both Monte-Carlo and real data (not the subject of this paper). The estimation of models without random parameters converged quite quickly. However, with random parameters, the alogrithms were slower. Our procedures and those made available by Kenneth Train had approximately the same rates of convergence. Furthermore, ensuring convergence with real data was often more problematic, with estimation sometimes doubling, relative to when Monte-Carlo data was used. The rates of convergence also depended on the types of tranformations $g(\beta_j)$ that were used. Models that included misreporting were slower to compute (perhaps an increase of up to 50%), because of the extra steps required to compute the model, as well as the increased dependence that was evident in the values generated by the sampler.

In an effort to improve the speed of convergence of our MCMC algorithms, we employed a number methods discussed in the Bayesian literature (e.g., Gilks and Roberts, 1996). In particular, we found that the performance of the independence sampler was generally superior to the random walk algorithm. At equivalent acceptance rates, the typical dependence of the sampler was reduced. But, as noted by Roberts (1996), independence samplers are unlikely to be an optimal 'stand alone' algorithm. So we used a mixture of the proposal densities, such that at each iteration, a proportion of b_j were randomly assigned a random walk proposal density, and others were assigned the independence proposal density. Also for our larger models we employed a method called "heating the chain" periodically, which is then followed by 'burn in' phases, before returning to recording the output from the sampler (Chen et al, 2000). In problem cases, this strategy gave at least a four fold increase in the efficiency of the algorithms. Details of all estimation and procedures are available from the authors on request.

Convergence of the sampler is monitored in several ways. First, visual plots of the sampled values are produced as the sampler runs for the sequences of α, Ω and λ. Second, the degree of dependence of the sampled values is examined by estimating the autocorrelation coefficients of the sequential values of the sampler. The 'skip' (only every 'skipth' iteration is recorded) was then set so as to allow a lesser degree of dependence should autocorrelation be too high (i.e. if coefficients have a first order correlation of more than 0.975 it is unlikely that an accurate value for the coefficients will be obtained when taking the mean or median, even with a sample of 10,000). Third, a modified t-test for the hypothesis of 'no-difference' between the first and second half of the sampled values (with a subset eliminated from the middle) was conducted on the sequence of α parameters. This used an estimate of the long-run covariance matrix (the spectral density matrix of the sequence at frequency 0) provided by the spectral kernel methods outlined in Andrews (1991). Our procedures enable the user to continue the run of the MCMC after a preliminary examination of the results should the sampler not have passed the tests for convergence.

3.1. Choice of Priors

In previous work Train (2003) and Train and Sonnier (2005) use non-informative (improper) priors for the mean and variance of α, and informative priors for the inverse Wishart distribution on Ω,setting these equal to $v_0 = k$ and $T_0 = k.I_k$. Our Monte-Carlo work indicated that setting $T_0 = k.I_k$ inflated estimates of the covariance matrices Ω generated by the sampler and inflated values of α also. This tendency depended on the values of Ω used to generate the data, the number of attributes, the sample size and the number of choice sets given to each respondent. However, when all parameters are random, setting $T_0 = 0.1 v_0 I_k$ and $v_0 = \frac{k_b(k_b+1)}{2}$, with $\frac{k_b(k_b+1)}{2}$ being the number of free elements in the covariance matrix, improved model performance, in that it did not over inflate the estimates of the covariance matrices Ω and were dominated by the data in cases where the elements of Ω were larger.

Proper priors are used for α because the marginal likelihood values cannot be computed without them. Fully non-informative priors can be obtained by setting the diagonal elements of A_0 to very large values. However, the priors employed here are set more informatively, with $A_0 = 10I_k$ and $\mu = 0$. In a standard linear regression framework these priors would, in most cases, be considered highly informative. However, in the context of the Logit we would argue that they are only weakly informative. Given a choice set ($\mathbf{X}$), we can explore what priors on α and Ω imply about priors on $p_{j,s,n}$. In terms of a prior for the probability of choosing a particular option, these priors are very similar. However, this prior information can sometimes have a substantive impact in restricting the absolute values of α, and in doing so also substantially improve the performance of the sampler. This is particularly useful in specifications that attempt to truncate the distributions of b_j (such as the truncated normal in Train and Sonnier, 2005) since, the whole of the distribution can become massed at a point of truncation, and α can become non-identified. In such circumstances, an informative prior can prevent this parameter wandering into non-identified regions indefinitely.

3.2. Calculating the Marginal Likelihood

One of the main aims of this paper is to evaluate alternative specifications. This requires the calculation of the marginal likelihood in [13]. The ratio of the marginal likelihoods gives the posterior odds for the two models(measuring the relative support for these models) given that the prior odds are even.

While theoretically straightforward, marginal likelihood calculations can be practically problematic in cases where the parameter space has many dimensions. Raftery (1996) provides a good discussion of methods available to calculate the marginal likelihood. For example, the method of Gelfand and Dey (1994) (GD), estimates the marginal likelihood using

$$\ln \hat{\mathcal{M}}(\mathbf{D}, M) = -\ln \left[G^{-1} \sum_{i=1}^{G} \frac{\psi(\Theta_i)}{P(\Theta_i)\ L(\Theta_i, \mathbf{D}, M)} \right] \tag{21}$$

or alternatively, if there are latent data,

$$\ln \hat{\mathcal{M}}(\mathbf{D}, M) = -\ln \left[G^{-1} \sum_{i=1}^{G} \frac{\psi(\Theta_i, \mathbf{B}_i)}{P(\mathbf{B}_i|\Theta_i)\, P(\Theta_i)\ L_f(\mathbf{B}_i, \Theta_i, \mathbf{D}, M)} \right] \tag{22}$$

where Θ_i and $\mathbf{B}_{s,i}$ are draws from their posterior distributions. The 'tuning functions' $\psi(\Theta_i)$ or $\psi(\Theta_i, \mathbf{B}_i)$ are densities with tails that are sufficiently thin so that the fractions within the expressions [21] or [22] are bounded from above. Alternatively, the tuning functions can be set equal to the priors, in which case the expressions [21] and [22] collapse to harmonic means. While this approach simplifies matters, harmonic means are known to be unstable and generally give poor estimates of the marginal likelihood (Raftery, 1996).

The second estimate [22] is the easier to calculate, for a given choice of $\psi(\Theta_i, \mathbf{B}_i)$. It does not require recording draws of $\mathbf{B}$, which would be memory intensive, providing $\psi(\Theta_i, \mathbf{B}_i)$ and $P(\mathbf{B}_i|\Theta_i)$ are recorded when running the sampler. However, Raftery (1996) suggests that the GD method tends to give poor estimates in high dimensional problems. As $\mathbf{B}$ contains up to $J \times k$ elements, our approach is to use [21] in preference to [22] in order mitigate the negative impacts of this dimensionality.

In performing the calculation of [21], as in Classical estimation, $L(\Theta, \mathbf{D}, M)$ can be simulated by making successive draws of $\mathbf{B}_t$ (t=1,......T) from $f(\mathbf{B}|\, \alpha, \Omega, \mathbf{Z})$ and calculating the likelihood as outlined in Chapter 10, of Train (2003). In our work we used a truncated normal tuning function (see Koop 2003 pp.104-106, for details). Computational difficulties arise using this method, since $L(\Theta_i, \mathbf{D}, M)$ requires computation by simulation for each value. Halton sequences, as described in Train (2003), greatly improve the efficiency of the simulated likelihood and these can be employed in simulating the likelihood at each of the posterior points facilitating the calculation of the marginal. Therefore, the results in this paper employ Halton sequences for this purpose. Monte-Carlo simulation suggested that using 500 replications in conjunction with Halton sequences yielded accurate values. In conducting the work in the paper, the GD likelihood calculations in [21] appeared to be able to work well in discriminating between different specifications generated using Monte-Carlo methods.

Finally, it is evident from equation [21] that estimation of the marginal likelihood is an average of the quantity $f(\Theta_i) = \frac{\psi(\Theta_i)}{P(\Theta_i)\, L(\Theta_i, \mathbf{D}, M)}$. As such, even in large samples, the numerical error in the estimate of the marginal likelihood needs to be considered when conducting model comparisons. An estimate of the numerical error can be obtained, but, the dependence in $f(\Theta_i)$ induced by the dependence of the sequence Θ_i needs to be accounted for in producing an estimate of this standard error. In this paper we employ a stationary bootstrap described in Li and Maddala (1997). A stationary bootstrap is a random length block bootstrap (e.g. Efron and Tibshirani, 1993) that accounts for the potential dependence in the sequence of likelihoods.

4. Empirical Section

4.1. Data.

The empirical work in this paper employs a new data set derived from a CE used to estimate the WTP for pesticide reductions in wheat used to produce bread. The use of bread as the product of interest is in keeping with Foster and Mourato (2000, 2002). For the application presented here, no explanatory variables are used for the parameters in the Utility function. In effect, therefore, $z_j = 1$ for all individuals.

Briefly, all CE respondents were presented with three choice cards, each consisting of three agricultural production practices:

- Policy A: current farming practices and national levels of pesticide applications;
- Policy B: a green policy employing less pesticides than under the status quo; and,
- Policy C: a nationwide ban on pesticide use.

The payment level was selected to be typical of consumer prices in the UK. The price of the 'standard' loaf was identified following an overview of price ranges as advertised in store and on the websites of the main UK grocers (around 50 pence for *status quo* loaf of bread). The alternatives and choice sets were constructed using a fractional factorial design that yielded 24 choice sets that were grouped in blocks of three choice cards. The survey instrument was piloted before being distributed by post to 3,000 households. The sample was stratified according to age, income and county of residence. The total number of analysable questionnaires was 420 and comparing the sample to national average figures indicates that our sample is reasonably representative of the UK population.

The attributes used in the study were percentage reductions insecticides, herbicides and fungicides. This is different to existing studies which only consider pesticide in gnereal. The reason for taking this approach was so that reductions for each of the types of pesticide could be considered so as to capture different affects on the environment. This allows to estimate the WTP for a reduction in a specific type of pesticide, something which has not been done before in the literature.

There were also three policies identified on each choice card. A and C are constant in each of their attribute levels (zero percentage reductions in all three attributes for A and 100% reduction in all three attributes for C). Policy B, however, varied across the choice sets. For this reason there are two other potential 'attributes' that can be added to the model, which treats 'Greenness' and 'Banning' as a different attributes, with intrinsic qualities that are not embodied in the values of the attribute levels associated with policy B. An alternative (i.e., attribute) specific constant cannot be included for Policy A as it would be perfectly collinear with the other attributes in the system. This approach is commonly employed in the environmental as well as marketing orientated choice modelling literatures before (e.g., Bjorner, *et al*, 2004). Since within each choice set, there were three policies, not only can potential misreporting be identified, but the direction of misreporting, should it exist, can be identified in favour of policies, A, B, C. A complete description of the survey instrument and data is presented in Chalak *et al* (2006).

4.2. Results

Our procedures allow us to estimate models with five alternative distributions for the parameters in preference and/or WTP space: normal; log-normal; truncated normal (with values below zero massed at zero); and the SB distributions (see Train and Sonnier, 2005, for more details). As discussed this paper extends these options to include fixed specifications on some or all of the coefficients. This enables us to test for the dominance of the ML over the fixed parameter Logit model. In addition, we estimate the models in WTP space as well including misreporting and all models estimated are compared using marginal likelihoods. We did not explore using truncated distributions or SB distributions as we do not, *a priori*, wish to restrict the WTP for the attributes in the choice sets to be positive, whereas distributions such as the log-normal and truncated normal involve positivity constraints. The SB distribution can usefully impose bounds on the parameters, but this requires a reasonable idea as to where the bounds of this distribution lie, and we have no secure prior knowledge in this regard. Consequently, alternative distributions are explored for the payment coefficient only.

Our model results are reported in Table 1. Using the methods discussed in section 3, all models converged tolerably. Each model is denoted as B and there are 12 different versions presented. Models B1 and B1* are the fixed parameter Logit, and the remaining models are various ML specifications. Marginal likelihood values were calculated at all points in the sample, using 500 Halton sequences at each point, and these are used to derive the relative rank of the models estimated.

{Approximate Position of Table 1}

As we can see from Table 1 all the ML specifications outperform the fixed parameter Logit on the basis of the marginal likelihood estimates. Moreover, the specification of a fixed coefficient on the intercept is also not supported. The marginal likelihood for models B1 and B1* in Table 1 are much smaller than for any other specification, and the next worst performing model (of those that include all the attributes) are models B5 and B5*. The fixed parameter specification B1* identifies significant misreporting on behalf of respondents, with 39% ($\pi = 0.61$) of respondents being estimated to misreport. The posterior distribution for this parameter is presented Figure 1.

{Approximate Position of Figure 1}

As can be seen from Figure 1, the distribution is unimodal, with its mode being away from its boundaries. The estimated direction of misreporting (mean of posterior, not presented in tables) is 22%, 38% and 40% in favour of options 1, 2 and 3 respectively. Thus, this model suggests that a large proportion of the population has a tendency to misreport in favour of the Green (B) policy or the Total Ban (C) policy which are always associated with options 2 and 3.

However, support for misreporting is significantly reduced for all the ML specifications. There is support when we consider models B2 and B2*, where the marginal likelihood for the misreporting model B2* is larger than its standard version B2. But, the proportion of respondents misreporting is extremely small proportion at 1.27%. The direction of misreporting for B2* (mean of posterior, not presented in tables) is 25% 27% and 48% in favour options 1, 2 and 3, respectively, again suggesting that those people who misreport, do so in favour of a total ban. However, the standard error on the marginal likelihoods are large, and we cannot be confident that B2* is truely preferred using this model. Likewise, for model B5 and B5*, the marginal likelihood could not be estimated accurately enough to discriminate between the standard and misreporting counterpart, though the point estimate of the marginal likelihood supports the specification misreporting. Further MCMC trials might allow us to discriminate between the models B2 and B2*, and B5 and B5*. However, since these models are dominated by B3, B3*, B4 and B4*, we do not conduct any further analysis on these models.

Focusing on B3, B3*, B4 and B4*, our superior performing models, we find that the data do not support a model where all parameters have normal distributions. These models all have log-normal distributions for the price coefficient. The models B3 and B3* are estimated in preference space, and the models B4 and B4* are estimated in WTP space. As can be observed from Table 1, the models estimated in WTP space have higher marginal likelihoods than their counterpart specifications estimated in preference space. Both sets of models with log-normal coefficients fail to support misreporting on the basis of their marginal likelihood estimates.

We also considered dropping the attribute specific constants for policies B and C. In this case, models B6 and B6*are not supported on the basis of the marginal likelihoods. Interestingly, B6* estimates that around 32% of respondents misreport their preferences. The direction of misreporting (mean of posterior, not presented in tables) is estimated to be predominantly in the direction of the Green policy B, (around 98%) with around 1% of people misreporting in favour of each other option. This is consistent with the findings in Table 2 that suggest that the Green Policy seems to be the attribute with the highest WTP. Summarising the model selection results:

i) ML models are preferred over the fixed parameter Logit model;

ii) Models that restricted the heterogeneity in the price were inconsistent with the data;

iii) Misreporting potentially appears to be partly a function of inappropriate restrictions (i.e., mis-specification) on heterogeneity as misreporting was supported by the 'all normal' model;

iv) The models with log-normal coefficients on the price attribute were preferred to other specifications, and of these the model estimate in WTP space was preferred.

Next we examine the WTP estimates. Table 2 presents the WTP estimates for the two top models (B3 and B4), as selected using the marginal likelihoods.

{Approximate Position of Table 2}

As already noted, these are the two models with log-normal distributions for the price coefficients, and normal distributions for all other parameters. The price coefficient in the second column of Table 2 is not a WTP estimate. It is the coefficient of the price in the utility function. This parameter only plays the role of a 'scale' parameter. Examining the median values of the WTP distributions for model B4, the value 23.5 pence in the second column of the row headed by 'Insecticide Reduction', means that consumers are prepared to pay, on average, an additional 23.5 pence on a 50 pence loaf of bread (an approximate 50% increase) for a 100% reduction in insecticides (and half this amount for a 50% reduction *etc.*). Herbicide and fungicides are estimated to have positive but smaller median WTPs at 9 and 15 pence respectively. Interestingly, policies B and C are estimated to have even higher WTP values than direct reductions in pesticides. The attribute with the highest WTP is that of the 'Green Policy'. Consumers are estimated to be prepared to pay an additional 56.7 pence on a basic 50 pence loaf of bread if it was produced as part of a wider policy that involved partial reductions in pesticide usage. On the other hand, consumers are less enthusiastic on the idea of having a total ban on pesticide usage, being estimated to be WTP only half that of the 'Green Policy' (at around 27 pence).

The stronger support for a Green Policy over a Ban as revealed by the relative WTP estimates is interesting. We might speculate that this occurs because there is maybe a reluctance to support a ban on pesticides because this will impact consumer choice. This argument has been advanced previously in the literature by Hamilton et al (2003) and Canavari et al (2005).

Notably, for model B4 estimated in WTP space, the mean and median values are almost identical. Unsurprising, in that each of the WTP parameters are assumed to be normally distributed. However, turning to model B3 and the preference space results in the last two columns of Table 2, it is evident that the median values (in the last column) are very similar to the median and mean results produced by the WTP space model. However, the mean results are very different even though the distribution for the price coefficient was constrained to be log-normal. The mean WTPs are higher across the board. This reflects the relatively large leverage of small values of the denominator (the price coefficient) in the WTP calculation that is the ratio of two coefficients. In our view, the WTP values derived from the WTP space estimation are quite high, and the mean WTP estimates produced by the preference space model are very high. Moreover, the ordering of the attributes according to their WTP values changes whether the mean or median is used in the preference space model. This is most likely due to the different correlations for each of the attribute coefficients with the price coefficient.

How do these WTP estimates compare to those previously reported in the literature? The most obvious comparison is with those for the UK. With Foster and Mourato (2000), for example, they estimated that UK consumers are WTP £ 1.15 (or 191 percent extra) for a "green" loaf of bread in order to reduce to zero cases of ill health per year and the number of declining farmland bird species jointly. Balcombe et al. (2007) looked at food choice for a whole basket of goods found that older females who classified themselves as either food safety aware or environmentally sensitive were WTP 150 percent more for the non-pesticide food. In contrast young males who described themselves as price sensitive yielded a WTP of almost zero. Overall, the sample average was 90 percent. So our WTP estimates are plausible in as much as they fall within the bounds of those previously reported.

Finally, in Table 3 we examine the correlation coefficients for the top performing model B4.

{Approximate Position of Table 3}

From Table 3 we can see that the price coefficient is negatively correlated to the other WTP coefficients. As already noted, the price coefficient plays the role of a 'scale' parameter. However, its value may conflate a number of effects including the individuals income level and marginal utility of income. This parameter does have meaning, at least from a statistical perspective. If two individuals have the same WTP for all attributes, then the person with the higher scale parameter is predicted to make a choice with more certainty than an individual with a smaller scale coefficient (since the relative variance of the Gumbel error is smaller). Train and Weeks (2005) discuss the interpretation of the scale parameter in detail. The negative correlations between the scale coefficients and the other coefficients suggest that those with a lower WTP for the model attributes are no less predictable than those with higher WTPs. The correlations between the WTP values are mainly positive with the exception of herbicide and insecticide. This is consistent with *a priori* expectations that an individual who is prepared to pay a larger than average amount for a reduction in one type of pesticide would also be WTP a larger than average amount for a reduction in another, or support a green or total ban policy. However, in general these are small. The 'ban' coefficient is the most positively correlated coefficient with the coefficients of the herbicide, fungicide and green policy attributes. Therefore, those supporting a total ban were generally individuals that were also prepared to pay higher amounts for reductions in the fungicidal and herbicidal reductions.

5. Summary and discussion

This paper has generalised existing approaches to the estimation of the ML and employed an original data set designed to obtain WTP estimates for reduced pesticides in bread production. Several questions were addressed. First, whether there was evidence that respondents misreported their preferences. Second, whether the ML logit was supported by over the fixed parameter Conditional Logit. Third, whether models parameterised in WTP space model were preferred to those in preference space and, finally, we also tested a range of model specifications using the transformations outlined in Train and Sonnier (2005). In order to address these questions, the marginal likelihood was simulated by employing Halton sequences. Simulation of the marginal likelihood proved practical, generally doubling the required computation time, with tolerable standard errors that enabled us to usefully distinguish between models.

With regard to misreporting. Previous findings using the fixed parameter Conditional Logit (i.e., Balcombe *et al*, 2007) were confirmed. There appeared to be a substantive proportion of respondents misreporting when using the fixed parameter Conditional Logit. Around 25% of respondents were estimated to be misreporting 35% of the time. However, when allowing for heterogeneity within the ML framework, it was found that this apparent misreporting largely disappeared. This finding was generally supported regardless of the transformations employed for the distributions of the parameters. However, any attempts to restrict heterogeneity (such as restricting the intercept to be constant) increased the support for the misreporting result.

It is worth noting that the way in which we deal with misreporting in this paper can be complemented by the introduction of additional questions as part of the survey instrument. For example, it has become common place in Contingent Valuation studies (e.g., Alberini *et al*, 2003 and Vossler and Poe, 2005) to ask respondents to indicate their level of response uncertainty. The degree of uncertainty expressed by respondents is then used to calibrate (i.e., reduce) the initial WTP estimates. There is no reason why we cannot do the same thing as part of a CE. Indeed, it would be interesting to see if the degree of misreporting estimated econometrically is similar to the magnitude as measured by an additional certainty follow up question.

Another key finding of this study, is that the ML was emphatically supported over the standard Logit using Bayesian methods. There was also no support for 'fixing' the payment coefficient. This is an important finding as current practice in the ML literature is to assume a fixed price coefficient. Based on the findings presented in this paper we consider it essential that this assumption is always tested in empirical studies.

Our data also supported a model estimated in WTP space over those estimated in preference space. The point estimates of WTP using this model were quite robust. Also, the median estimates of the WTP generated by a similar model, estimated in preference space, were almost the same. The mean and median WTP estimates were practically the same when using WTP space estimation. However, the mean estimates of a model estimated in preference space were very different to both the median estimates produced by the preference space model, or to the mean and medians estimates from the WTP space model. This was in spite of the fact that the payment coefficient was specified to be log-normal in both models. Our conclusion is, therefore, if estimation is performed using the ML in preference space, WTP estimates should be median rather than mean estimates.

Finally, as a practical observation, we suggest that practitioners using the Bayesian ML pay considerable attention to the sequence of values that are generated by the sampler. Visual observation and formal tests for convergence are both valuable in this respect. In this data set we found that a very large number of iterations were often needed before the distributions of the parameters were accurately mapped.

Acknowledgements

The authors acknowledge that the *GAUSS* procedures used herein incorporated much of the orginal code distributed by Kenneth Train on the Web. They also thank Kenneth

Train and Riccardo Scarpa for commets on an earlier version of this paper.

References

Alberini, A., Boyle, K. and Welsh, M. (2003). Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty, *Journal of Environmental Economics and Management*, 45: 40-62.

Allenby, G.M. and Rossi, P.E. (1999). Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, 89: 57-78.

Andrews, D.W.K (1991). Heteroscedastic and Autocorrelation Consistent Covariance Estimation. *Econometrica*, 59: 817-858.

Balcombe K., Bailey, A. Chalik A, and Fraser I. (2007) Bayesian Estimation of a Willingness to Pay where Respondents Mis-report their *Oxford Bulletin of Economics and Statistics,* 69 (3): 413-437.

Bjørner, T. B., Hansen, L. G. and Russell, C. S. (2004). Environmental labeling and consumers' choice - an empirical analysis of the effect of the Nordic Swan. *Journal of Environmental Economics and Management* 47(3): 411-434.

Cameron T. (1988) A new paradigm for valuing non-market goods using referendum data. Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management,* 15: 355-379.

Cameron T. and M. James (1987). Efficient estimation methods for closed-ended contingent valuation survey data. *Review of Economics and Statistics,* 69: 269-276.

Canavari, M., Nocella, G. and Scarpa, R. (2005). Stated Willingness-to-Pay for Organic Fruit and Pesticide Ban: An Evaluation Using Both Web-Based and Face-to-Face Interviewing, *Journal of Food Products Marketing*, 11(3): 107-134.

Caudill S. B. and Mixon, F. G. (2005). Analysing discrete responses: a Logit model based on misclassified data, *Oxford Bulletin of Economics and Statistics*, 67: 105-113.

Chalak, A., Balcombe, K., Bailey, A. and Fraser, I. (2006). Pesticides, Preference Heterogeneity and Non-Market Values: A Latent Class Model, Paper Presented to the 80th Annual Agricultural Economics Society Conference, Paris, March.

Chen M.H. Shao, Q.M, and Ibrahim J.G. (2000). *Monte Carlo Methods in Bayesian Computation.* Springer Series in Statistics. Springer, New York.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Ferrini, S. and Scarpa, R. (2007). Designs with a priori Information for Nonmarket Valuation with Choice Experiments: A Monte Carlo Study, *Journal of Environmental Economics and Management*, 53(3): 342-363.

Florax, R. J. G. M., Travisi, C. M. and Nijkamp, P. (2005). A meta-analysis of the willingness to pay for reductions in pesticide risk exposure. *European Review of Agricultural Economics*, 32(4): 441-467.

Foster, V. and Mourato, S. (2000). Valuing the multiple impacts of pesticide use in the UK: a contingent ranking approach. *Journal of Agricultural Economics,* 51(1): 1-21.

Foster, V. and Mourato, S. (2002). Testing for Consistency in Contingent Ranking Experiments, *Journal of Environmental Economics and Management,* 44: 309-328.

Gelfand, A. and Dey, D. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society Series* B, 56: 501-504.

Gilks W.R. and Roberts G.O. (1996). Strategeis for improving MCMC. Chapter 6 of *Markov Chain Monte Carlo in Practice.* Eds Gilks W.R. Richardson S. and D.J. Spiegelhalter. Chapman and Hall, 2006.

Hamilton, S.F., Sunding, D.L. and Zilberman, D. (2003). Public Goods and the Value of Product Quality Regulations: The Case of Food Safety, *Journal of Public Economics*, 87: 799-817.

Hausman, J.A., Abrevaya, J. and Scott-Morton, F. M. (1998). Misclassification of the dependent variable in a discrete-response setting, *Journal of Econometrics*, 87: 239-269.

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

Huber, J. and Train, K. (2001). On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths, *Marketing Letters*, 12(3): 259-269.

Koop G. (2003). *Bayesian Econometrics.* Wiley Publishers, West Sussex, England.

Layton, D.F. and Levine, R.A. (2003). How Much Does the Far Future Matter? A Hierarchical Bayesian Analysis of the Public's Willingness to Mitigate Ecological Impacts of Climate Change, *Journal of the American Statistical Association*, 98(463): 533-544).

Layton, D.F. and Levine, R.A. (2005). Bayesian Approaches to Modeling Stated Preference Data, in *Applications of Simulation Methods in Environmental and Resource Economics*, R. Scarpa and A. Alberini (Eds.), Kluwer Academic Publishing.

Li H. and G.S. Maddala (1997) Bootstrapping cointegrating regressions, *Journal of Econometrics* 80, 297-318.

Louviere, J.L. (2006). What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments, *Environmental and Resource Economics*, 34: 173-188.

McFadden, D. and Train, K. (2000). Mixed MNL Models for Discrete Response, *Journal of Applied Econometrics*, 15: 447-470.

Raftery A.E. (1996). Hypothesis testing and model selection. Chapter 10 of *Markov*

Chain Monte Carlo in Practice. Eds Gilks W.R. Richardson S. and D.J. Spiegelhalter. Chapman and Hall, 2006.

Revelt, D. and Train, K.E. (1998). Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level, *Review of Economics and Statistics*, 80(4): 647-657.

Roberts, G.O. (1996) Markov chain concepts related to sampling algorithms. In Chapter 3 of *Markov Chain Monte Carlo in Practice*. Eds Gilks W.R. Richardson S. and D.J. Spiegelhalter. Chapman and Hall, 2006.

Rossi, P.E., McCulloch, R. and Allenby, G.M. (1996) The Value of Purchase History Data, *Journal of Marketing Science*, 15,4, 321-340.

Ruud P. (1996) Simulation of the multinomial probit model: An analysis of covariance matrix estimation' Working Paper, Dept of Economics, University of California.

Scarpa, R. and Alberini, A. (Eds) (2005). *Applications in Simulation Methods in Environmental Resource Economics*, Kluwer Academic Publisher.

Smith, M.D. (2005). State Dependence and Heterogeneity in Fishing Location Choice, *Journal of Environmental Economics and Management*, 50: 319-340.

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

Train, K.E. (2003) *Discrete Choice Methods and Simulation*. Cambridge University Press.

Train, K. and G. Sonnier (2005). Mixed Logit with Bounded Distributions of Partworths, in *Applications of Simulation Methods in Environmental Resource Economics*, R. Scarpa and A. Alberini (Eds.), Kluwer Academic Publishing.

Train, K. and M. Weeks (2005). Discrete Choice Models in Preference Space and Willingness-to-Pay Space, in *Applications of Simulation Methods in Environmental Re-*

source Economics, R. Scarpa and A. Alberini (Eds.), Kluwer Academic Publishing.

Vossler, C.A. and Poe, G.L. (2005). Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty: A Comment, *Journal of Environmental Economics and Management*, 32: 219-232.

Table 1: Main Model Results

	pr	in	hb	fg	gr	bn	wp	mr	rk	MargL	se
B1	0	0	0	0	0	0	.	.	12	-1253.88	0.003
B1*	0	0	0	0	0	0	.	.612	11	-1251.88	0.231
B2	1	1	1	1	1	1	.	.	6	-919.90	0.601
B2*	1	1	1	1	1	1	.	.987	5	-917.72	0.516
B3	2	1	1	1	1	1			3	-914.36	0.629
B3*	2	1	1	1	1	1		.987	4	-914.79	0.714
B4	2	1	1	1	1	1	y		1	**-905.90**	0.191
B4*	2	1	1	1	1	1	y	.984	2	-911.75	0.731
B5	0	1	1	1	1	1	.		8	-936.91	0.463
B5*	0	1	1	1	1	1	.	.984	7	-935.32	0.942
B6	2	1	1	1	.	.	y	.	10	-1189.42	0.071
B6*	2	1	1	1	.	.	y	.682	9	-1058.4	0.156

Notes: pr= price, in=insecticide reduction, hb=herbicide reduction,

fg=fungicide reduction, gr=green policy (B), bn=ban policy (C)

WP=WTP Space

MR=Misreporting, Rk=ranking (1 best, 12 worst)

MargL=Marginal Likelihood,

se=Bootstrap Standard Error on the Estimated Marginal Likelihood.

For columns 2 to 7 numbers represent the following distributions

0=fixed, 1=normal, 2=log-normal

Table 2: WTP Estimates (£) - Models B3 and B4

	WTP Space (B4)		Prefrence Space (B3)	
	Mean (Stdv)	Median (Lower-Q) [Upper-Q]	Mean (Stdv)	Median (Lower-Q) [Upper-Q]
Price Coefficient	11.638 (5.124)	10.651 (7.982) [14.207]	1	1
Insecticide Reduction	0.235 (0.352)	0.237 (-0.001) [0.472]	0.608 (1.447)	0.225 (0.078) [0.579]
Herbicide Reduction	0.091 (0.451)	0.091 (-0.211) [0.398]	0.446 (1.400)	0.0767 (-0.004) [0.360]
Fungicide Reduction	0.152 (0.416)	0.151 (-0.131) [0.433]	0.734 (2.333)	0.132 (-0.003) [0.573]
Green Policy	0.567 (0.413)	0.563 (0.288) [0.847]	1.290 (1.999)	0.693 (0.329) [1.445]
Complete Ban	0.276 (0.665)	0.276 (-0.172) [0.721]	1.325 (3.572)	0.373 (0.095) [1.152]

Notes: Lower Q=lower quartile; Upper Q=Upper quartile

Table 3: Correlations in Transformed Coefficients

	in	hb	fg	gr	bn
pr	-0.166	-0.056	-0.156	-0.300	-0.243
in	.	-0.117	0.085	0.151	0.063
hb	.	.	0.175	0.056	0.242
fg	.	.	.	0.167	0.416
gr	.	.	.	.	0.372

Notes: pr= price

in=insecticide reduction

hb=herbicide reduction,

fg=fungicide reduction

gr=green policy (B)

bn=ban policy (C)

GAUSS Thu May 25 09:57:12 2006

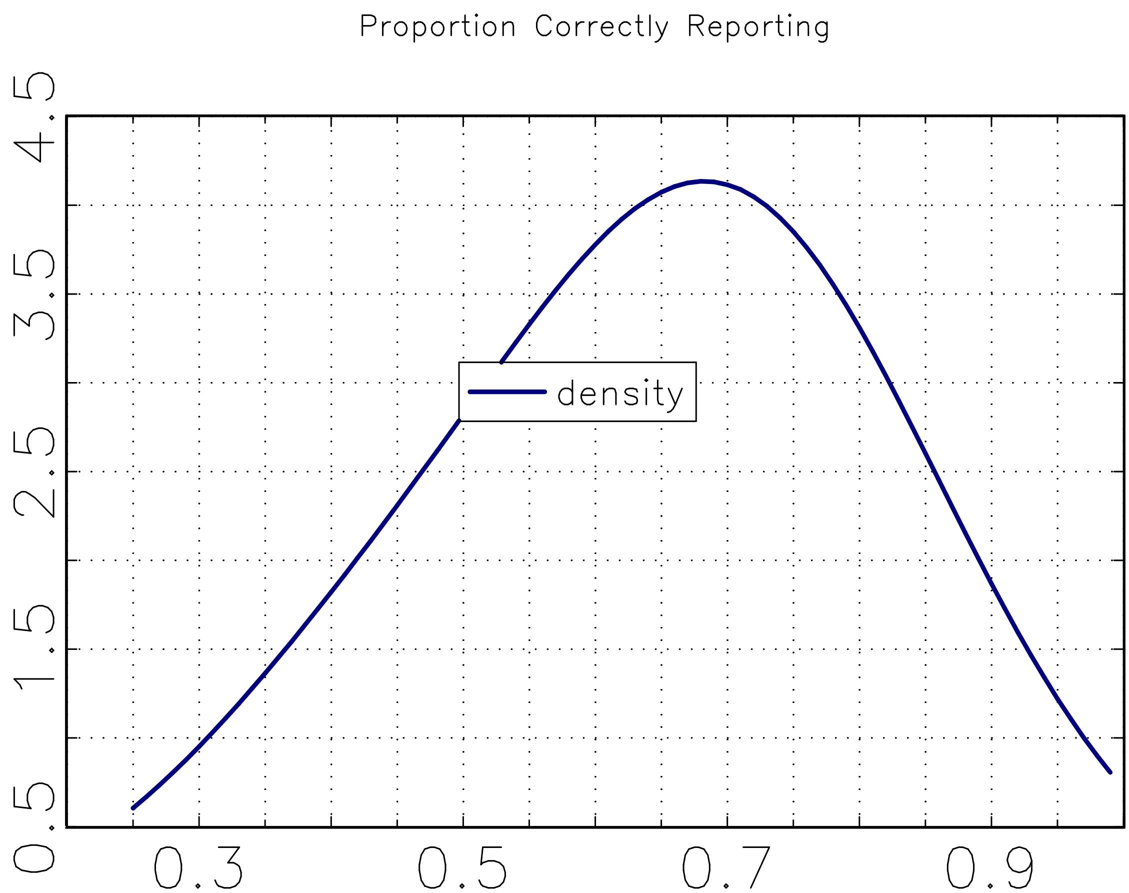

Figure 1: Proportion Correctly Reporting Bread Data