How Indecisiveness in Choice Behaviour affects the Magnitude of Parameter Estimates obtained in Discrete Choice Models

Abstract

Parameter estimates ($\hat{\beta}$) obtained in discrete choice models are confounded with the scale parameter of the logistic distribution. The scale parameter captures the indecisiveness (or errors) made by individuals in the choice of experiment alternatives. Therefore the magnitude of $\hat{\beta}$ is determined by both location (mean) and scale (indecisiveness) effects. This has important implications for marketing research. Variability in $\hat{\beta}$ across sub-groups (e.g. markets, segments or research samples) is often used to indicate differences in preference. However, these differences may be due simply to scale, and not average preference. Incorrect interpretation could result in the adoption of inappropriate marketing strategies. Through the analysis of two simple choice experiments, this paper illustrates how errors in individual choice behaviour lead to a logistic distribution of choice probabilities, and the consequent counfounding of these errors with $\hat{\beta}$. The sampling distribution of $\hat{\beta}$ is identified, and further research recommended into its implication for Discrete Choice Modelling.

Introduction

In the interpretation of discrete choice models, differences in estimated beta parameter estimates ($\hat{\beta}$), say across segments, markets or research studies, are commonly used to indicate differences in consumer preference. Unfortunately, the correct interpretation is not so straightforward. Apparent differences in $\hat{\beta}$, rather than being simply an indicator of relative preference (location effects), may instead be the result of the indecisiveness, or inconsistency, of the subjects in the choice experiment (scale effects). The objective of this paper is to highlight this relationship between indecisiveness and $\hat{\beta}$, and illustrate it with data collected from two simple choice experiments.

The implication of this relationship for effective marketing decision-making is considerable. For example, a marketer may be interested in whether one consumer segment prefers a product more than another, or whether price sensitivity varies between markets. In both these situations it is insufficient simply to compare $\hat{\beta}$s. In the Multinomial Logit (MNL) model, $\hat{\beta}$ is confounded with the estimated scale parameter of the logistic distribution ($\hat{b}$). Consequently, $\hat{\beta}$ can be large either because the alternative in the choice experiment is preferred on average more often or because the decisiveness (or confidence) with which the choice was made in the experiment was high. Scale effects might therefore cause (false) differences to be concluded in segment product preference or market price sensitivity, resulting in the adoption of inappropriate marketing strategies. Instead, the accurate comparison of preference between sub-groups from estimated MNL models (and all other limited dependent-variable models), requires consideration of both location and scale effects so that the impact of scale can be removed from $\hat{\beta}$. Louviere and Swait (1996) have discussed an approach for dealing with these comparisons and this is recommended when sub-group comparisons are being made, for at present independent estimation of scale and location effects is not possible using existing MNL procedures available in commonly used marketing research software. For the sake of clarity it is noted that scale effects are irrelevant if no comparison of $\hat{\beta}$ is being made across sub-groups. In this simpler (and more common) one model situation, the magnitude of $\hat{\beta}$ is not being considered, and the confounding of $\hat{\beta}$ with $\hat{b}$ presents no problems in the interpretation of model outputs.

How Inconsistencies (Errors) in Preference lead to Variability in Parameter Estimates

Random Utility Theory postulates that utility can be separated into systematic (representative) and random (individual idiosyncrasies of taste) components. The utility U of the i^{th} alternative for the q^{th} individual can therefore be stated as

$$U_{iq} = V_{iq} + \varepsilon_{iq}$$

Utility maximising individuals are assumed to choose the alternative that yields the greatest utility. In other words, individual q will choose alternative i iff

$$V_{iq} + \varepsilon_{iq} > V_{jq} + \varepsilon_{jq} \ \forall \ j \neq i \in A, \quad \text{or by rearrangement,} \quad V_{iq} - V_{jq} > \varepsilon_{jq} - \varepsilon_{iq}$$

Rearranging the observables and the unobservables (errors) together highlights why the distribution of the difference of the two error components (ε_j - ε_i) is so fundamental to choice modelling. Of course the analyst does not know (ε_j - ε_i), so a probability that (ε_j - ε_i) < (V_i - V_j) has to be assumed. In MNL the assumption is that each error component is Independently and Identically Distributed (IID) Extreme Value Type I (EV1). This paper focuses on examining this distribution of the error terms, and illustrates why the EV1 is often an appropriate distributional assumption, leading to a logistic distribution of choice probabilities and an ensuing variability of $\hat{\beta}$ determined by this distribution's scale parameter.

Real market choice behaviour, however, is complex and there are many sources of error that are captured by ε. So to demonstrate the effect of ε on $\hat{\beta}$ more clearly, a much simpler experimental context was chosen. Illustration of variability in $\hat{\beta}$ arising from this situation would suggest an amplified version in the more complex real-world environment. Instead of looking at utility maximisation (preference) behaviour, the simpler case of colour perception was examined in two experiments; that is, the choice of the '*most red*' in response to two boxes of colour presented on a computer screen. Again for simplicity, these experiments were designed to limit the source of error to 'within-subject' perception error - i.e. the difference between subjective perception and the objective characteristic of a presented alternative. Louviere, Hensher and Swait (2000) have stated the functional relationships implied by the choice process, and Table 1 sets out a modified version of their flowchart as a summary of the objectives of the two colour perception experiments.

Table 1: The Functional Relationships of the Colour Choice Process

Step	*Measure*	*Description*	*Assumed Error Distribution*
0	r_x	The measured level of red in colour *x*.	No distribution, as assumed to be measured without error.
1	$s_x = f(r_x)$	The perceived level of red in colour *x*	Exponential, ~Exp(θ,σ)
2	$u_x = g(s_x, s_y)$	The perception that s_x is more red than s_y	Extreme Value Type I ~EV1(ξ,θ)
3	$p(x) = h(u_x, u_y)$	The likelihood of colour u_x being chosen.	Logistic distribution ~Logistic(*m*,*b*)

The Two Experiments

The first experiment (Experiment 1) was designed to capture the distribution of errors in s_x (Step 1). The second experiment (Experiment 2) was designed to capture the distribution of errors in u_x (Step 2). Further analysis of Experiment 2 data was then conducted to understand the distribution of the choice probabilities *p*(*x*) (Step 3), and hence the distribution of $\hat{b}$ and $\hat{\beta}$ resulting from the preceding error distributions. Each treatment was a simple computer screen containing two boxes of colour, and the subject was asked to click the '*most red*' box. Both experiments had the same appearance, although the experimental design behind each was different. The treatments were evaluated by a single subject, on the same computer, to minimize uncontrolled sources of error. The two colours assigned to stimuli *A* and *B* can be defined as follows. Let U be the universe of all possible colours definable on an RGB scale. Let T ⊂ U be the subset of RGB colours chosen for the experiment. Let $\{x,y\} \in$ T be a pair of colours presented to the subject as the experiment treatment. Let the colour of *x* be defined as $x = f(r,g,b)$, where *r*, *g* and *b* are red, green and blue, respectively. (Colours with varying levels of red can therefore be generated by altering the parameter *r*, while holding the parameters *g* and *b* constant at zero). Let *r* be measured on an interval scale of range [0,255].

Experiment 1: Errors in perception (Step 1)

In Experiment 1, the midpoint of the pair $\{x,y\}$ was fixed at $r = 60$, and treatments were constructed by varying *y-x* over the range [0,60]. Colours *x* and *y* were then randomly allocated to treatment stimuli *A* and *B*. Each possible difference *y-x* was randomly presented 100 times (N = 6,100). Figure 1 shows the distribution of errors in choosing the '*most red*', over the range of colour differences. It can be seen that the distribution of error is well approximated by an exponential distribution. The average rate of error was nine percent. The results of Experiment 1 provide an approximate distribution of perception error given mean-centred differences ($\bar{r} = 60$) in the actual level of red in the two colours. This distribution can now be used as a selection probability that yields a subset of colours S ⊂ T such that each element $x \in$ S enters the set in proportion to its likelihood of being perceived the same as $r = 60$. In this sense, S represents the domain of error for $r = 60$, and its use in Experiment 2 enables accurate measurement of errors made in the perception of a maximum when evaluating the function u_x. Its use is also necessary due to the effect of Weber's Law over different colour intensities. Fixed differences in colour become easier to discriminate between the more red they are, and hence a non-constant error variance exists over the colour spectrum. The subset S thus provided the appropriate experiment frame for Experiment 2.

Experiment 2: Errors in the perception of a maximum (Step 2)

This experiment focussed on the distribution of errors around the choice of the '*most red*' for pairs of colours randomly selected from the domain of error S; that is, the distribution of r_{chosen} - $r_{notchosen}$ when r_{chosen} is not an objective maximum. Experimental treatments were designed by randomly selecting pairs of colours from subset S. As assumed in MNL, Figure 2 shows that this distribution is well approximated by the EV1 distribution. This is not surprising, as Johnson, Kotz and Balakrishnan (1995) have shown that the EV1 distribution *is* the distribution of sample maxima taken from a parent distribution whose tail decreases at least at an exponential rate. This result has therefore captured the distribution of the error component ε_j - ε_i where V_i was perceived a maximum. A total of N = 7,000 treatments were evaluated. The average error rate was 50%.

Fig.1: Plot of % Error in Experiment 1

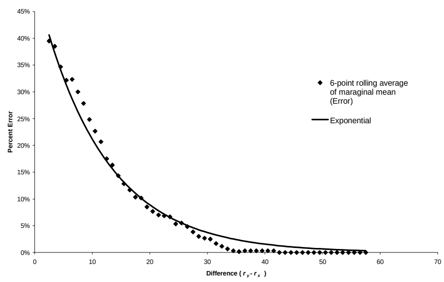
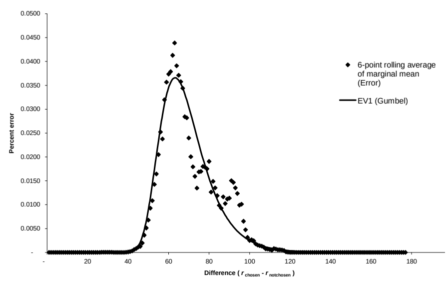


Fig.2: Plot of % Error in Experiment 2



The choice of A or B: Errors in the choice of an alternative (Step 3)

Experiment 2 demonstrated the use of the EV1 distribution to approximate the distribution of errors around r_{chosen} - $r_{notchosen}$. However, in the MNL choice model we are interested in the likelihood of a particular alternative being chosen: the probability (in Step 3) that *x* is chosen given the perception $u_x > u_y$. Johnson, Kotz and Balakrishnan (1995) have shown that the difference between two IID EV1 variables is the logistic distribution. This is supported both in the marginal means plot of the choice of *A* given the difference $r_B - r_A$ (Figure 3) and in the linearity of the plot of estimated regression coefficients versus the midpoints of sextiles of this difference (Figure 4). Although treatments *A* and *B* were the '*most red*' with equal frequency, there was approximately a 60/40 preference for *A*, possibly due to the subject's left-eye dominance, or the slight relative ease of centre-to-left over centre-to-right mouse movements.

Fig.3: Plot of % Choice of Treatment *A*

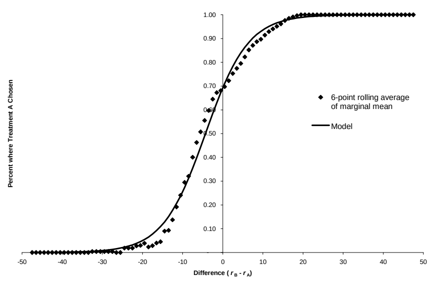
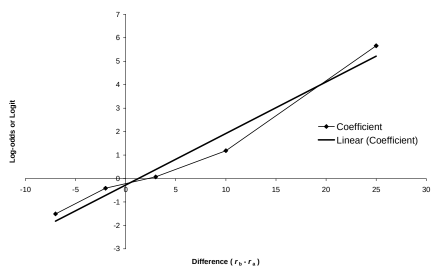


Fig.4: Plot of $\hat{\beta}$ for Sextiles of $r_B - r_A$



We note that the data support the use of a logistic distribution of error differences (ε_B - ε_A) as close first approximations appear to be consistent with the theory as postulated. Subsequent examination of the functional form of the logistic regression model $\pi(x)$ and its linear form created through the logit link function g(x) [i.e. taking the log of the odds ratio], provides a notational explanation of how each β is confounded with the scale parameter.

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+B_p x_p)/b}}{1+e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+B_p x_p)/b}}, \quad \text{where} \quad g(x) = \lambda\beta_0 + \lambda\beta_1 x_1 + \lambda\beta_2 x_2 + \cdots \lambda\beta_p x_p, \quad \text{given} \quad \lambda = \frac{1}{b}$$

The variability of parameter estimates – the effect of lambda

Finally, the collected data can be treated as a population, and a series of MNL models run on subsets of the data, simulating a hypothetical segment, market or multiple-study situation. By taking repeated random samples (N = 10,000), with replacement, the sampling distribution of $\hat{b}$ and $\hat{\beta}$ was bootstrapped. Figures 5 and 6 detail these results, indicating that $\hat{b}$ is distributed normally, and $\hat{\beta}$ is distributed EV1 with a large variability in the estimated parameter.

Fig.5: Bootstrap Plot of $\hat{b}$

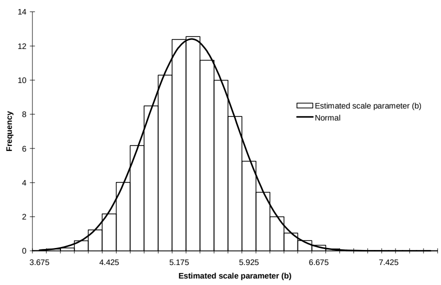
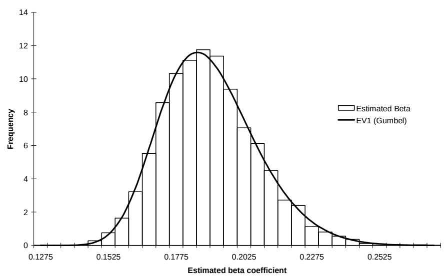


Fig.6: Bootstrap Plot of $\hat{\beta}$



Limitations, Conclusions and Further Research

Although a high level of replication for each subject is appropriate for examining within-subject error, a limitation of this study is that it was conducted on a single subject only. These results would therefore benefit from demonstration of consistency of the error distributions (although not the estimated parameters) across individuals. Despite this, it is hoped that sufficient evidence has been provided to conclude that $\hat{\beta}$s obtained in MNL models are confounded with the scale parameter of the logistic distribution, and that the errors captured in the scale parameter dramatically impact the magnitude of $\hat{\beta}$. The sampling distribution of $\hat{\beta}$ has not been widely considered in discrete choice modelling, and further research is recommended to understand the implication of this result.

References

Hosmer, D., Lemeshow, S., 2000. Applied Logistic Regression, 2nd ed. New York: John Wiley & Sons.

Johnson, N., Kotz, S., & Balakrishnan, N., 1995. Continuous Univariate Distributions Volume 2, 2nd ed. New York: John Wiley & Sons.

Louviere, J., Hensher, D., & Swait, J., 2000. Stated Choice Methods. Cambridge: University Press.

Louviere, J., Swait, J., 1996. Searching for Regularities in Choice Processes, or the Little Constant that Could. Working paper, Sydney: University of Sydney, Australia.

Luce, R., 1959. Individual Choice Behaviour. New York: John Wiley & Sons