# Australian School of Business

# Working Paper



Australian School of Business Research Paper No. 2012 ECON 47

## The confounding effects of consumer heterogeneity on model-based inference of attribute non-attendance

Hong il Yoo

# The confounding effects of consumer heterogeneity on model-based inference of attribute non-attendance

HONG IL YOO

*School of Economics, Level 4, Australian School of Business Bldg.*

*University of New South Wales, Sydney, NSW 2052. Australia*

*(e-mail: h.yoo@unsw.edu.au)*

15 June 2012

## Abstract

Several empirical studies conclude that a majority of economic agents ignore some of observed product attributes when choosing among discrete alternatives. Many of these findings are based on latent class logit with partially constrained support points wherein the share of each point is interpreted as the probability of ignoring particular attribute(s). We note that because the logit kernel is mixed over these points to approximate unmodeled interpersonal taste variation during the estimation stage, the interpretation of estimated shares is necessarily ambiguous. Using simulated examples, we explain why common forms of unobserved consumer heterogeneity can be confounded with attribute non-attendance.

**JEL classification:** C25, C52, C81

**Keywords:** attribute non-attendance, gmnl, latent class, consumer heterogeneity, mixed logit, information processing rule

**Word count:** <3300 words

# I. Introduction

Econometric models of discrete choice are usually specified and interpreted as if economic agents derive utility directly from product attributes, and compare all attributes across all available products during the utility maximization process. Such behavioral hypothesis has been challenged by practitioners who conjecture instead that agents may adopt various decision strategies to simplify choice tasks. For example, the well-established literature on consideration set formation (Haab and Hicks, 1997; Chiang *et al.*, 1998; Gilbride and Allenby, 2004; Li and Trivedi, 2012) investigates screening rules whereby some products are excluded from detailed evaluation as candidates for the final choice, while a rapidly growing body of literature (Arana *et al.*, 2008; Scarpa *et al.*, 2009; Hensher and Greene, 2010; Campbell, Aravena and Hutchinson, 2011; Campbell, Hensher and Scarpa, 2011; Hole, 2011; Hensher *et al.*, 2012; Lagarde, 2012) analyzes attribute non-attendance (ANA) which arises when some of the attributes observable to the modeler are ignored by agents who evaluate a given set of products.

Several extensions of multinomial logit (MNL) have been proposed in the latter literature to accommodate the incidence of ANA. A popular empirical approach involves specification of latent classes ignoring different subsets of attributes by constraining utility weights on the ignored attributes to zero, and testing the resulting model against MNL. The population share of each class is estimated along with unconstrained utility weights to infer from observed choices the prior probability that each subset is ignored. The related studies have analyzed data from stated choice experiments, following much of the recent literature on choice modeling in environmental economics, health economics and transportation research. The estimated incidence of ANA is startling, with some studies finding less than a 10% chance of attending to all attributes (Scarpa *et al.*, 2009; Hole, 2011; Campbell, Hensher and Scarpa, 2011) and many reporting substantial shares of classes non-attending to multiple attributes simultaneously. Taken at face value, these findings warrant a further investigation into whether the prevalence of ANA is an artefact of eliciting hypothetical choices or reflects heuristics also applied in real-life choice situations.

This paper questions the reliability of model-based inference of ANA in the latent class framework and discusses two potential pitfalls associated with it. First, ANA becomes analytically indistinguishable from cases where agents attend to all attributes but genuinely find changes in some attributes irrelevant to their choices. Second and more importantly, ANA may be empirically confounded with unmodeled consumer heterogeneity which does not necessarily induce zero utility weights. The fundamental source of confounding is clear. A latent class ANA model is, by McFadden and Train (2000)'s definition, a mixed logit model with discrete mixing. The particular discrete distribution operationalizing ANA is less flexible than discrete mixtures considered by Train (2008) but can still approximate interpersonal taste variation. Accordingly, the significant class shares and the superior fit of latent class ANA models over MNL should be interpreted broadly as evidence on the importance of specifying a flexible model. The results do not necessarily constitute evidence that ANA is a widespread behavioral phenomenon.

The remainder of this paper is organized as follows. Section 2 critically reviews the latent class approach to modeling ANA. Section 3 presents simulated examples to explain why ANA can be confounded with scale heterogeneity conceptualized by Fiebig *et al.* (2010) and traditional coefficient heterogeneity. Section 4 concludes.

# II. Models

Suppose that each of $N$ agents chooses amongst $J$ alternatives on $T$ different choice occasions. Alternative $j$ in agent $n$'s choice set $t$ is described by a $K$-vector of attributes, $\mathbf{x}_{njt}$. The probability of observing her actual choice, $P_{nt}(\boldsymbol{\beta})$, depends on her utility weights on these attributes, $\boldsymbol{\beta}$, and is assumed to take the MNL form; if alternative $i$ has been chosen, $P_{nt}(\boldsymbol{\beta}) = \exp(\boldsymbol{\beta} \cdot \mathbf{x}_{nit})/[\sum_{j=1}^{J} \exp(\boldsymbol{\beta} \cdot \mathbf{x}_{njt})]$.

In the latent class framework for modeling ANA, different classes of agents are assumed to consider different subsets of $\mathbf{x}_{njt}$, and the class ignoring a particular subset is defined by constraining the corresponding elements in $\boldsymbol{\beta}$ to 0. For example, non-attendance to "fuel cost" and "seating capacity" of an automobile is accommodated by constraining the utility weights on these attributes to 0. We use $c = 1, 2, \cdots, C$ to index the classes, where $C \leq 2^K$. The unconditional likelihood of agent $n$'s sequence of choices is specified as:

$$P_n(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_C, \rho_1, \cdots, \rho_C) = \sum_{c=1}^{C} \rho_c \prod_{t=1}^{T} P_{nt}(\boldsymbol{\beta}_c) \tag{1}$$

where $\boldsymbol{\beta}_c$ collects utility weights for class $c$, $\rho_c$ is the population share of this class, and $\sum_{c=1}^{C} \rho_c = 1$. All coefficients in $\boldsymbol{\beta}_1$ are unconstrained, as class 1 is assumed to consider all attributes or "fully attend". A common practice is to constrain the utility weight on an attribute to be identical across all classes attending to it, assuming that they have the same underlying tastes. With $K = 2$, four possible classes include $\boldsymbol{\beta}_1 = (\beta_A, \beta_B)$, $\boldsymbol{\beta}_2 = (\beta_A, 0)$, $\boldsymbol{\beta}_3 = (0, \beta_B)$ and $\boldsymbol{\beta}_4 = (0, 0)$.

The model specification described so far involves $K$ utility weights and up to $C - 1$ class shares, and has been implemented by Scarpa *et al.* (2009), Campbell, Hensher and Scarpa (2011), Hensher *et al.* (2012) and Lagarde (2012). We use ANA-MNL to denote this specification and a similar model due to Hole (2011). The latter imposes restrictions $\rho_c = \prod_{k \in \Omega_c} \pi_k \prod_{k \notin \Omega_c} (1 - \pi_k)$ for each class $c$, where $\pi_k$ is the binary probability of attending to attribute $k$ and $\Omega_c$ collects indices for the attributes class $c$ considers. Arana *et al.* (2008) and Hensher and Greene (2010) analyze ANA along with other decision strategies after allowing for some form of preference heterogeneity, while Campbell, Aravena and Hutchinson (2011) examine whether the cheaper of two alternatives is more likely to be subject to ANA; the crux of our subsequent discussion applies to these studies too because they model and interpret the incidence of ANA in essentially the same way as the ANA-MNL applications.

It is analytically clear that ANA-MNL would overestimate the true incidence of ANA whenever some agents fully attend but find changes in certain attributes irrelevant to their choices. That is, whenever some agents genuinely have the utility weights of 0 on those attributes. For instance, many non-Muslim agents would find the Halal status of a breakfast cereal irrelevant in a cereal choice experiment or alternatively, some agents may not respond to the observed range of variation in an attribute due to discontinuity in preferences.

A major threat to the model-based inference of ANA, however, comes not from these special cases but from a general form of interpersonal variation in tastes, $\boldsymbol{\beta}$. Consider a mixed logit (MIXL) model (McFadden and Train, 2000) which has driven much of the recent literature on modeling individual heterogeneity.[1] The MIXL likelihood of agent $n$'s choices is usually specified as in Revelt and Train (1998):

$$P_n(\boldsymbol{\theta}) = \int \prod_{t=1}^{T} P_{nt}(\boldsymbol{\beta}) f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta} \tag{2}$$

where density $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ describes as a function of $\boldsymbol{\theta}$ the mixing distribution that captures

[1] Empirical evidence to date suggests that the amount of unobserved taste heterogeneity in choice data tends to be substantial. See Train (2009) and references therein.

interpersonal variation in $\boldsymbol{\beta}$. When $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ is discrete, MIXL becomes latent class logit (LCL) with the likelihood identical to equation (1) except no constraint is placed on any $\boldsymbol{\beta}_c$. ANA-MNL is thus LCL (also MIXL) wherein a behavioral hypothesis motivates a particular specification of discrete mixture. This link tends to be noted only in the context of implementing estimation, but carries far-reaching implications for interpreting the estimation results.

ANA can be empirically confounded with unmodeled taste heterogeneity as ANA-MNL is estimated, even when no agent has a zero utility weight on any attribute. Most of modeling and survey design implications derived from the related studies depend on the sharp interpretation of estimated class shares as estimated frequencies of ignoring different attributes. But the discrete mixture specified for ANA-MNL is not exempt from the general property that any mixing distribution potentially captures the structure of unobserved heterogeneity other than what it is motivated to capture (Cherchi and Ortuza, 2010). The class shares $(\rho_1, \cdots, \rho_C)$ estimated for ANA-MNL in equation (1) would reflect how best the MNL likelihood, $\prod_{t=1}^{T} P_{nt}(\boldsymbol{\beta}_c)$, can be mixed over partially constrained support points $(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_C)$ to approximate mixing over the true distribution of individual heterogeneity. As we illustrate in Section 3, that the support points operationalize a specific behavioral hypothesis does not preclude such approximation. The economic significance of these shares, and the consequential improvement of model fit and changes of substantive results in comparison with MNL, should be interpreted broadly as evidence on the importance of specifying a flexible model, not that of modeling and quantifying ANA.

It may be tempting to suggest that potential confounding can be avoided by estimating an extension of ANA-MNL with the following likelihood:

$$P_n(\boldsymbol{\theta}, \rho_1, \cdots, \rho_C) = \sum_{c=1}^{C} \rho_c \int \prod_{t=1}^{T} P_{nt}(\boldsymbol{\beta}_c^*) f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta} \tag{3}$$

where $\boldsymbol{\beta}_c^* = \mathbf{d}_c \cdot \boldsymbol{\beta}$ and $\mathbf{d}_c$ is a $K$-vector of zero-one binary indicators which equal 0 for the attributes ignored by class $c$. Now ANA is seemingly distinguished from taste heterogeneity because the latter is described by $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ while the former is captured by a separate discrete mixture. The problem still remains, however, unless all interpersonal taste variation can be adequately accommodated by $f(\boldsymbol{\beta}|\boldsymbol{\theta})$. The model in equation (3) is equivalent to MIXL using a more flexible mixture than $f(\boldsymbol{\beta}|\boldsymbol{\theta})$, and may achieve a significant improvement in log-likelihood over MIXL using $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ alone even when all agents fully attend. The situation as such can arise, for example, if $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ is specified as a unimodal distribution when the true taste distribution is bimodal.

# III. Simulated examples

That a substantial proportion of economic agents are estimated to ignore some attributes is often highlighted as a key finding in studies adopting the latent class approach to modeling ANA. In fact, this approach loses much of its appeal unless the relative incidence of ANA *per se* is of interest to the researcher. Among MIXL models, ANA-MNL in equation (1) and its potential extension in equation (3) are less flexible *a priori* than alternative specifications which require a similar amount of computational effort, for example those analyzed by Train (2008), because the support points must be partially constrained to the origin and relative to one another. On the other hand, ANA as defined for ANA-MNL and its variants can be accommodated by more flexible MIXL models with positive coefficient densities at and/or near the origin.

In this section, we use three simulated data sets to illustrate how unmodeled interpersonal heterogeneity may mislead inference on the incidence of ANA. The data generating process (DGP) for each example is a special case of Fiebig *et al.*'s (2010) generalized multinomial logit II (GMNL-II).[2] We simulate $N = 300$ utility-maximizing agents who choose among $J = 2$ alternatives on each of $T = 10$ choice occasions. Each alternative is described by two observed attributes, $x_{Anjt}$ and $x_{Bnjt}$, drawn independently from the standard normal distribution. The utility agent $n$ derives from alternative $j$ on occasion $t$ is:

$$U_{njt} = \sigma_n(\beta_{An}x_{Anjt} + \beta_{Bn}x_{Bnjt}) + \varepsilon_{njt} \tag{4}$$

where the idiosyncratic error $\varepsilon_{njt}$ is independently type I extreme value distributed,

[2] In usual MIXL models following the tradition of Revelt and Train (1998), a coefficient can be decomposed into its population mean and agent-specific random deviation around it. GMNL-II specifies an agent-specific positive random parameter multiplying both components of all coefficients. This parameter is motivated by, but need not be tied to, scale heterogeneity discussed in Section 3.1. Fiebig *et al.* (2010) note that random scaling can be more broadly justified as a parsimonious approach to increase flexibility of an initially specified coefficient distribution. They also formulate GMNL-I, wherein only the mean components are randomly scaled, and GMNL, which nests GMNL-I and GMNL-II.

$\sigma_n$ is a positive scalar describing the agent-specific scale of utility, $\beta_{An}$ and $\beta_{Bn}$ are agent-specific coefficients on attributes $A$ and $B$ respectively. The scale and coefficient heterogeneity parameters follow different specifications across data sets, as discussed later. We use the same realized draws of $x_{Anjt}$, $x_{Bnjt}$ and $\varepsilon_{njt}$ throughout all examples, to focus on the confounding effects of different forms of heterogeneity while controlling for variations in observed and unobserved attributes.

Table 1 summarizes four models estimated for each data set: multinomial logit (MNL), Fiebig *et al.* (2012)'s GMNL-II with lognormal-normal mixing, the most common latent class MNL tailored to capture ANA (ECL-MNL) and finally a similar model, due to Hole (2011), with additional constraints on class shares (EAA-MNL).[3]

## Attribute non-attendance and scale heterogeneity

As well known, economically relevant coefficients in nonlinear choice models are inversely proportional to the unnormalized variance of the idiosyncratic error. Interpersonal scale heterogeneity arises when this variance differs across agents, for example because the influence of unobserved attributes relative to that of observed attributes varies.

Table 2 reports estimation results for the simulated data featuring only this form of heterogeneity. DGP is Fiebig *et al.*'s (2010) Scaled MNL (SMNL). We specify $(\beta_{An}, \beta_{Bn}) = (3, 1)$ for all $n = 1, 2, \cdots, 300$ so that there is no real incidence of ANA. $\sigma_n$ varies across $n$ and is drawn from a log-normal distribution with mean $-0.5\tau^2$ and variance $\tau^2$, where $\tau = 1.5$; $\sigma_n$ thus equals 1 in expectation. The chosen level of $\tau$ in relation to the magnitude of $(\beta_{An}, \beta_{Bn})$ reflect empirical SMNL estimates reported in Fiebig *et al.* (2010).

GMNL-II obtains the best BIC and Wald test statistics correctly point towards SMNL. MNL is conclusively rejected in favour of EAA-MNL and ECL-MNL at any conventional significance level using a likelihood ratio test; the latter two models also provide a substantial improvement in BIC. Both models suggest that a large fraction of agents ignore at least one attribute. ECL-MNL indicates that the vast majority either fully attend (55.5%) or fully non-attend (31.3%). EAA-MNL suggests that 38% fully

[3] ECL stands for 'equality constrained latent class' (Scarpa *et al.*, 2009). Hole (2011) dubs his variation endogenous attribute attendance (EAA) model.

attend and 30.5% ignore attribute $B$ only, with the rest being almost evenly split over the other two classes. The two sets of ANA-MNL results differ because EAA-MNL uses constrained class shares; to describe the high probability of full attendance, $\pi_A\pi_B$, both binary probabilities of attending to $A$, $\pi_A$, and $B$, $\pi_B$, need to be reasonably high and in consequence the estimated probabilities of attending to $A$ only, $\pi_A(1-\pi_B)$, and $B$ only, $(1-\pi_A)\pi_B$, are higher than their ECL-MNL counterparts.

But in this data set, neither the gain in fit over MNL nor the precisely estimated shares can be explained by the real incidence of ANA. Why is scale heterogeneity confounded with ANA? ECL-MNL is MIXL with discrete mixing over four partially constrained support points, $\boldsymbol{\beta}_1 = (\beta_A, \beta_B)$, $\boldsymbol{\beta}_2 = (\beta_A, 0)$, $\boldsymbol{\beta}_3 = (0, \beta_B)$ and $\boldsymbol{\beta}_4 = (0, 0)$ where $\beta_A$ and $\beta_B$ are the coefficients on attributes $A$ and $B$. Now suppose that $\beta_A > \beta_B > 0$. Given the DGP, the economically relevant coefficient vector takes form $\sigma_n(3, 1)$ for all agents. When $\sigma_n$ is close to or larger than unity, agent $n$'s choice behavior may be best described as full attendance, $\boldsymbol{\beta}_1$, while when $\sigma_n$ is close to zero, agent $n$'s choice behavior resembles full non-attendance, $\boldsymbol{\beta}_4$. When $\sigma_n$ is much smaller than unity and larger than zero simultaneously, say 0.3, agent $n$'s scaled coefficient vector $(1, 0.3)$ somewhat resembles non-attendance to attribute $B$ only, $\boldsymbol{\beta}_2$, and exactly corresponds to a linear combination of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_4$. Mixing the MNL likelihood mainly over $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_4$ with $\beta_A > \beta_B > 0$ thus provides an approximation to mixing over the true distribution of scale heterogeneity, resulting in large probability masses at these points. Essentially the same explanation holds for EAA-MNL, except that $\boldsymbol{\beta}_3$ must obtain a non-negligible share too as a consequence of the constrained class shares.

## Attribute non-attendance and coefficient heterogeneity

Table 3 presents estimates for the second data set featuring traditional coefficient heterogeneity as modeled by Revelt and Train (1998). DGP is MIXL with independently normal mixing and there is no scale heterogeneity: $\sigma_n = 1$ for every $n$ and $\beta_{An}$ is drawn from $N(3, 1)$ while $\beta_{Bn}$ is drawn from $N(1, 1)$, where $N(m, v)$ denotes a normal distribution with mean $m$ and variance $v$. Hence, $\beta_{An}$ is positive except for extremely rare cases, while $\beta_{Bn}$ can be negative for a non-negligible proportion of agents. For a motivating example, consider attribute $A$ as $-1$ times price and $B$ as the length of stay in a holiday package choice problem. The real incidence of ANA is almost ruled out because the probability of drawing a coefficient of 0 from each distribution is 0.

GMNL-II achieves the best BIC and aptly fails to find significant scale heterogeneity, while each of EAA-MNL and ECL-MNL obtains a very significant improvement in log-likelihood over MNL. Both ANA-MNL models suggest that about 56% of agents fully attend while 41% consider attribute $A$ only. The model-based inference on the incidence of ANA appears contaminated by coefficient heterogeneity, much as it was by scale heterogeneity.

The confounding effects of coefficient heterogeneity can be explained by typical draws from the coefficient distribution. As before, suppose that $\beta_A > \beta_B > 0$. Since $\beta_{An} \sim N(3, 1)$, this coefficient lies between 1 and 5 for about 95% of agents. Classes non-attending to attribute $A$, namely $\boldsymbol{\beta}_3 = (0, \beta_B)$ and $\boldsymbol{\beta}_4 = (0, 0)$, become poor support points in consequence and obtain negligible shares. Now because $\beta_{Bn} \sim N(1, 1)$, this coefficient takes negative or sufficiently small positive values for a nontrivial fraction of agents, whose behavior would then be best described as non-attendance to $B$ only, $\boldsymbol{\beta}_2 = (\beta_A, 0)$. For somewhat over 50% of agents, however, $\beta_{Bn}$ would be greater than 1 or slightly below it, making their behavior best captured by full attendance, $\boldsymbol{\beta}_1 = (\beta_A, \beta_B)$. Mixing the MNL likelihood mainly over $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ with $\beta_A > \beta_B > 0$ thus provides an approximation to mixing over the true distribution of coefficient heterogeneity.

## Attribute non-attendance and scale & coefficient heterogeneity

Table 4 reports estimation results for the third data set featuring both scale and coefficient heterogeneity. Neither form of heterogeneity can be ruled out *a priori* and this set-up arguably mimics the real choice data most closely. DGP is GMNL-II which combines SMNL and independent normal MIXL specified above: $\ln \sigma_n \sim N(-\frac{1}{2}1.5^2, 1.5^2)$, $\beta_{An} \sim N(3, 1)$ and $\beta_{Bn} \sim N(1, 1)$. For every $n$, we use the same realization of each parameter as drawn earlier.

GMNL-II unsurprisingly obtains the best BIC but the estimated standard deviation of $\beta_{An}$ is economically and statistically insignificant; with the coefficient's tight population distribution, it is empirically difficult to disentangle the coefficient deviation parameter, $\sigma_{\beta_A}$, from the scale variation parameter, $\tau$. EAA-MNL and ECL-MNL are again preferred to MNL using a likelihood ratio test, and indicate significant incidence of ANA. The pattern of ANA suggested by ECL-MNL is as expected; 35% of agents attend to attribute $A$ only (due to confounding coefficient heterogeneity), 33.1% non-attend to both attributes (due to confounding scale heterogeneity), while almost all of the rest (31.5%) fully attend. EAA-MNL suggests a somewhat different pattern which can be expected from the constrained class shares; to describe the high frequency of attending to $A$ only, the probability of considering $A$, $\pi_A$, must be high and that of considering $B$, $\pi_B$, low, but to capture the high frequency of full attendance simultaneously, $\pi_B$ must remain reasonably large, requiring an offsetting increase in $\pi_A$.

The confounding effects of joint scale and coefficient heterogeneity can be understood intuitively as follows. When $\sigma_n$ moderately deviates from 1, the explanation provided in the preceding subsection still holds while when $\sigma_n$ is close to zero, agent $n$'s choice behavior resembles full non-attendance, $\boldsymbol{\beta}_4 = (0, 0)$. Two major possibilities exist when $\sigma_n$ is very large. Agent $n$'s behavior may be described as full attendance, $\boldsymbol{\beta}_1 = (\beta_A, \beta_B)$, when both $\beta_{An}$ and $\beta_{Bn}$ are positive while it may be better approximated as non-attendance to $B$ only, $\boldsymbol{\beta}_2 = (\beta_A, 0)$, when $\beta_{An}$ is positive but $\beta_{Bn}$ is negative. Mixing the MNL likelihood mainly over $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ with $\beta_A > \beta_B > 0$ thus provides an approximation to mixing over the joint distribution of scale and coefficient heterogeneity.

# IV. Discussion

Several recent studies adopt a latent class approach to modeling attribute non-attendance (ANA) and conclude that a majority of economic agents ignore one or more of the observed product attributes when making discrete choices. These studies formulate discrete mixture logit with a behaviorally motivated selection of support points wherein the share of each point is interpreted as the relative incidence of ignoring particular attribute(s). We note that because the logit kernel is mixed over these points to approximate unmodeled interpersonal taste variation empirically, the estimated shares cannot be interpreted as estimated incidence. Evidence from our simulated data shows that scale and coefficient heterogeneity indeed can be mistaken for ANA when this type of discrete mixture logit is estimated.

The general message of our analysis is that an estimated mixing distribution does not provide convincing statistics which can be narrowly associated with behavioral hypotheses motivating the specification of that distribution. While our immediate focus has been on model-based inference of ANA, this message and related discussion could prove useful in assessing similar attempts at quantifying the relative incidence of other information processing rules. For example, Fok *et al.* (2010) hypothesize that respondents assign arbitrary rankings to worst $H$ alternatives in a choice set when ranking all alternatives from best to worst, and formulate a latent class exploded logit model allowing $H$ to vary across classes. In this model, arbitary ranking assignment is operationalized by constraining all coefficients of the relevant MNL multiplicands in the exploded logit formula (p.157, Train, 2009) to 0, in the same manner as how full ANA is operationalized.

# References

Arana, J. E., Leon, C. J. and Hanemann, M. W. (2008). 'Emotions and decision rules in discrete choice experiments for valuing health care programmes for the elderly', *Journal of Health Economics*, Vol. 27, pp. 753-769.

Campbell, D., Aravena, C. D. and Hutchinson, W. G. (2011a). 'Cheap and expensive alternatives in stated choice experiments: are they equally considered by respondents?', *Applied Economics Letters*, Vol. 18, pp. 743-747.

Campbell, D., Hensher, D. A. and Scarpa, R. (2011). 'Non-attendance to attributes in environmental choice analysis: a latent class specification', *Journal of Environmental Planning and Management*, Vo. 54, pp. 1061-1076.

Cherchi, E. and Ortuza, J. D. (2010). 'Can mixed logit reveal the actual data generating process? Some implications for environmental assessment', *Transportation Research Part D: Transport and Environment*, Vol. 15, pp. 428-442.

Chiang, J., Chib, S. and Narasimhan, C. (1998). 'Markov chain Monte Carlo and models of consideration set and parameter heterogeneity', *Journal of Econometrics*, Vol. 89, pp. 223-248.

Fiebig, D. G., Keane, M.P., Louviere, J. and Wasi, N. (2010). 'The generalized multionomial logit model: Accounting for scale and coefficient heterogeneity', *Marketing Science*, Vol. 29, pp. 393-421.

Fok, D., Paap, R. and Van Dijk, B. (2010). 'A rank-ordered logit model with unobserved heterogeneity in ranking capabilities', *Journal of Applied Econometrics*, In press. DOI: 10.1002/jae.1223

Gilbride, T. J. and Allenby, G.M. (2004). 'A choice model with conjunctive, disjunctive and compensatory screening rules', *Marketing Science*, Vol. 23, pp. 391-406.

Haab, T. C. and Robert L. H. (1997). 'Accounting for choice set endogeneity in random utility models of recreation demand', *Journal of Environmental Economics and Management*, Vol. 34, pp. 127-147.

Hensher, D. A. and Greene, W. H. (2010) 'Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification', *Empirical Economics*, Vol. 39, pp. 413-426.

Hensher, D. A., Rose, J. M. and Greene, W. H. (2012). 'Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design', *Transportation*, Vol. 39, pp. 235-245.

Hole, A. R. (2011). 'A discrete choice model with endogenous attribute attendance', *Economics Letters*, Vol. 110, pp. 203-205.

Lagarde, M. (2012). 'Investigating attribute non-attendance and its consequences in choice experiments with latent class models', *Health Economics*, In press. DOI: 10.1002/hec.2824

Li, Q. and Trivedi, P. K. (2012). 'Medicare health plan choices of the elderly: a choice-with-screening model', *Journal of Business and Economic Statistics*, Vol. 30, pp. 81-93.

McFadden, D. and Train, K. (2000). 'Mixed MNL models for discrete response', *Journal of Applied Econometrics*, Vol. 15, pp. 447-470.

Revelt, D. and Train, K. (1998). 'Mixed logit with repeated choices: households' choices of appliance efficiency level', *Review of Economics and Statistics*, Vol. 80, pp. 647-657.

Scarpa, R., Gilbride, T. J., Campbell, D. and Hensher, D. A. (2009). 'Modelling attribute non-attendance in choice experiments for rural landscape valuation', *European Economic Review of Agricultural Economics*, Vol. 36, pp. 151-174.

Train, K. (2008). 'EM Algorithms for Nonparametric Estimation of Mixing Distributions', *Journal of Choice Modelling*, Vol. 1, pp. 40-69.

Train, K. (2009). *Discrete choice methods with simulation*, Cambridge University Press, New York.

# TABLE 1

*A summary of estimated models*

| *Model* | *Likelihood for agent n* | *Notes* |
|---|---|---|
| MNL | $\prod_{t=1}^{T} P_{nt}(\boldsymbol{\beta})$ | $\cdot\boldsymbol{\beta} = (\beta_A, \beta_B)$<br>$\cdot$Estimate $\beta_A$ and $\beta_B$. |
| GMNL-II | $\int\int\prod_{t=1}^{T} P_{nt}(\sigma\boldsymbol{\beta})g(\sigma\|\tau)f(\boldsymbol{\beta}\|\boldsymbol{\theta})d\boldsymbol{\beta}d\sigma$ | $\cdot\boldsymbol{\theta} = (\overline{\beta}_A, \overline{\beta}_B, \sigma_{\beta_A}, \sigma_{\beta_B})$<br>$\cdot\boldsymbol{\beta} = (\beta_A\, \beta_B) \sim N((\overline{\beta}_A, \overline{\beta}_B), diag(\sigma^2_{\beta_A}, \sigma^2_{\beta_B}))$<br>$\cdot \ln\sigma \sim N(-0.5\tau^2, \tau^2)$<br>$\cdot$Estimate $\tau$ and $\boldsymbol{\theta}$ |
| ECL-MNL | $\sum_{c=1}^{4} \rho_c \prod_{t=1}^{T} P_{nt}(\boldsymbol{\beta}_c)$ | $\cdot\boldsymbol{\beta}_1 = (\beta_A, \beta_B), \boldsymbol{\beta}_2 = (\beta_A, 0)$<br>$\boldsymbol{\beta}_3 = (0, \beta_B), \boldsymbol{\beta}_4 = (0, 0)$<br>$\cdot$Estimate $\beta_A$, $\beta_B$, $\rho_1$, $\rho_2$ and $\rho_3$. |
| EAA-MNL | $\sum_{c=1}^{4} \rho_c \prod_{t=1}^{T} P_{nt}(\boldsymbol{\beta}_c)$ | $\cdot\boldsymbol{\beta}_c$ is as defined for ECL-MNL.<br>$\cdot\rho_1 = \pi_A\pi_B, \rho_2 = \pi_A(1-\pi_B)$<br>$\rho_3 = (1-\pi_A)\pi_B, \rho_4 = (1-\pi_A)(1-\pi_B)$<br>$\cdot$Estimate $\beta_A$, $\beta_B$, $\pi_A$ and $\pi_B$. |

The GMNL-II likelihood is simulated by taking 600 draws of each random parameter.
$g(.|.)$ and $f(.|.)$ denote log-normal and multivariate normal densities respectively.
Other notations are as defined in Section 2.

TABLE 2

*DGP with scale heterogeneity*

| | *MNL* | *EAA-MNL* | *ECL-MNL* | | *DGP* | *GMNL-II* |
|---|---|---|---|---|---|---|
| $\beta_A$ | 0.939*** | 1.928*** | 2.163*** | $\overline{\beta}_A$ | 3 | 2.696*** |
| | (0.063) | (0.139) | (0.154) | | | (0.456) |
| $\beta_B$ | 0.294*** | 0.664*** | 0.765*** | $\overline{\beta}_B$ | 1 | 0.800*** |
| | (0.036) | (0.107) | (0.105) | | | (0.144) |
| Relative incidence ($\rho_c$) of: | | | | Heterogeneity | | |
| ($\beta_A$, $\beta_B$) | | 0.380*** | 0.555*** | $\sigma_{\beta_A}$ | 0 | 0.529 |
| | | (0.066) | (0.070) | | | (0.397) |
| ($\beta_A$, 0) | | 0.305*** | 0.104 | $\sigma_{\beta_B}$ | 0 | 0.210 |
| | | (0.062) | (0.067) | | | (0.144) |
| (0, $\beta_B$) | | 0.175*** | 0.0272 | $\tau$ | 1.5 | 1.528*** |
| | | (0.033) | (0.025) | | | (0.143) |
| (0, 0) | | 0.140*** | 0.313*** | | | |
| | | (0.034) | (0.039) | | | |
| Parameters | 2 | 4 | 5 | | | 5 |
| Log-likelihood | -1641.277 | -1575.980 | -1558.195 | | | -1528.953 |
| BIC | 3293.963 | 3174.774 | 3144.910 | | | 3086.425 |
| LR statistic | | 130.596 | 166.164 | | | 224.648 |

The LR statistic assumes MNL under the null. Standard errors are in parentheses.
*, **, *** denote statistical significance at the 10%, 5% and 1% level respectively.

TABLE 3

*DGP with coefficient heterogeneity*

| | *MNL* | *EAA-MNL* | *ECL-MNL* | | *DGP* | *GMNL-II* |
|---|---|---|---|---|---|---|
| $\beta_A$ | 2.028*** | 2.591*** | 2.587*** | $\overline{\beta}_A$ | 3 | 2.984*** |
| | (0.104) | (0.117) | (0.117) | | | (0.252) |
| $\beta_B$ | 0.678*** | 1.520*** | 1.513*** | $\overline{\beta}_B$ | 1 | 1.032*** |
| | (0.061) | (0.121) | (0.121) | | | (0.114) |
| Relative incidence ($\rho_c$) of: | | | | Heterogeneity | | |
| ($\beta_A$, $\beta_B$) | | 0.566*** | 0.563*** | $\sigma_{\beta_A}$ | 1 | 0.869*** |
| | | (0.047) | (0.047) | | | (0.089) |
| ($\beta_A$, 0) | | 0.408*** | 0.411*** | $\sigma_{\beta_B}$ | 1 | 0.975*** |
| | | (0.047) | (0.047) | | | (0.118) |
| (0, $\beta_B$) | | 0.0149** | 0.0184* | $\tau$ | 0 | 0.241 |
| | | (0.007) | (0.011) | | | (0.207) |
| (0, 0) | | 0.0108** | 0.00755 | | | |
| | | (0.005) | (0.008) | | | |
| Parameters | 2 | 4 | 5 | | | 5 |
| Log-likelihood | -1096.507 | -1031.969 | -1031.854 | | | -1021.673 |
| BIC | 2204.421 | 2086.753 | 2092.228 | | | 2071.864 |
| LR statistic | | 129.075 | 129.304 | | | 149.668 |

The LR statistic assumes MNL under the null. Standard errors are in parentheses.
*, **, *** denote statistical significance at the 10%, 5% and 1% level respectively.

TABLE 4

*DGP with scale and coefficient heterogeneity*

| | *MNL* | *EAA-MNL* | *ECL-MNL* | | *DGP* | *GMNL-II* |
|---|---|---|---|---|---|---|
| $\beta_A$ | 0.840*** | 1.768*** | 2.007*** | $\overline{\beta}_A$ | 3 | 2.653*** |
| | (0.056) | (0.144) | (0.155) | | | (0.492) |
| $\beta_B$ | 0.271*** | 1.183*** | 1.426*** | $\overline{\beta}_B$ | 1 | 0.914*** |
| | (0.037) | (0.179) | (0.169) | | | (0.199) |
| Relative incidence ($\rho_c$) of: | | | | Heterogeneity | | |
| ($\beta_A$, $\beta_B$) | | 0.231*** | 0.315*** | $\sigma_{\beta_A}$ | 1 | 0.230 |
| | | (0.042) | (0.042) | | | (0.408) |
| ($\beta_A$, 0) | | 0.455*** | 0.350*** | $\sigma_{\beta_B}$ | 1 | 1.056*** |
| | | (0.043) | (0.044) | | | (0.223) |
| (0, $\beta_B$) | | 0.106*** | 0.00445 | $\tau$ | 1.5 | 1.584*** |
| | | (0.020) | (0.012) | | | (0.159) |
| (0, 0) | | 0.208*** | 0.331*** | | | |
| | | (0.033) | (0.036) | | | |
| Parameters | 2 | 4 | 5 | | | 5 |
| Log-likelihood | -1701.444 | -1631.755 | -1616.578 | | | -1593.076 |
| BIC | 3414.296 | 3286.326 | 3261.675 | | | 3214.671 |
| LR statistic | | 139.377 | 169.732 | | | 216.736 |

The LR statistic assumes MNL under the null. Standard errors are in parentheses.
*, **, *** denote statistical significance at the 10%, 5% and 1% level respectively.