# Simple Ways To Estimate Choice Models For Single Consumers[1]

*Bart Frischknecht, University of Technology Sydney, bart.frischknecht@uts.edu.au*

*Christine Eckert, University of Technology Sydney, christine.ebling@uts.edu.au*

*Jordan Louviere, University of Technology Sydney, jordan.louviere@uts.edu.au*

## Abstract

We show with Monte-Carlo simulations and empirical choice data sets that we can quickly and simply refine choice model estimates for individuals based on methods such as ordinary least squares regression and weighted least squares regression to produce well-behaved in-sample and out-of-sample predictions of choices. We use well-known regression methods to estimate choice models, which should allow many more researchers to estimate choice models and be confident that they are unlikely to make serious mistakes.

Keywords: Individual-level choice models

[1] This is a working paper. Please do not distribute or cite without the authors' permission.

# Simple Ways To Estimate Choice Models For Single Consumers

## Introduction

Choice models were first proposed by Thurstone (1929) for pairs of options. Models for multiple choice options are due to Luce (1959) and McFadden (1974). Except for laboratory choice experiments in psychology, it is rare to see discrete choice models estimated for single people; and after Chapman (1984), there was little work on ways to measure and model single person choices in survey applications until recently. Finn and Louviere (1993) proposed a new measurement model and associated way to elicit choices called Best-Worst Scaling that gives individual-level estimates of preferences, values, etc. Marley and Louviere (2005) prove the measurement properties of the Best-Worst approach; and Louviere, et al (2008) extended the approach to allow one to estimate choice models for single persons. They used simulated and empirical data to show that one can use several estimation methods to model individuals.

The purpose of this paper is to show that one can use simple methods to model single person choices, which extends Louviere, et al (2008) to estimation methods familiar to most academics and practitioners, such as ordinary least squares regression and weighted least squares regression. These two methods yield biased estimates of the choice probabilities, but we demonstrate that one can "correct" these estimates rather simply. More specifically, we show with Monte-Carlo simulations and empirical choice data sets that we can refine simple estimates to produce well-behaved in-sample and out-of-sample predictions of choices. The contribution of the paper is to describe and discuss these methods to allow many more researchers to estimate choice models and be confident that they are unlikely to make serious mistakes.

The objective of the paper needs to be seen in proper context. For approximately the last 15 years there has been the equivalent of an arms race in academic marketing and applied economics. To wit, ever more sophisticated statistical models have been proposed and applied, with the result that the barriers to entry to do frontier academic research and sophisticated applied research are now extremely high. One result has been an increasing reliance of practitioners on commercial software that “does” choice experiments and allows one to produce the associated statistical models. Considering that the closest analogue to discrete choice models are models for statistical (quantum) mechanics in physics, one might well ask what our colleagues in the physical sciences would think of many under-trained practitioners “doing” quantum mechanics using off-the-shelf commercial software. A second outcome of this arms race is a gradual decline in new academics entering the field, which also has led to a fairly significant decline in papers on choice modelling presented at conferences like EMAC, Marketing Science and ANZMAC. Thus, the field could benefit from simple, easy-to-use methods that are highly likely to produce correct answers.

Of course, correctly analysing choice data is only one part of a larger whole associated with academic or applied work in choice modeling. One must first conceptualise academic and practical problems involving consumer and other choices and understand how to design and implement choice experiments. This paper does not address the precursors to choice data analysis, which require education, training and experience; but given that academics and practitioners can acquire these skills, the approaches proposed in this paper should allow many researchers to estimate choice models for single people and predict choice outcomes in many circumstances of academic and practical interest.

## Background

In the following, we will discuss the motivation for modelling individuals. We start with a brief history on modelling individual choices and provide some background on data expansion methods that provide the basis for estimating individual level models (ILMs).

### Modelling individuals

Early efforts to model individuals can be found in Beggs, Cardell , & Hausman (1981) and Chapman (1984), who show how individual-level preference rank order information from conjoint data can be expanded to estimate ILMs . However, apart from those early attempts, there has been little research on estimation of individual level choice models since. One reason for this is tradition. That is, following McFadden (1974), virtually all choice models were top-down models. That is, one makes assumptions about the form of the indirect utility function (decision rule) and its parameters (fixed or random) and the associated error distribution, and derives the associated choice model give these assumptions. One then applies the derived model form to choice data for samples of individuals to estimate the model effects.

Historically, research on to-down models evolved from simple, fixed effects models to more complex random parameter forms, and more recently, model forms that allow for various types of more complex latency, such as estimating the terms that can be identified in the variance-covariance matrix of the parameter estimates associated with an assumed distribution. So, marketing and economics academics began to focus more on incorporating latent heterogeneity into choice models. A seminal paper by Kamakura and Russell (1998) showed how to estimate segment-specific parameters for a finite number of latent segments. If one weights these parameter estimates by an individual's segment membership probability, the segment specific parameter estimates can be used to derive individual-specific parameters. The continuous distribution, or random parameters model form, such as the Mixed Logit model, assumes that consumer preferences follow a specific (continuous) distribution. Like the Kamakura and Russell (1998) latent class model, individual individual observations can be used to infer a individual's position on this preference distribution (see Train 2003, Rossi, Allenby& McCulloch, 2006). However, in the case of both finite mixture and random parameters models individual specific parameter estimates mayindividual be biased towards the segment or aggregate population mean, depending on the number of observations available per individual.

Recently there has been renewed interest in directly modelling individualindividual choices. For example, Louviere et al. (2008) show how to use observations from Best-Worst choice experiments to estimate individual preference parameters with WLS regression. Marshall, Chan, and Curry (2011) describe and discuss a comparison of the WLS approach with standard HB estimation of a Mixed Logit model, and report that both predict equally well. Finally, Ebling, Frischknecht, and Louviere (2010) compare several individual level modelling techniques in a recent conference paper.

These papers are motivated by the fact that discrete choice experiments require the selection of one alternative from several, which is the observed dependent variable. Statistical models for the analysis of such data are known as "latent dependent variable" models because the dependent variable of interest is some underlying continuous latent variable like strength of preference, but one observes only an indicator of the unobserved, latent measure. Typically, responses to choice experiments are coded one (1) for a chosen option and zero (0) for all unchosen option(s).All latent dependent variable models have a formal identification problem, namely that the standard deviation of the error component is perfectly inversely correlated with the model estimates. This poses no problem in predicting choice probabilities; but, as noted by Swait and Louviere (1993), it poses issues for comparisons of model estimates across data sources. Further, as noted by Fiebig, et al (2010), if the standard deviation of the error component is not constant across people, this can lead to seriously biased model estimates and misleading inferences. The latter consideration motivated us to find ways to account for choice variability differences across people. One way to account for these differences is to estimate models for single individualsindividual.

## Data expansion models

Estimation of ILMs usually fails due to an insufficient number of degrees of freedom to estimate parameters reliably. One way to overcome this limitation is to expand the available full or partial ranking information to artificially create implied new choice set and choice observations. That is, data expansion methods use full or partial preference rankings  of choice options to infer chosen options in implied choice sets. ExpansionRank order expansion to implied choices was first discussed by Luce and Suppes (1965), and first used in marketing by Chapman & Staelin (1982). The two prior references relied on a full ranking of choice options in a choice set, but more recent papers (e.g., Louviere et al. 2008, Ebling, Frischknecht & Louviere 2010) show that choice responses from best-worst questions asked about each choice set in discrete choice experiments (Louviere, Hensher & Swait 2000, Marley & Louviere 2005) can be used to generate partial rankings that can be expanded to implied new choice sets. An example of best-worst questions being used in a discrete choice experiment (DCE) is shown in Figure 1.

### Figure 1: Example of Best-Worst Choice Experiment

| | Solar Panel Product A | Solar Panel Product B | Solar Panel Product C | Solar Panel Product D |
|---|---|---|---|---|
| Capacity | 2.5 kW | 1.0 kW | 2.0 kW | 1.5 kW |
| Government Rebate | $2,600 | $5,500 | $8,500 | $2,200 |
| Production Output Warranty | 15 Years | 30 Years | 15 Years | 30 Years |
| Purchase Price | $16,250 | $5,500 | $17,000 | $11,250 |
| Payback Time | 8.8 Years | 0 Years | 4.8 Years | 6.8 Years |

Which ONE of the 4 products above is your MOST preferred choice? (tick one)

Product A   Product B   Product C   Product D

Of the remaining 3 products, which ONE is your LEAST preferred choice? (tick one)

Product A   Product C   Product D

In the above example, product B is chosen as best and product D is chosen as worst. This gives the partial ordering B>A, C>D (“>” denotes “is preferred to”). This partial ordering allows us to expand the data in the following two ways:

1) OLS-Expansion: Knowing that B is preferred to A, C, and D, we get three different choice sets, namely choice set {A,B,C,D}, where B is chosen as most preferred, choice set {A,D}, where A is chosen as most preferred, and choice set {C,D}, where C is chosen as most preferred.
2) WLS-Expansion: Alternatively, we can create all the possible  choice sets associated with the four alternatives (see also Louviere et al. 2008). There are 15 such non-empty choice

sets; in these 15 sets the most preferred option B appears 8 times, and it should be chosen each of these times, or 8 times. The least preferred option D also appears 8 times, but it should only be chosen once, namely when it appears as a singleton option with no other competing options. . The remaining two options A and C also appear in 8 choice sets, but they can be chosen only if the most preferred option B is not present, and we cannot tell from the partial order whether A is better or worse than C. Thus, we assign both the average of the remaining two expected choice counts, or 3 choices each. In this way, we can associate expected choice frequencies of 8, 3, 3, and 1 to alternatives B, A, C, and D, respectively.

We thus use rank order expansion procedures to construct r-1 choice sets from a subset of r rankings of N options. Rank order expansion also can be seen as a way to aggregate choices; i.e., 1, 0 choices are fully disaggregated because they are associated with the most elemental level of choices, namely a particular person, a particular choice set and a set of particular choice options (alternatives). Hensher and Louviere (1984), Louviere and Woodworth (1983) and Louviere, et al (2008) (among others) show how to aggregate choices across people or choice sets by calculating and analysing choice counts or choice proportions. Non-experimental analogues of these aggregate choice counts were considered by (among others) Berkson (1944) and Theil (1969) for cases where real market choices are aggregated by choice alternatives and choice sets or other differences in choice observations. If one has such choice counts or one can transform rankings into such counts (via rank order expansion, Louviere, et al 2008), one can apply WLS regression to estimate choice model parameters. Green (1984) discusses using WLS as the first step of the maximum likelihood estimator, which is the approach used by Louviere and Woodworth (1983). That is, the dependent variable is the natural log of the expected choice frequencies, and a weight vector, which is the expected choice frequencies is used to correct for the heteroscedastic errors associated with the multinomial choice proportions or frequencies.

## Proposed Modelling Approaches

In the following we illustrate how to estimate a separate model for each individual to obtain each individual's preferences without biasing towards a population mean.

In particular, we investigate ordinary least squares regression models estimated directly from the 1, 0 choice indicators, ordinary least squares regression models estimated from the 1, 0 choice indicators of the OLS-exploded choice sets, and weighted least squares regression models estimated from the expected choice frequencies associated with the partial rankings of the choice options in each set (hereafter termed "WLS" models). The OLS approach is motivated by the work of the Nobel Laureate James Heckman and co-author Synder (1998), who proposed that OLS regression models (i.e., "Linear Probability Models") may be a better way to model binary discrete choices if errors are asymmetric; naturally, no one knows if errors in latent utility functions are symmetric or not, so this approach is worth considering. The WLS approach is motivated by rank order expansion of choices, originally discussed by Luce and Suppes (1965) and first used in marketing by Chapman and Staelin (1982). Parameter estimates from OLS and WLS are then used to predict choice shares via a logit link function.

In the case of both OLS and WLS, the model estimates are on the wrong "scale" to correctly predict the choice probabilities. That is, the parameter estimates are too small or too large and systematically under- or over-predict the observed choice probabilities, and so must be corrected. We use a simple correction that seems to work quite well under a variety of simulated and actual conditions for both in- and out-of-sample predictions. Specifically, one estimates an OLS or WLS model for a single person and uses the model to predict the observed dependent variable for each choice option in each choice set of interest. Then, one calculates the associated residual mean squared error for each person. One then multiplies all OLS or WLS model estimates for an individual by a 1/(mean squared error), which is the correction factor. The "corrected" parameters are used to make choice predictions for each person using a logit link function.

The following formulas describe the procedure for OLS and WLS estimation in more detail. We make no distinction between OLS applied to "normal" versus expanded choice sets because the estimation approach does not differ, except that there are different numbers of choice observations used for estimation.

## OLS Estimation & Rescaling

Let *i* represent the individual, *r* represent the choice set, *j* represent the alternative in the choice set, and *k* represent other alternative in the choice set. Then for each individual we estimate a model where the dependent variable is

$$y_{rj} = \begin{cases} 1, & \text{if alternative } j \text{ chosen in choice set } r \\ 0, & \text{if alternative } j \text{ not chosen in choice set } r \end{cases}$$

$\hat{\beta}_{i1}$ $\delta_r$

For each individual we estimate preference parameters$\beta$ and “choice set-specific” parameters$\delta$ via minimizing

$$\min_{\boldsymbol{\beta}_i,\delta_i} S_i = \sum_{r=1}^{R}\sum_{k=1}^{K}(y_{irj} - \hat{y}_{irj})^2$$

where

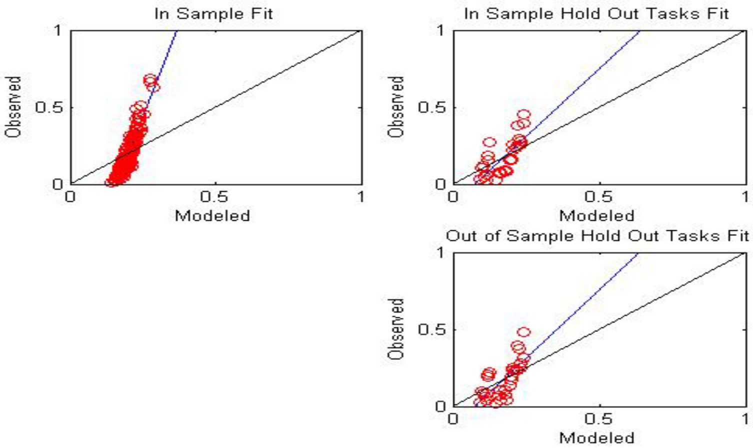
$$\hat{y}_{irj} = \hat{\beta}_{i1}x_{1rj} + \hat{\beta}_{i1}x_{2rj} + \ldots + \hat{\beta}_{i1}p_{rj} + \delta_r + \varepsilon_{irj}.$$

If we do not rescale the preference parameters and use the logit link function

$$\hat{P}_{irj} = \frac{e^{\hat{\boldsymbol{\beta}}_i^{\mathrm{T}}\mathbf{x}_{rj}}}{\sum_{k=1}^{K} e^{\hat{\boldsymbol{\beta}}_i^{\mathrm{T}}\mathbf{x}_{rk}}}$$

to predict choices, we obtain biased results, such as those shown in Figure 2.

**Figure 2: Predicted Aggregate OLS Choice Shares Without Rescaling of Parameters**




Thus, using “uncorrected” OLS to estimate the model parameters from the observed choices produces parameters that are on the wrong scale. In order to correct for this bias, we rescale the parameters from the first-stage OLS estimation as described above. In particular, we define the mean square error as a function of the regression residuals

$$MSE_i = \frac{1}{RJ}\sum_{r=1}^{R}\sum_{j=1}^{J}\left(y_{irj} - \hat{y}_{irj}\right)^2$$

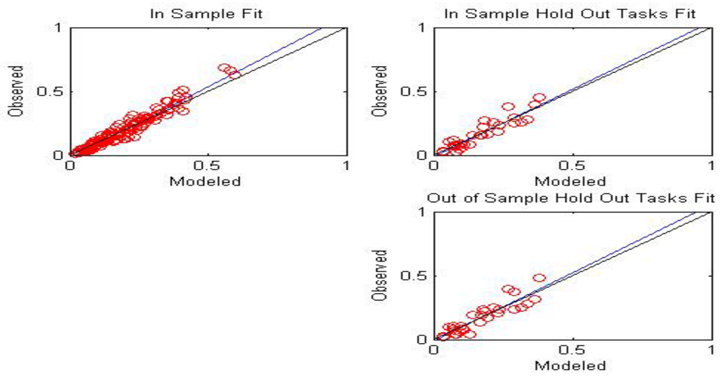
and define the individual rescaling coefficient as $\alpha_i = \frac{1}{MSE_i}.$

We then rescale model parameters as $\hat{\boldsymbol{\beta}}_i' = \alpha_i \hat{\boldsymbol{\beta}}_i$, and again apply the logit link function to predict the choice probabilities, as shown below:

$$\hat{P}_{irj}' = \frac{e^{\hat{\boldsymbol{\beta}}_i'^{\mathrm{T}} \mathbf{x}_{rj}}}{\sum_{k=1}^{K} e^{\hat{\boldsymbol{\beta}}_i'^{\mathrm{T}} \mathbf{x}_{rk}}},$$

The above yields very precise predictions of the aggregate choice shares, as shown in Figure 3:

### Figure 3: Predicted Aggregate OLS Choice Shares After Rescaling




### WLS Estimation & Rescaling

Again, let *i* represent the individual, *r* represent the choice set, *j* represent the alternative in the choice set, and *k* represent other alternative in the choice set. Then for each individual we estimate a model where the dependent variable is as shown below for the case of choice sets containing four alternatives, and only having a partial ranking of best and worst available:

$$y_{rj} = \begin{cases} \ln(8), \text{ if alternative } j \text{ chosen "best" in choice set } r \\ \ln(3), \text{ if alternative } j \text{ not chosen in choice set } r \\ \ln(1), \text{ if alternative } j \text{ chosen "worst" in choice set } r \end{cases}$$

For each individual we estimate preference parameters $\hat{\beta}_{i1}$ and choice set specific parameters $\delta_r$ via minimizing

$$\min_{\boldsymbol{\beta}_i, \delta_i} S_i = \sum_{r=1}^{R} \sum_{k=1}^{K} W_{rij} (y_{irj} - \hat{y}_{irj})^2$$

where

$$\hat{y}_{irj} = \hat{\beta}_{i1} x_{1rj} + \hat{\beta}_{i1} x_{2rj} + \ldots + \hat{\beta}_{i1} p_{rj} + \delta_r + \varepsilon_{irj}$$

and we use the weights below to correct for heteroscedastic errors:

$$W_{rj} = \begin{cases} 8, \text{ if alternative } j \text{ chosen "best" in choice set } r \\ 3, \text{ if alternative } j \text{ not chosen in choice set } r \\ 1, \text{ if alternative } j \text{ chosen "worst" in choice set } r \end{cases}$$

We again must rescale the parameters obtained from the procedure above. In particular, we define mean square error to be a function of the regression residuals.

$$MSE_i = \frac{1}{RJ} \sum_{r=1}^{R} \sum_{j=1}^{J} \left( y_{irj} - \hat{y}_{irj} \right)^2$$

and define an individual's rescaling coefficient as $\alpha_i = \dfrac{1}{\dot{M}SE_i}$

We then rescale the model parameters as $\hat{\boldsymbol{\beta}}_i' = \alpha_i \hat{\boldsymbol{\beta}}_i$ , and again apply a logit link function:

$$\hat{P}_{irj}' = \frac{e^{\hat{\boldsymbol{\beta}}_i'^{\mathrm{T}} \mathbf{x}_{rj}}}{\sum_{k=1}^{K} e^{\hat{\boldsymbol{\beta}}_i'^{\mathrm{T}} \mathbf{x}_{rk}}}.$$

## Simulation Study

We now describe and  discuss a simulation study that demonstrates the validity of our approach under a wide array of conditions likely to occur in real applications.

The simulation varied three different factors:

1) Sample size: 2050 versus 3000 individuals, out of which 2000 individuals were only used for hold-out performance measurement.
2) Number of choice sets (we held set size constant at 4 alternatives) per individual: 4, 8, 16, 32, 64 choice sets. All simulated individuals saw identical choice tasks.
3) The underlying preference distribution: fixed, normal, normal with higher variance, and tri-modal mixture of normal distributions

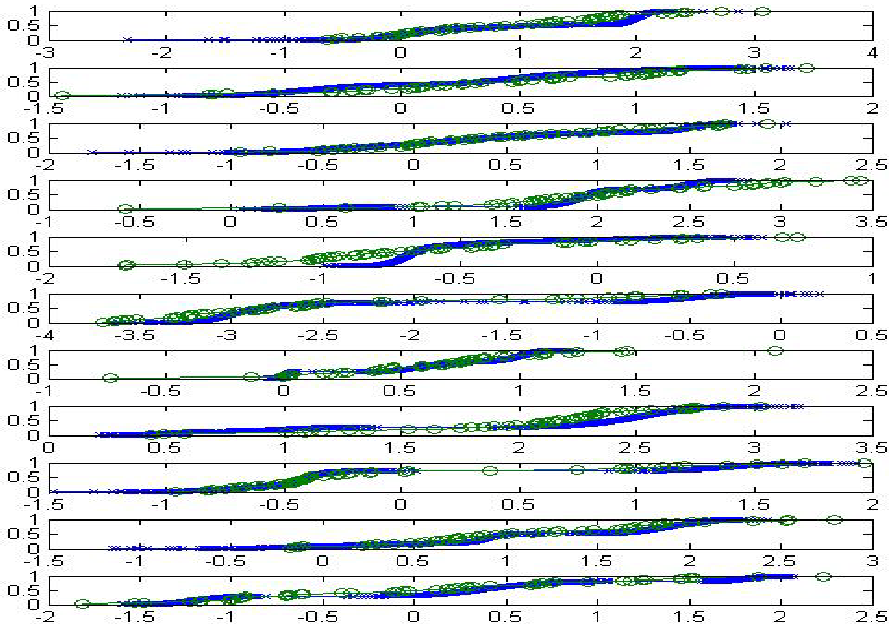
The alternatives in our simulated data set had 11 attribute levels. For each individual we simulated parameters by drawing from the joint distribution of parameters. We added an EV type 1 error term to each alternative in each choice set for each individual to obtain logit choice probabilities (for each alternative in each choice set for each individual) that we used to simulate individuals' choices.

We compared the performance of the OLS individual-level models against a commonly used Mixed Logit specification with normally distributed priors and a diagonal covariance. We estimated the Mixed Logit via Hierarchical Bayes using Ken Train's code for Matlab.

We compared the performance of the two OLS methods (i.e., OLS on non-expanded and expanded choice sets) based on three different comparison measures:

1) To measure population parameter distribution recovery, we use the sum of the Kolmogorov Smirnov (KS) test statistic for all 11 parameter distributions. The KS test statistic is a measure for one-dimensional distributions that compares the simulated CDF with the estimated CDF (see also Figure 4). For each parameter distribution the KS test statistic yields a value between 0 and 1 with a value closer to 0 implying that the simulated CDF is closer to the true CDF. Thus, in our simulation, the sum of the KS test over the 11 dimensions can take on values between 0 and 11. Of course, this measure is only available in simulations, when the true preference CDF is known.

### Figure 4: Example of Simulated versus Estimated CDF




2) In order to assess the precision of population level prediction, we calculated the mean squared error for differences between aggregate observed choice shares and aggregate estimated choice shares across alternatives and choice sets (MSE). A lower MSE means that a model fits the observed aggregate choices better. This measure is available in simulations or experimental data with holdout tasks or holdout samples.
3) Finally, to capture the accuracy of individual predictions, we calculated the mean of the individuals' root likelihoods. The root likelihood takes on values between 0 and 1 with a value closer to 1 implying that the model fits the observed choices better. The root likelihood is the geometric mean of the estimated choice probabilities for the observed choices and is again available in simulations or experimental data with holdout tasks.

We averaged the results of these measures over 50 population draws of 50 or 1000 in-sample individuals and 2000 out-of-sample individuals. We assessed in-sample and out-of-sample choice probability predictions using the method of sample enumeration where we average the choice share for a particular alternative across all the predicted individual choice probabilities for both in-sample and out-of-sample individuals, respectively.

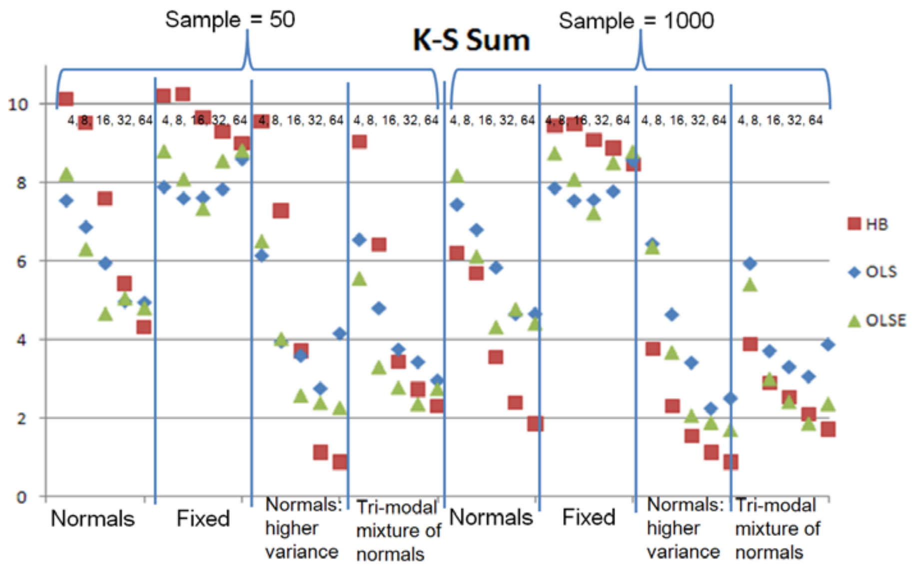
The results of the simulation study are shown in Figure 5 and Figure 6. These results show the changes in parameter recovery and changes in MSE for hold out choices when a) the number of choice sets increases, b) the sample size increases, or c) the underlying distribution of preferences changes.

The K-S sum in Figure 5 shows that for HB as the number of choice sets per individual increases, the overall ability to recover the parameter distribution increases. This holds for all underlying preference distributions, but it is significantly less pronounced when the underlying preferences are homogenous across individuals (i.e., fixed). In this case, HB only improves slightly with more observations per individual. It is also interesting to note that HB does a good job in recovering the tri-modal mixture of normal distributions. The K-S sum also shows that HB's ability to recover the true parameter distributions increases with sample size: For example, regardless of the underlying true distribution, if HB has 1000 rather than 50 individuals for estimation, the recovery of the true distribution improves. The latter result is interesting in its own right in so far as it has interesting implications for the number of choice sets per individual versus number of individuals trade-off in typical surveys used in field settings.

The ability of individual level models (OLS and OLSE) to recover the true parameter distribution improves, but this is not associated with increasing sample sizes. Whereas individual models recover the true parameter distribution better for small sample sizes, HB is better for large sample sizes. Interestingly, when there is no consumer heterogeneity in preferences (i.e., fixed parameters), the individual level models recovery of the true parameters exhibits a u-shaped relationship with the number of choice sets[2].

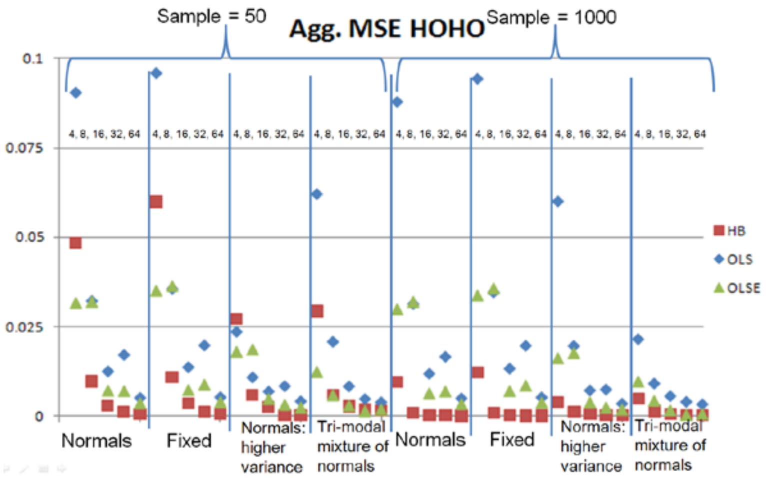
[2] It is important to note that the CDF in case of an underlying homogenous preference function is a step function.

**Figure 5: KS sum to measure of parameter distribution recovery of in-sample individuals versus true hold-out individuals**




Whereas the KS-sum is a measure of parameter distribution recovery, the aggregate MSE measures the deviation between predicted and true choice shares across individuals. Figure 6 shows this MSE measure when applied to individuals and choice sets not used for estimation; thus, it is ensures that we are not measuring overfitting. A value closer to zero indicates a better model prediction. Again, one can see that more choice sets improves the predictive ability of all models. However, when  there are few choice sets, HB does a much better job in predicting aggregate choice shares of individuals and choice tasks not used in the estimation. Figure 6 also shows that OLS performs poorly when only four choice sets are used for observation. Finally, Figure 6 also reveals that HB benefits also from larger sample sizes, whereas the individual level models do not.

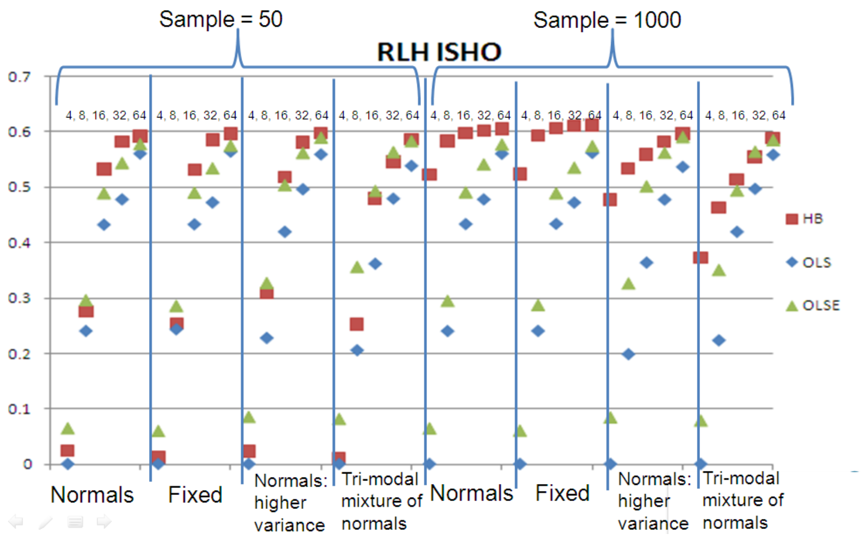
### Figure 6: Aggregate MSE for hold-out individuals and hold-out choice tasks




sho

The root likelihood shown in Figure 7 shows the ability of the models to predict the individuals' choices with a higher root likelihood (RLH) indicating a better fit. Figure 7 shows the RLH for the hold-out choices of the individuals used in the estimation. Again, ability improves with more choice sets; HB also improves with larger sample sizes.

### Figure 7: Root likelihood for in-sample individual hold-out choice tasks




## Empirical examples

We used OLS, WLS and HB to estimate model parameters and predict observed choices in four empirical data sets1) choices of car insurance options, 2) choices of cross-country airline flights; 3) delivered pizzas; and 4) digital cameras. In the airline and car insurance data we estimate the models from choices of 150 out of 200 individuals using 12 choice sets with four options each. The choice sets were constructed using the Street and Burgess (2007) optimal design approach. We use observed and predicted shares for each choice alternative in each choice set (48 total) to compare models in- and out-of-sample using root likelihood (RLH) and mean squared error (MSE) (150 individuals for in-sample; the remaining 50 individuals for out-of-sample). We used an additional four choice sets available to us to assess holdout task fits, again using RLH and MSE. To minimise bias due to particular individuals selected for estimation, we averaged our fit measures over ten different random draws of 150 individuals.

For the pizza and camera datasets we had 600 respondents with 20 choice sets each on which we estimated the models. For each of those in-sample respondents we also observed choices for five hold-out choice tasks. In addition, we collected choice data from 600 additional respondents only on these five hold-out choice tasks, i.e. for these respondents we had no information for the choice sets that were used for estimation.

**Table 1: Empirical data sets model fit**

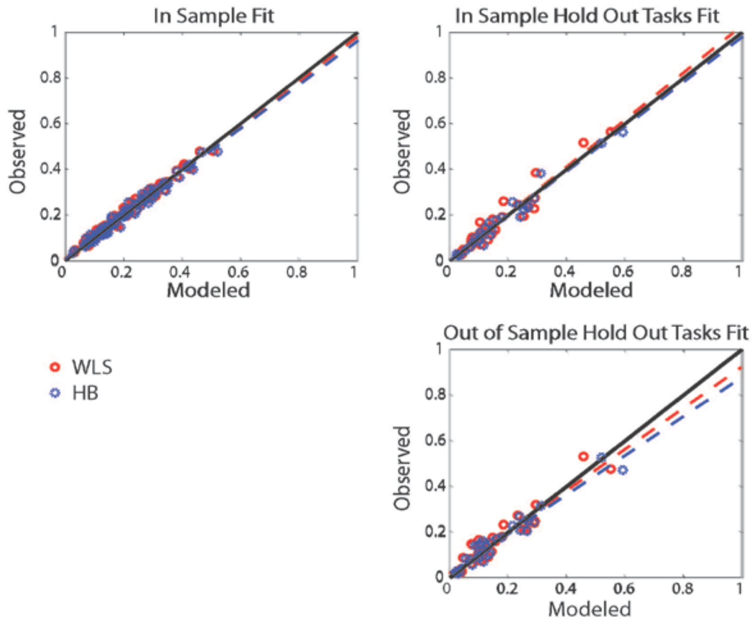
| | **WLS** | | **OLS** | | **HB** | |
|---|---|---|---|---|---|---|
| | **RLH** | **MSE** | **RLH** | **MSE** | **RLH** | **MSE** |
| **Pizza** | | | | | | |
| In-sample design sets | 0.67 | 0.000 | 0.67 | 0.000 | 0.72 | 0.000 |
| In-sample holdout sets | 0.19 | 0.001 | 0.33 | 0.003 | 0.42 | 0.001 |
| Out-of-sample design sets | | | | | | |
| Out-of-sample holdout sets | 0.21 | 0.002 | 0.20 | 0.003 | 0.21 | 0.001 |
| **Camera** | | | | | | |
| In-sample design sets | 0.63 | 0.001 | 0.62 | 0.001 | 0.67 | 0.001 |
| In-sample holdout sets | 0.06 | 0.001 | 0.22 | 0.002 | 0.29 | 0.001 |
| Out-of-sample design sets | | | | | | |
| Out-of-sample holdout sets | 0.21 | 0.002 | 0.21 | 0.003 | 0.22 | 0.001 |
| **Airline** | | | | | | |
| In-sample design sets | 0.77 | 0.000 | 0.76 | 0.001 | 0.80 | 0.001 |
| In-sample holdout sets | 0.28 | 0.004 | 0.39 | 0.005 | 0.40 | 0.001 |
| Out-of-sample design sets | 0.39 | 0.003 | 0.39 | 0.004 | 0.39 | 0.004 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Out-of-sample holdout sets | 0.38 | 0.005 | 0.38 | 0.007 | 0.39 | 0.004 |
| **Car Insurance** | | | | | | |
| In-sample design sets | 0.73 | 0.001 | 0.72 | 0.002 | 0.77 | 0.001 |
| In-sample holdout sets | 0.22 | 0.005 | 0.34 | 0.013 | 0.37 | 0.003 |
| Out-of-sample design sets | 0.34 | 0.004 | 0.34 | 0.005 | 0.34 | 0.004 |
| Out-of-sample holdout sets | 0.39 | 0.007 | 0.37 | 0.015 | 0.40 | 0.005 |

The results are summarised in Table 1, which reports the model performances for four cases: 1) in-sample, same people, same choice sets; 2) out of sample, same people, different choice sets; 3) out of sample, different people, same choice sets; and 4) out of sample, different people, different choice sets. Table 1 displays the root likelihood measure as and the aggregate mean squared errors of the choice share predictions.

Let us now consider the two performance measures: 1) Root likelihoods are always higher for HB when predicting the individuals who were used for estimation. However, there is little difference in WLS, OLS and HB when it comes to predicting hold-out individuals; 2) The mean squared error of predicted aggregate choice shares is uniformly low for all models and all situations, indicating that all models do a good job in predicting aggregate choice shares. This is nicely demonstrated in Figure 8 for the pizza category. Indeed, HB is only slightly better in the airline and car insurance categories.

**Figure 8: Aggregate choice share prediction HB versus WLS**




## Discussion and Conclusions

We proposed and investigated three simple ways to model individual choices, namely Ordinary Least Squares regression on expanded and non-expanded choice sets, and Weighted Least Squares regression as suggested by Louviere, et al (2008). We tested the performance of both estimation approaches against the popular HB approach to estimating Mixed Logit model using both simulations and four empirical datasets where we compared in- and out-of-sample performance measures. Our results suggest that OLSs and WLS models perform well in both real and simulated data once the parameter estimates are "corrected" by the inverse of the mean squared error for each individual. Because of simplicity and ease of use, both estimation approaches are likely to be attractive to researchers who want to better understand and model consumer choices, but find complex statistical models or black box commercial software challenging. In turn, this should allow both academics and practitioners to do a reasonable job of predicting choices without making big mistakes.

## References


Beggs, S., Cardell , S., & Hausman, J. (1981) Assessing the potential demand for electric cars. Journal of Econometrics, 16, 1-19.

Berkson, J. 1944. Application of the Logistic Function to Bio-Assay. Journal of the American Statistical Association 39, 357-365.

Chapman, R.G., Staelin, R. 1982. Exploiting Rank Order Choice Set Data Within the Stochastic Utility Model. Journal of Marketing Research 19, 288-301.

Chapman, R.G. 1984. An Approach to Estimating Logit Models of a Single Decision Maker's Choice Behavior. Advances in Consumer Research 11, 656-661.

Ebling C., Frischknecht, B., Louviere, J., 2010, “Keep It Simple: Easy Ways To Estimate Choice Models For Single Consumers,” Australian and New Zealand Marketing Academy Conference 2010, Christchurch, New Zealand.

Fiebig, D.G., Keane, M.P., Louviere, J., Wasi, N. 2010. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Marketing Science 29, 393-421

Finn, A., Louviere, J.J. 1992. Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. Journal of Public Policy and Marketing 11 (1) 12-25.

Green, P.J. 1984. Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. Journal of the Royal Statistical Society, Series B (Methodological) 46(2), 149-192.

Heckman, J.J., Snyder, J.M. 1997. Linear Probability Models of the Demand for Attributes With an Empirical Application to Estimating the Preferences of Legislators. Rand Journal of Economics 28, 142-189.

Hensher, D.A., Louviere, J.J. 1983. Identifying Individual Preferences for International Air Travel: An Application of Functional Measurement Theory. Journal of Transport Economics and Policy 17(3), 225-245.

Kamakura, W. A. & Russell, G. J. 1989. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research, Vol. 26, No. 4, 379-390.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and applications* Cambridge University Press.

Louviere, J.J., Woodworth, G. 1983. Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. Journal of Marketing Research 20, 350-367.

Louviere, J.J., Street, D., Burgess, L., Wasi; N., Islam, T., Marley, A.A.J. 2008. Modeling the Choices of Individual Decision-Makers by Combining Efficient Choice Experiment Designs with Extra Preference Information. Journal of Choice Modelling 1(1), 128-163.

Luce, R.D. 1959. Individual Choice Behavior: A Theoretical Analysis. New York: Wiley.

Luce, D., Suppes, P. 1965. Preference, Utility, and Subjective Probability. In Bush, R.R., Galanter, E. (eds.), Handbook of Mathematical Psychology, 249-410. John Wiley and Sons Inc., New York.

Marley A.A.J., Louviere, J.J. 2005.Some Probabilistic Models of Best, Worst, and Best-Worst choices. Journal of Mathematical Psychology 49, 464-480.

Marshall, D., Chan, S., & Curry, J. (2011). A head-to-head comparison of the traditional (top-down) approach to choice modeling with a proposed bottom-up approach. 2010 Sawtooth Software Conference, Newport Beach, CA. 309-319.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (ed.), Frontiers In Econometrics, Academic Press: New York, 105-142.

Rossi, P., Allenby, G., McCulloch, R. (2006) Bayesian Statistics and Marketing. Wiley.

Street, D., Burgess, L. 2007. The Construction of Optimal Stated Choice Experiments: Theory and Methods. New York: Wiley.

Swait, J.D., Louviere, J.J. 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. Journal of Marketing Research 30, 305-314.

Theil, H. 1969. A Multinomial Extension of the Linear Logit Model. International Economic Review 10, 251-259.

Thurstone, L.L. 1927. A Law of Comparative Judgement. Psychological Review 34, 273-286.

Train, K. (2003) Discrete choice methods with simulation. Cambridge University Press.