AN INTEGRATED MODELLING APPROACH EXAMINING THE INFLUENCE OF GOALS, HABIT AND LEARNING ON CHOICE USING VISUAL ATTENTION DATA

ABSTRACT

Previous economics literature has explored the role of visual attention on choice in isolation without accounting for other influences such as habits and goals or learning effects, nor their interrelationship. In this paper, we: (i) develop a novel joint framework to explore the relationship between visual attention, observed heterogeneity from stated habits and goals, and choice outcomes while accounting for shorter- and longer-term learning effects; and (ii) investigate whether accounting for these relationships improves model predictive power and behavioral insights. The empirical analysis used an eye-tracked discrete choice experiment on sugar-sweetened beverage purchasing (n=152 adults with 20 choice tasks). Results suggest that habits, goals, and shorter-term learning are key drivers of information acquisition whereas cumulative choices (longer-term learning) affect subsequent choice outcome. Importantly, ignoring the joint relationship between habits, visual attention and choice may exaggerate the role of visual attention, leading to incorrect behavioral insights and reduced prediction accuracy.

KEYWORDS: Eye-tracking; Habit; Sugar-sweetened beverage; Choice; Preference; Joint-modelling

JEL CODES: D830 Search; Learning; Information and Knowledge; Communication; Belief; Unawareness; C35 Discrete Regression and Qualitative Choice Models; Discrete Regressors; Proportions; C33 Panel Data Models; Spatio-temporal Models; L66 Food, Beverages, Cosmetics, Tobacco, Wine and Spirits; I120 Health Behavior

ABBREVIATIONS: AOI, area of interest; AR, auto-regressive; DCE, discrete choice experiment; MNP, multinomial probit; SSB, sugar-sweetened beverages;

1.0 INTRODUCTION

A better understanding of human decision-making behavior is fundamental to successful prediction and understanding the drivers of economic choices. Cognitive process tracing methods, such as eye-tracking, are well-established methods of seeking insight into the complex processes occurring within the 'black box' of consumer decision-making (Schulte-Mecklenbeck et al. 2019). In the last decade, eye-tracking data have been used increasingly in the fields of psychology, neuroscience, marketing, health economics and food and agricultural economics to penetrate this 'black box' to explore (i) how 'bottom-up' influences in the visual environment (e.g., Orquin et al. 2019) and experimental constraints (e.g., Fenko et al. 2018; Ryan et al. 2018) affect visual attention and thereby affect choices; (ii) how 'top-down' habits and goals guide visual attention and thereby affect choices (e.g., Reutskaja et al. 2011; Büttner et al. 2014; Balcombe et al. 2015; Meißner et al. 2016; van der Laan et al. 2016); and (iii) the potential improvement in the predictive power of choice models using visual attention data (Orquin & Loose 2013; Towal et al. 2013; Balcombe, Fraser & McSorley 2015; Spinks & Mortimer 2015; van der Laan et al. 2015; Mullett & Stewart 2016; Krucien et al. 2017; Meyerding 2018; Van Loo et al. 2018b; Vass et al. 2018; Yegoryan et al. 2019).

The first of the three sets of above studies focus on how choice experiments should be designed to minimize the influence of lexicographical biases. The second set of studies focus on explaining the underlying decision-making process through the use of visual attention data. The third set of studies focus on improving predictions by utilizing information about attribute attendance through eye-tracking data. In the latter studies, few explicit assumptions are made about the relationships between visual attention, information acquisition and decision processes. However, the implicit assumption is that visual attention has a down-stream effect on choice. For a detailed review of eye-tracking measurement and factors affecting visual attention in choice experiments see Orquin & Loose (2013) and Yegoryan, Guhl & Klapper (2019), and in the food preferences literature, see Van Loo et al. (2018a).

The literature to date has therefore mainly explored the role of eye-tracking data for different decision-making strategies in isolation without accounting for top-down influences such as habit. Failure to control for unobserved heterogeneity across different model components (e.g., habit and goals, visual attention and choice), and the feedback effect due to learning in repeated choices over time, may lead to a spurious effect of visual attention on choice; it may also worsen model predictive power. For example, Camerer et al. (2004) and Gabaix et al. (2006) have reported that accounting for the effect of previous choices on subsequent choices improves model prediction. Our proposed model extends the previous literature which has mainly considered the influences of 'top-down' and 'bottom-up' process pathways, heuristic processing, and the influence of previous choice on subsequent choice in isolation, and allows us to account for the interactions between these processes.

The potential interaction of 'top-down' and 'bottom-up' processing pathways in consumer decision-making has significant implications for business, including in the design of product packaging (Orquin, Bagger, Lahm, Grunert & Scholderer 2019) and store layout and product positioning (Valenzuela et al. 2013; Orquin, Bagger, Lahm, Grunert & Scholderer 2019). For example, there is emerging literature to suggest that weight-consciousness is associated with both increased visual attention to nutritional information on food products and increased willingness to pay for nutritional information (Ran et al. 2015). Further understanding of these interactions may help guide retail practices. For example, in order to promote sales for products that promote healthy weight, but are often not perceived as such by consumers, like nuts, manufacturers could consider displaying nutrition information more prominently to engage weight-conscious customers while keeping prices the same so as not to discourage purchases by customers who are not weight conscious.

In the current study, we address this gap and add to the health economics and business literature by developing a joint model to account for the influence of top-down factors on visual attention pathways. This paper advances the previous literature by accounting for the effects of ‘top-down’ influences on choice, as well as the interrelationships and feedback loops between these ‘top-down’ influences, visual attention and choice. We focus on improving predictions and quantifying the effect of visual attention on choices, after controlling for potential confounds. We assume top-down goal-driven control of visual attention (also referred to as the “endogenous effect”) (Corbetta & Shulman 2002; Theeuwes 2010). The current study is motivated by the apparent paucity of consideration of the endogenous effect in previous prediction studies on the effects of goal, habits, visual attention and choice (i.e., goals and habits may direct visual attention, which subsequently has a down-stream effect on choice) (Van Loo, Grebitus, Nayga Jr, Verbeke & Roosen 2018a). This endogenous effect is our principal focus. We formulate a comprehensive econometric framework and provide a computationally feasible estimation process. Although controlling for unobserved factors in a multilevel model is not difficult, the estimation of such models becomes near impossible using the usual full information likelihood or Bayesian approaches (see Bhat & Dubey 2014 for a detailed discussion on issues related to estimation of multilevel models) for a detailed discussion on issues related to estimation of multilevel models). Our proposed estimation method circumvents these difficulties.

In this study, we use eye-tracking data from a discrete choice experiment (DCE) on the effect of changing volume and price on non-alcoholic beverage purchases (n=152) to investigate the effect of factors influencing inherent preferences (including habits and health goals) on choice and examine the mediating effect of visual attention using an integrated modelling approach. This is the largest eye-tracking sample size we are aware of to date in the health economics and food marketing literature. We address the above-highlighted research gaps and develop a comprehensive framework for analysing multilevel choice data with supplementary eye-tracking information to answer the following questions: Does accounting for the endogenous relationship between goals and habits, visual attention, and choice improve the predictive power of and insights drawn from choice models? To what extent is modifying visual stimuli of beverage alternatives predicted to change preferences and behaviour?

More generally, we add to the advancement of multilevel choice data analysis by developing a comprehensive framework that connects various components (visual attention, habits and goals, and choice) of models using a fully-specified covariance structure. We incorporate feedback loops in both visual attention and choice components in a parsimonious fashion through the use of a first-order lag structure.

The rest of the paper is organized as follows: in the next section we summarise the existing literature on visual attention, and existing models of the effects of habit and learning on choice, as well as highlight relevant literature gaps; we then outline our empirical example, followed by a detailed description of the methodology. We report and compare the model results and out-of-sample prediction statistics followed by discussion and concluding remarks.

2.0 LITERATURE REVIEW

2.1 The Relationship Between Visual Attention And Choice

There are two main ways in which visual attention is posited to affect choice: the ‘top-down’ goal-driven pathway, and the ‘bottom-up’ stimulus-driven pathway. In the former cognitive process, individuals focus their attention on relevant cues based on goals and pre-defined preferences (Land et al. 1999; Hayhoe 2000; Hayhoe et al. 2003). Previous research by van der Laan et al. (2016) and Orquin and Scholderer (2011) found increased attention on food options that correspond to respondents’ intended health goals. It is plausible that pre-defined goals and habits direct visual attention towards relevant products, for inclusion or exclusion from the choice consideration set (Souza 2015), according to ‘Choice Set Formation’ theory (Swait 1984; Ben-Akiva & Boccara 1995).

The second, or ‘bottom-up’ process considers choice as stimulus-driven. The bottom-up process assumes that by making an alternative more salient, one can affect the choice. The stimulus-driven process presents an opportunity to change health behaviours through modifying the stimulus. Recent research has demonstrated the importance of salience of product packaging elements (Chandon et al. 2009; Orquin, Bagger, Lahm, Grunert & Scholderer 2019) on consumer attention, independent of consumer health goals (Orquin, Bagger, Lahm, Grunert & Scholderer 2019). Eye-tracking provides a useful opportunity to examine the influence of goals and habits on choice, mediated through visual attention.

Further, decision-making heuristics may drive visual attention and thereby choice. For example, sequential visual attention movement across the ‘row’ in a traditional tabular choice task layout may suggest an ‘elimination by aspects’ strategy whereby a given attribute is compared to a threshold or compared across alternatives (Tversky 1972). Alternatively, visual attention that moves sequentially down a column may suggest an ‘additive compensatory-model’ approach in which all attributes for a given alternative are considered before moving on to the next alternative (Keeney & Raiffa 1993). For example, Ares et al. (2014) found in an eye-tracked choice experiment that consumers who reported more ‘analytical’ thinking styles had longer attention on nutritional information in order to differentiate between yoghurt alternatives. Those who reported more ‘intuitive’ thinking styles spent relatively more time looking at the label background. It is possible that there is a causal relationship between consumer health goals and the use of heuristics, but this needs to be established by future research. Nonetheless, the conjecture that goals may cause heuristics adoption is a plausible one.

Improved understanding of the cognitive processes that lead to choice decisions may enhance the real-world applicability of data from experimental studies, and potentially identify levers for intervention when the goal is to change choices through altering preferences. This paper examines the effects of neglecting the endogenous relationship between goals and habits, visual attention and choice may introduce bias in the parameter estimates and exaggerate the effect of habits, goals and visual attention on choices.

2.2 The Effect Of Learning On Visual Attention And Choice

Over time, individuals learn to separate relevant and irrelevant cues through practise and experience (Haider & Frensch 1999). Previous studies have established that decision-makers become more efficient over time when making repeated or similar choices, potentially due to learning (Payne et al. 1988; Meißner & Decker 2010; Meißner, Musalem & Huber 2016). The availability of eye-tracking along with choice data opens an avenue to disentangle the effect of shorter-term choices (choices made in the last one or two choice occasions) and longer-term choices (cumulative count of various choices made until the last choice occasion in a stated preference (SP) study).

In this study, we refer to the effect of past choices on subsequent choices as “learning”. However, we acknowledge that there are several potential explanations for this effect. The ‘drift diffusion model’ in psychology (Krajbich & Rangel 2011) describes the accumulation of information over time in favour of a particular alternative, until evidence in favour of that alternative exceeds a threshold. Similarly, the ‘choice perseveration model’ (Senftleben et al. 2019) posits that previous choices of an alternative cumulatively bias a respondent towards that alternative.

One way to capture the learning effect is to regress the exogenous variables of the previous time periods (e.g., experimental constraints such as previous price levels) on the current choice (see Erdem et al. 1999 for further discussion). Although this approach is easy to incorporate, it may cause explosion of parameters for a moderate to high number of alternatives and attributes. An alternative could be regression of the past utility value on the current utility in order to reduce the number of parameters. However, a simple utility regression approach may induce bias in parameter estimation due to the need to regress both observed and unobserved utility portions (Bhat 2015).

In order to obtain unbiased estimates, a 'lag structure' on utility (both observed and unobserved) is used widely in spatial econometrics and time series analysis (LeSage & Pace 2009). The use of a lag structure is elegant but challenging due to estimation of high dimensional integrals (see Anselin 2001 for a detailed discussion of pertinent issues). Instead, eye-tracking researchers outside of health (this issue has been ignored to date in the health literature) have used simple regression by either incorporating previous choices (e.g., Meißner, Musalem & Huber 2016) or previous attribute values as explanatory variables (e.g., Ben-Elia & Shiftan 2010). These approaches may cause bias in parameter estimates if an auto-regressive component is present in the data generation process.

On the other hand, incorporating learning effects requires capturing the effect of past choices and contexts on present choices. Abstracting the potential availability of data, the econometric challenge in representing learning models lies in accounting for unobserved factors across choice occasions, which imply that choices (utilities) are not independent over time. In this paper, we incorporate a first-order autocorrelation process in our econometric framework to quantify the impact of full (systematic and stochastic) prior preferences. To our knowledge, this is the first such specification in the eye-tracking literature. We develop a parsimonious model with improved predictive power compared to extant practice. Below, we apply our model to decision-making in a beverage choice task with health policy and retail practice implications.

3.0 MATERIAL AND METHODS

3.1 Empirical application

There is increasing consumer and government interest in reducing the consumption of sugar-sweetened beverages (SSBs), which are a major cause of excess energy consumption and contribute significantly to the global burden of chronic disease, including obesity (Singh et al. 2015). Understanding the mechanisms for consumer beverage choices may help guide retail changes or policy development to decrease the purchase and consumption of less healthy beverages and to increase the consumption of healthier beverages.

The relationship between visual attendance and participant demographics, beverage preference and choice characteristics was explored using an eye-tracked DCE. Details of the DCE without the addition of eye-tracking data have been published (Blake et al. 2018; 2019)which report on the DCE applied to different, larger samples than used in the eye tracking dataset used in this current study. Briefly, the primary purpose of the DCE was to explore heterogeneity in consumer beverage preferences and price responsiveness over key socioeconomics characteristics including income levels and usual SSB consumption frequency. This eye-tracked dataset provides the opportunity to investigate the effect of factors influencing inherent preferences (self-reported habits, goals and experimental constraints) on choice, and to then examine the mediating effect of visual attention and to do so accounting for learning effects.

3.1.1 Participants

Participants completed the DCE while being monitored at an eye-tracking laboratory in Melbourne, Australia. Participants were Australian residents 18 years or older. Recruitment targets were set for this sample so as to reflect the Australian adult population in age and gender. A minimum of 70% of participants who had consumed a SSB purchased from a convenience store at least “a few times” in the past month was set. Participants were recruited from a database of past participants at the research center, through the university staff newsletter, social media, local newspaper advertising, and direct recruitment through local community organisations. Participants provided written informed consent and were given an AU$30 supermarket gift card for their time. Ethical approval was received from Monash University Human Research Ethics Committee (approval number CF15/4153 - 2015001760).

3.1.2 Experimental Design

In the labelled DCE, participants selected a beverage within a hypothetical convenience store setting. Each participant completed 20 choice tasks involving three SSB alternatives (energy drink, flavored milk, regular soft drink (i.e., “soda”)), four non-sugar-sweetened alternatives (non-SSBs: plain low-fat milk, fruit juice, diet soft drink, bottled water), and a “no drink” alternative (meaning that they would “consume no drink on this occasion”). Each beverage was described by alternative-specific prices and generic volume attributes which each varied over four levels. An orthogonal design was generated using Ngene software (Rose et al. 2009). Prior to completing the choice tasks, half of participants were randomly exposed to a real-world educational message designed to discourage selection of SSBs. See Web Appendix A for further detail on experimental design and an example choice task and list of attribute levels for each alternative.

Following the DCE, participants completed questions on stated attendance to attributes and alternatives as well as strength of SSB consumption habit. This included an 11-point scale of readiness to consider reducing SSB intake based on a validated tool to assess readiness to quit smoking (Biener & Abrams 1991) and the Self-Report Behavioral Automaticity Index, a

4-item measure of habit strength measured on a 5-point Likert scale with higher scores signifying a stronger habit (Gardner et al. 2012).

3.1.3 Eye-Tracking Data

A discrete, web-cam like device tracked eye movements (Tobii Pro, 2011, Tobii TX300; Stockholm, Sweden). Participant visual attention to Areas of Interest (AOIs) was defined using a continuous measure (fixation duration) (Krucien, Ryan & Hermens 2017). AOIs were defined for each attribute 'row', each alternative 'column' and for each individual choice task table cell. For each participant, choice tasks with less than two fixations were excluded from the analysis to reduce data noise from random eye-movements.

3.2 Model Overview

We describe the model here with further detail including relevant estimation approach provided in Web Appendix B.

3.2.1 Econometric Details

Let j=1,…,8 be labelled alternatives, where j=8 represents the "no drink" option. Each respondent completes T tasks, each task t having a choice set C_t= {1,2,…,8} of all beverages. A beverage is presented as a constant label (e.g., fruit juice, flavored milk, see Web Appendix A- Fig. A.1), a generic size for all beverage types S_j (varying across four levels) in milliliters, and a varying alternative specific price (p_{jt}) in Australian dollars (see Web Appendix A- Table A.1 for price levels). With this preamble, the model specified in Fig. 1 can be defined econometrically.

--*Insert Fig. 1 about here*---

Let the utility U_{jt} (subscript for person n is omitted for clarity, but should be assumed throughout) be given as

(1) $U_{jt}=\alpha_j + \beta_j(S_j/p_{jt}) + \gamma_j d_{j,t-1} + \delta_j D_{j,t} + \varphi_j \mathit{ln}(Y_{jt}) + \varepsilon_{jt}$, j=1,…,8, t=1,…,T,

where

α_j is the alternative-specific constant for beverage j;

β_j is the marginal impact of the volume to price ratio for beverage j, expected to be positive;

$d_{j,t-1}$ =1 if beverage j chosen in the prior task (t-1), =0 otherwise, used to proxy for shorter-term learning within the task;

γ_j is the utility impact of $d_{j,t-1}$;

$D_{j,t}$ $=\sum_{t=1,\ldots,t-1} d_{j,t}$ is the cumulative choice of beverage j in all prior tasks to t, which proxies for longer-term learning within the task;

δ_j is the utility impact of $D_{j,t}$;

Y_{jt} is the visual attention the respondent gave to beverage j during task t, which is defined as the total time (msec) spent on the label, volume and price, used in the model with a natural log transform to reflect the assumption of diminishing marginal impact of visual attention on utility (see Orquin & Loose 2013);

φ_j is the utility impact of $\ln(Y_{jt})$;

ε_{jt} is the additive stochastic utility for j at task t.

As we noted earlier, we assume that ε_{jt} is auto-regressive AR(1). An AR(1) process allows for the possibility that time previously spent on an alternative partly determines how much time will be spent on it currently, combining the possibility that both present and past conditions help to establish present behaviour.:

(2) $\varepsilon_{jt}=\lambda_j\varepsilon_{j,t-1} + \eta_{jt}$, j=1,…,8, t=1,…,T,

λ_j is the one-period autoregression coefficient, with a range from -1 to +1;

η_{jt} is a contemporaneous stochastic utility that has no time dependence to it.

This assumption allows stochastic sources of utility for a beverage to be correlated over trials. The link between utilities U_{jt}, for all j, and observed choice d_{jt} is given through the relationship

(3) $d_{jt}=1$ if $U_{jt} \geq \max(U_{kt}, k \neq j)$, =0 otherwise, for j=1,…,8, t=1,…,T,

implying that choice is made on the basis of utility maximization. Since the utilities are stochastic, it is necessary that we specify the distributional law followed by errors η_{jt} to specify the link between utilities and observed choices. We assume that

(4) $\eta_t \sim MVN(0_\eta | \Omega_\eta), \mathrm{t} = 1, \dots, \mathrm{T},$

where MVN(a|B) is the multivariate normal distribution with mean a and covariance matrix B;

η_t is a 8x1 vector of stochastic utilities;

0_η is a 8x1 vector of zeroes;

Ω_η is the contemporaneous covariance matrix for the stochastic utilities (note that there is no temporal component to this matrix).

We estimate the visual attention (continuous) model which is later integrated into the choice model. The visual attention model is given by the following equation:

(5) $Y_{jt}=a_j + \rho_j Y_{j,t-1} + \sum_{l=1..3}\kappa_{jl} H_l + \sum_{k=1..6}\pi_{jk}\Psi_k + \theta_j d_{j,t-1} + \xi_{jt}$, j=1,…,8, t=1,…,T,

where

a_j is the intercept of visual attention time for beverage j;

ρ_j is the AR(1) coefficient for the previous time spent on beverage j, ranging in the interval [-1, +1];

H_l is the individual's habit, a count of *l*={strongly disagree, disagree, neutral, agree and strongly agree} across four scale items (see definition in note for Table 1);

κ_{jl} is the marginal time impact of scale value H_l on visual attention given to j;

Ψ_k is equal to 1 if the individual's score or response on an item measuring the intention to drink less SSBs on a 10-point scale (1=no thought of drinking less to 10=taking action to drink less) is equal to k, k=1,…,6, and Ψ_k =0 if k=7,…,10;

π_{jk} is the marginal time impact of the k-th dummy variable Ψ_k on beverage j;

θ_j is the time impact of $d_{j,t-1}$;

ξ_{jt} is a stochastic source of visual attention time arising from other sources than those enumerated in (5).

To complete the specification of model (5), we need to stipulate the density for

(6) $\xi_t \sim MVN(0_\xi | \Omega_\xi)$, t=1,…,T,

which has an analogous interpretation to the terms defined for expression (4). Finally, we specify that error terms (η_t,ξ_t) may covary across beverages in the same task. Since both stochastic vectors are MVN, we can specify this as follows:

(7) $\begin{pmatrix}\eta_t \\ \xi_t\end{pmatrix} \sim MVN\begin{pmatrix}0_\eta & | & \Omega_\eta & \\ 0_\xi & | & \Omega_{\eta\xi} & \Omega_\xi\end{pmatrix}$, t=1,…,T,

where $\Omega_{\eta\xi}$ is the covariance matrix for stochastic covariation between (η_t,ξ_t); other quantities as previously defined.

To summarize, the model system depicted in Fig.1 has the following characteristics which together significantly advance the current approach to visual attention data and choice model analysis:

a) The choice component is a Multinomial Probit (MNP) model with contemporaneous covariation given through the covariance matrix Ω_η, which is 8*x*8, thus allowing beverage utilities to be correlated positively or negatively for the same task, and for utility variances to differ across beverages. In addition, the MNP model allows for an AR(1) error at the beverage level.

b) The visual attention time, Y_{jt}, is a nonlinear predictor (through the natural logarithm transformation) of the attractiveness/utility of a beverage. The natural logarithm reflects the *a priori* conjecture that the marginal impact of visual attention on utility of beverage j diminishes with increasing time.

c) Y_{jt} is influenced by past visual attention to beverage j through an AR(1) specification, in addition to which habit, health goal and learning can impact the attention given to a beverage during any task.

d) Visual attention is correlated across beverages, through the covariance matrix Ω_{ξ}, which is 8*x*8, making it possible that consistent patterns of time allocations to beverage pairs (whether increasing or decreasing) be captured within a task.

e) Finally, contemporaneous stochastic utilities η_{jt} and stochastic visual attentions ξ_{jt} for a given beverage j during task t can covary, through covariance matrix $\Omega_{\eta\xi}$, also 8*x*8.

We tested the following models where attention/AOI time is modelled as a driver of preference, where V represents the observed part of utility and E is the unobserved part of utility. A 'Joint' model refers to models where the habit, visual attention and choice outcomes are linked by the covariance structure and has the properties a) to e) as described above. An 'Independent' model refers to a model which does not assume a correlation between visual attention time and choice through an error structure:

- Joint-AR(1)VE: Joint model with AR(1) structure on both observed and unobserved parts of utility
- Joint-AR(1)V: Joint model with AR(1) structure on observed part of utility
- Joint-AR(1)E: Joint model with AR(1) structure on unobserved part of utility

- Independent-AR(1)E: Independent model with AR(1) structure on unobserved part of utility

We also tested the following models where time is used to capture screening behavior through a penalty function (P), to be detailed later:

- Joint-AR(1)VEP: Joint model with AR(1) structure on both observed and unobserved parts of utility and penalty function
- Joint-AR(1)EP: Joint model with AR(1) structure on unobserved part of utility and penalty function

Please note that for all the models, the continuous (visual attention) component has AR(1) structure on both observed and unobserved portions of propensity.

Identification of this model system requires that a number of restrictions be imposed. With respect to the choice model, it is necessary that one of the Alternative Specific Constants (ASCs) be normalized, so we set $\alpha_1=0$ (for j=1, bottled water). Additionally, it is necessary to restrict elements of covariance matrix Ω_η since at most 7*8/2=28 of its 8*9/2=36 elements can be identified (Bunch 1991), with at least one of the 28 elements being normalized to unity (in this case, the variances of the differences of stochastic utility of energy drink and bottled water, j=1,2); accordingly, the cell (1,1) is set to 1.

The joint model described above is used to test whether visual attention is a driver of choice. To test whether habits, goals, and constraints work as *screening mechanisms*, we still use visual attention Y as an explanatory variable in expression (1), but with a different functional form that lets it serve as a penalty to utility. Specifically, we rewrite the utility function of beverage j as follows. Note that the penalty function of each alternative differs:

(8) $U_{jt}=\alpha_j + \beta_j(S_j/p_{jt}) + \gamma_j d_{j,t-1} + \delta_j D_{j,t} + \mathit{ln}\tau_{jt} + \varepsilon_{jt}$, j=1,…,8, t=1,…,T,

where

τ_{jt} $(1 + \exp(Y_{jt}))^{-1}$ is the penalty term associated with beverage j in task t.

The logistic parameterization of the penalty τ ensures that its value is bounded between 0 and 1, so in expression (8) the penalty is bounded between $-\infty$ (Y_{jt} small, near zero) and 0 (Y_{jt} large). Thus, an alternative is *screened* out (i.e., becomes unavailable) because its utility grows very negative as visual attention decreases. Note that there is no further stochastic component in the penalty function other than that implied through the logistic functional form.

While we assume the direction of causality to be from goals and habits to visual attention, which subsequently informs preferences through choice, it is plausible that other causal relationships may co-exist. For a model with three dependent variables, a total of six different causality directions may co-exist. For example, goals and habits may affect choices which can then direct visual attention. In this paper, we do not model all possible causality directions. Researchers can simultaneously model multiple causality directions by embedding the proposed multilevel framework in a latent class framework where each class represent a causality direction.

3.2.2 Parameter Estimation by Composite Maximum Likelihood

The full vector of parameters to be estimated is quite extensive due to the dimensionality of the three covariance matrices, even after accounting for identification restrictions that must be imposed:

$$\Gamma_C=\{(\alpha_1,\ldots,\alpha_8)', (\beta_1,\ldots, \beta_8)', (\gamma_1,\ldots,\gamma_8)', (\delta_1,\ldots,\delta_8)', (\varphi_1,\ldots,\varphi_8)', (\lambda_1,\ldots,\lambda_8)'\}$$

$$\Gamma_Y=\{(a_1,\ldots,a_8)', (\rho_1,\ldots,\rho_8)', (\kappa_{11},\ldots,\kappa_{83})', (\pi_{11},\ldots, \pi_{86})', (\theta_1,\ldots, \theta_8)'\} \quad (9)$$

$$\Gamma_\Omega=\{\Omega_\eta, \Omega_\xi, \Omega_{\eta\xi}\},$$

This dimensionality imposes a significant computational burden in using traditional likelihood-based estimation methods, reflecting the complication of a first-order auto-regressive MNP choice model, plus the lagged, linear visual attention models. This causes difficulties both theoretical and computational in nature (e.g., choice probabilities near zero). By itself, the MNP choice probability is a well-known challenge in the literature (Connors et al. 2014). Simulated maximum likelihood methods (e.g., Geweke- Hajivassiliou-Keane (GHK) simulator, Hajivassiliou et al. 1996) can calculate MNP probabilities accurately only up to a limited number of dimensions (Sándor & András 2004) and suffer from long computational times (Train 2000; Craig 2008). Therefore, it is challenging to estimate the full set of parameters using maximum likelihood.

A competing method is to use a Bayesian approach to evaluate the complex likelihood function, which would involve sampling from a complex series of conditional distributions. A review of literature involving the MNP kernel shows that the Bayesian approach has often not performed as expected in terms of recovering parameters and their standard errors (Franzese et al. 2010; Patil et al. 2017), though some studies have found the performance of Bayesian approach to be quite good (Daziano 2015). Faced with these polarized results, we opted not to pursue this path.

Instead, we use the composite marginal likelihood (CML) approach. This has been established in the last decade as a powerful approach for parameter estimation involving likelihood functions with high dimensional integrals. A comprehensive discussion on the CML approach is outside the scope of this paper and readers are referred to the literature for background (Varin & Vidoni 2005; Varin 2008; Varin et al. 2011), and to Bhat and colleagues (Bhat & Dubey 2014; Bhat et al. 2016) for its application in the context of discrete choice models. Bhat and colleagues have performed extensive simulation testing using the CML approach for complex econometric models and have observed highly accurate results.

One of the practical advantages of the CML method for our problem is that it reduces the dimensionality of integration of likelihood function terms to calculations based on pairs of random variables. To our knowledge, this is the first time CML has been applied in the eye-tracking literature. The details of the CML likelihood function and our estimation method are provided in Web Appendix B.

4.0 RESULTS

4.1 Sample Description

Between November 2015 and March 2016, 160 eligible adults completed the eye-tracked DCE (see Web Appendix C Fig. C.1 for participant flow diagram). Eye movements were recorded on every choice task for 139 participants (used for main analysis) and during at least one choice task for 13 participants. These 13 individuals were excluded from the main analysis but used to test out-of-sample prediction. Mean duration of the study (DCE and post-DCE questions) was 24.6 mins (SD 7.8), and the DCE alone 4.4 mins (SD 2.2).

Participant demographics are summarized in Web Appendix D, Table D.1. The convenience sample by design approximately reflected the Australian population based on age and gender. There was a higher proportion of those in the lowest income quintile compared to the Australian population income distribution. Sixteen percent reported that they never drink SSBs. Participants scored a mean 9.6/20 (SD 4.3) on the Self-Report Behavioral Automaticity Index habit measure, meaning that on average participants had a moderately strong SSB consumption habit (Gardner, Abraham, Lally & de Bruijn 2012). Forty-eight percent of participants reported currently taking action or considering how to drink fewer SSBs.

4.2 Description of Visual Attendance

While total fixation duration per choice task nearly halved from mean 18.2 secs (SD 10.0) in the first choice task to 10.3 secs (SD 8.0) in the final choice task, the proportion of time spent looking at *relevant information* (choice set task), increased from mean 71% fixation duration (SD 16%) to 82% fixation duration (SD 17%) in the final choice task.

Visual non-attendance of beverage types was highest for energy drink, and lowest for bottled water. Non-attendance on all beverage types increased through subsequent choice tasks, although non-attendance was temporarily decreased after the 10th choice scenario when participants were presented with a message reminding them to "consider their options carefully". Most people attended to volume and price in every choice task. Further descriptive results of visual attendance data are found in Web Appendix E.

4.3 Model Estimation Results

In this section, we first present fixation duration results (Table 1), followed by choice component results (Table 2). As noted in Web Appendix A, we tested the effects of the educational message using the model of best fit (Joint-AR(1)E, fully compensatory AR-1 Error model, described later) and found no significant effect on beverage choice, hence sub-samples were pooled and we used the full sample (n=139) in the estimation (n=13 used in out-of-sample predictions below).

-------Insert Table 1 about here-----

-------Insert Table 2 about here-----

We found evidence for the AR(1) structure on both observed and unobserved components of the fixation duration (continuous) model, combined with AR(1) structure on the unobserved portion of the choice component, as in the Joint-AR(1)E model. This implies that respondents do exercise their experience from previous tasks when acquiring information on alternatives and thus past fixation behavior guides current information acquisition strategy. Therefore, the results described below correspond to the Joint-AR(1)E model.

As shown in Table 1 and as anticipated, stronger SSB habits (H_l) are generally associated with positive (increased) visual attendance time on SSBs and negative (decreased) visual attendance time on non-SSB alternatives. For example, people with a moderate to strong habit of drinking SSBs are likely to spend less time looking at the attributes of bottled water as compared to attributes of regular soft drink. Some parameter estimates for visual attention are the same for different health goal categories (Ψ_k). For example, mild to moderate health goals with scores in the range of 1 to 6 out of 11 had the same association with visual attention to bottled water. Participants who reported a high intention to drink less SSBs spent more time looking at the attributes of SSBs compared to people who have a lower intention to change SSB consumption. Intuitively, it may suggest that a conscious decision to reduce consumption of SSBs leads to careful evaluation of various aspects of such beverages prior to choice. This could be a demonstration of ‘regret regulation’ (Pieters & Zeelenberg 2007), which posits that choices are made to minimize future regret, leading to a careful examination of products which they are trying to avoid.

We also observed that shorter-term learning results ($d_{j,t-1}$) suggested that respondents tended to spend more time on an alternative if it was chosen in the previous task occasion. Finally, the positive autoregressive coefficients (ρ_j) for all beverages (last row of Table 1) suggest that respondents do exercise their experience (reinforcing or discouraging from previous tasks) when acquiring information on alternatives, and thus past fixation behavior guides current information acquisition strategy. The AR structure parsimoniously captures the effect of past information (represented through habit, goal, past choices and other unobserved characteristics) on current information acquisition (visual attention time spent on attributes), and therefore operates as a feedback link between past and current tasks.

In Table 2 (MNP choice model results), the volume and price attributes are included in the model as a volume/price ratio to accommodate the trade-off between them. As per *a priori* expectations, the volume/price ratio (β_j) was significant and positive for all beverages, suggesting participants preferred beverages with higher volume per dollar ratios. We observed non-significant coefficients for the direct effect of shorter-term choices ($d_{j,t-1}$) indicated by last chosen beverage on the subsequent beverage selection, suggesting that shorter-term choices are an indirect driver of information acquisition through visual attention time to attribute and alternative information. However, we found a significant and positive effect of longer-term preference on the choice of all beverages including the “no drink” option (indicated by the cumulative sum of chosen alternatives until the last choice occasion, $D_{j,t,}$).

In addition to these findings from the Joint-AR(1)E model, both shorter and longer-term learning effects were found to be significant in both visual attention and choice components in the independent model (Independent-AR(1)E, the model which does not assume a correlation between visual attention time and choice through error structure).[1] The AR coefficient is positive and statistically significant, suggesting the presence of feedback loops between past and current choice occasions. Finally, time spent on beverage information has a positive effect on the likelihood of choice of a beverage. Thus, importantly, with the help of the joint model, we are able to disentangle the effect of shorter- and longer-term preferences on information acquisition and alternative selection (choice). Results for habit and goal parameters (direct effect of habit and goals on utility) were not significant.

Although this broad directional effect finding is in line with Balcombe, Fraser & McSorley (2015), in Table 3 we estimate the joint covariance matrix (Ω), along with inclusion of the autoregressive structure, which allows us to obtain the ‘true effect’ of structural endogenous factors such as fixation duration, short and longer-term choices, while allowing for better model fit. Estimates greater than zero indicate positive correlation between visual attention and choice, while estimates less than zero indicate negative correlation. For example,

[1] The detailed estimation result for the independent model is available from the authors on request.

utility of healthier alternatives like bottled water, plain low flat milk, diet soft drink and fruit juice are positively correlated with visual attention to bottled water. Our assumption that there exists a significant correlation between information gathering as observed through fixation duration (continuous model), habit and final decision-making (choice model) is reinforced by the covariance matrix. In addition, characterization of unobserved sources of dependence in information gathering across SSBs implies that we control for the bias in the model that would otherwise be created in the observed sources of dependence, and is generally ignored in the prior literature.

---*Insert Table 3 about here*---

---*Insert Table 4 about here*--

4.4 Data-fit Statistics

Table 4 displays the model fit statistics. We explored two decision-making mechanisms using eye-tracking data: (1) fully compensatory, and (2) two-step decision-making process where screening precedes the fully compensatory decision step. The fully compensatory behavior is captured by the model where fixation duration is used as an explanatory variable in the choice model. The second decision-making behavior is captured by introducing a penalty function in the choice model as a function of fixation time (as discussed in the Methodology section 3.1.3). The estimation results for the penalty models are similar to the fully compensatory models, including direction of signs of parameter coefficients, together with positive fixation duration parameters. The penalty value for a beverage alternative approaches a large negative number as the fixation duration increases. This suggests that participants may spend more time analyzing an option before eliminating it from the final consideration set in order to minimize choice regret (Pieters & Zeelenberg 2007).

Table 4 also provides the model fit measures for these two competing models and other tested models. Since the models were estimated using a CML approach, the non-nested models can be compared by the Composite Likelihood Information Criterion (CLIC), which is similar to the familiar AIC and BIC criteria (Varin & Vidoni 2005). The model with *higher* CLIC is preferred. Based on CLIC statistics, the current dataset is best represented by the fully compensatory model (Joint-AR(1)E) with CLIC of -4942922.25, compared to a CLIC of -4958377.42 for the screening model Joint-AR(1)EP, and CLIC of -4961982.10 for screening model Joint-AR(1)VEP. This suggests that the fully compensatory decision behavior is preferred in the current dataset, an eminently reasonable result given the low complexity of the choice task (eight alternatives with two varying attributes).

We then tested the performance of Joint-AR(1)E against the nested models using adjusted composite likelihood ratio test (ADCLRT) (equivalent to the likelihood ratio test in the CML approach; see Varin, Reid & Firth 2011). The Joint-AR(1)E model is superior to its competitors with the same compensatory behavior mechanism but with AR structure on observed utility (Joint-AR(1)VE and Joint-AR(1)V), and to the Independent-AR(1)E, in which the correlation in the unobserved part of utility between fixation duration and choice is zero (*p*-value .010).

Differences in model fit may be exaggerated due to the difference in log-likelihood values while in fact performing equally well in terms of in-sample or out-of-sample prediction. Table 5 demonstrates that the fully compensatory behavior model Joint-AR(1)E has better prediction accuracy for both in-sample (mean absolute error (MAE) of .031) and out-of-sample (MAE of .013) data compared to all other models. Interestingly, while there is a large discrepancy in data fit statistics, predictions are very similar for the fully compensatory behavior model Joint-AR(1)VE (.039 and .029 for in- and out-of-sample predictions, respectively) and Joint-AR(1)V (.038 and .031 for in- and out-of-sample predictions, respectively). Among all tested models, the Independent-AR(1)E model has the worst in- and out-of-sample prediction accuracy. These results support the need to capture screening processes to enhance the predictive power of eye-tracking models.

--*Insert Table 5 about here*---

--*Insert Table 6 about here*---

4.5 Elasticity Effects

To quantify the true magnitude of difference in discrete choice model estimations accounting for the possibility of screening during the decision-making process with those models that do not, we calculate the elasticity effects for fixation time with respect to beverage choice. For brevity, we only calculate and compare the elasticity effect of fixation for the fully compensatory model Joint-AR(1)E (preferred model) and its corresponding independent version (Independent-AR(1)E).

For the elasticity calculation, we increase the fixation time by 10% and calculate the implied change in share for each beverage. Since the model is based on a Probit kernel, the expression for elasticity effects does not take a closed form. Table 6 shows that elasticity values obtained from the two models are indeed statistically different (for all beverages, the *p*-value<.05). As expected, the implied shares are higher for the independent model than the joint model. Finally, the true effect of visual attention on choice (share from the joint model divided by share from the independent model) is around 56% to 65% for all beverages. This implies that if an analyst fails to consider the interrelationship between information gathering (visual attention) and information processing (decision-making), the result may be an overestimation of the impact of visual attention on actual choice.

5.0 DISCUSSION

In this study, we developed a model to analyze the relationship between habits and goals, visual attention and choice outcomes in a joint framework. We found habit, goal and longer-term learning effects to be significant drivers of decision-making processes independent of the effects of visual attention. We also found unobserved factors to be significant drivers of choice. Most importantly, we found that ignoring potential unobserved heterogeneity between habits, visual attention and choice outcomes may exaggerate the role of visual attention as a driver of choice leading to low prediction accuracy.

Taking account of each variable separately, we found that time spent on beverage alternative information was positively correlated with the likelihood of choice of that alternative, similar to findings of Balcombe et al. (2015) and others (e.g., Henderson et al. 2003), who did not simultaneously account for multiple drivers of choice, potentially masking unobserved heterogeneity.

Other authors outside of the eye-tracking literature (Camerer, Ho & Chong 2004; Gabaix, Laibson, Moloche & Weinberg 2006) have reported that Markov-like decision models, which consider the influence of previous information acquired on respondent information acquisition behaviours in subsequent choices, provide better data-fit than models which ignore such information acquisition behaviours. This improved predictive power is possibly due to accounting for the endogeneity inherent in such decision-making behaviours. Unlike prior modelling approaches, our more comprehensive approach allows both prior preferences and goal and constraint-based screening to co-exist simultaneously as drivers of choice within a probabilistic approach. While we did not find a significant direct effect of habit and goals on utility, our model allows for this mechanism to be explored in future studies. These advances could be used to identify the mechanism of effect of different cognitive and environmental influences on health or non-health behaviour and purchasing decisions, and thus identify targets for effective intervention. The high predictive power demonstrated by out-of-sample predictions further highlights the need for joint modelling of influences on decision-making, to better identify the potential effect of interventions and the influence of different goals and influences for targeting.

The superior fit of the joint model with AR(1) structure on the unobserved part of utility using time as a preference driver suggests that a significant portion of utility explanatory power is in the unobserved factors affecting choice. Of course, there are a number of decision-making heuristics that our model could be adapted to account for, while harnessing the strength of our model of also accounting for other competing influences on choice rather than considering eye tracking data in isolation. These include the influence of 'row-based' visual attention or 'elimination by aspects' strategy whereby a given attribute is compared to a threshold or between alternatives (Tversky 1972), and 'column-based' visual attention strategies suggesting an 'additive compensatory-model' approach in which all attributes for a given alternative are considered before moving on to the next alternative (Keeney & Raiffa 1993). Visual attention data could be used following our suggested approach to provide evidence for 'row' and 'column' behavioral processes jointly, while accounting for other influences on choice as we have done, aiding decision-making in health and non-health DCEs.

Our model provides evidence of several pathways whereby previous choices and attention may influence subsequent choice and attention. We observed that respondents tended to spend more time on an alternative if it was chosen in the previous task occasion. This may suggest that the previously chosen alternative works as an anchor in the shorter-term, and other options are then evaluated in comparison to the anchor in a binary fashion. This is similar to the 'drift diffusion model' in psychology (Krajbich & Rangel 2011). Independently, we found that the cumulative sum of choice of an alternative in previous choice tasks increased the probability of choice in subsequent tasks ($D_{j,t,}$). This is consistent with the choice perseveration model (Senftleben, Schoemann, Schwenke, Richter, Dshemuchadse & Scherbaum 2019) whereby previous choices cumulatively bias a respondent such that the likelihood of choosing an alternative increases with subsequent choices.

As discussed in our review of the literature, choice set formation theory proposes that such heuristics may be preceded by an initial screening step in which the set of alternatives to be further considered is narrowed (e.g., Swait 1984; Ben-Akiva & Boccara 1995). Pre-determined or 'inherent' preferences, habits and goals (Tversky & Thaler 1990; Simonson 2008) may drive this screening behavior. Variation in choice set formation behavior could be further explored using visual attention data by parameterizing the constraints as a function of visual attention as done in our penalty approach. Future comprehensive models should ideally extend our framework to accommodate multiple decision-making strategies simultaneously. Similarly, interactions with non-health goals could be explored, for example cost-saving. Further work should test the causal relationships between decision-making variables we have proposed using exogenous source of variation.

Finally, our findings suggest that visual attention time does influence choice in complex ways and our model provides a means of exploring the effect of intentionally varying visual attention duration on choice. Marketers or policy makers who wish to influence choice should consider the potential influence that shortening or lengthening consideration time may have on choice, or the influence of factors that may affect visual attention on choice, which in our case study might affect the healthiness of beverage purchases. For example, the removal of SSBs from display has been found to reduce sales of these beverages and increase sales of healthier alternatives in a real-world café setting (Huse et al. 2016).

The interaction of 'top-down' and 'bottom-up' processing pathways in consumer decision-making has significant implications for business, including in the design of product packaging (Orquin, Bagger, Lahm, Grunert & Scholderer 2019) and store layout and product positioning (Valenzuela, Raghubir & Mitakakis 2013). For example, observed retail practice of product positioning and consumers perceptions of product positioning strategies have been shown to interact to influence purchasing behaviour (Valenzuela, Raghubir & Mitakakis 2013). Not accounting for these interactions may cause poor predictions of consumer behaviour and sub-optimal category management. On the other hand, product positioning strategies could be optimised by better understanding this interaction. For example, Valenzuela et al. (2013) suggest initial positioning of products during an introductory period could be aligned with consumer expectations about the position of popular or cheaper products, which may later persist in future purchases due to learning effects, even after products have been moved to less salient (expensive) positions.

6.0 CONCLUSIONS

In this study, we developed an integrated model to analyze the relationship between information acquisition, inferred from visual attention and choice outcome while accounting for stated participant goals and habits. We observed that the frequent practice in previous literature of ignoring the effect of these top-down influences on both visual attention and choice may exaggerate the role of visual attention as a driver of choice. Most notably, we have added to the literature by developing a model that incorporates both observed characteristics (goals and habits) and unobserved characteristics and observed choice history. The model developed here enables researchers to test the guiding effect of observed and unobserved characteristics on visual attention thus providing insight into decision-making strategies and interventions to modify visual stimuli in health, business, and beyond. We hope that the current study will provide a framework to help health and non-health researchers establish the practical validity of eye-tracking data in the context of choice modelling while accounting for other competing influences on choice.

7.0 REFERENCES

Anselin, L. (2001). A companion to theoretical econometrics. In Baltagi, B. H. (Ed.), *Spatial econometrics* (pp. 310-330). Malden, MA: Blackwell Publishing Ltd.

Ares, G., Mawad, F., Giménez, A., & Maiche, A. (2014). Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference*, *31*, 28-37.

Balcombe, K., Fraser, I., & McSorley, E. (2015). Visual attention and attribute attendance in multi-attribute choice experiments. *Journal of Applied Econometrics*, *30*(3), 447-467.

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, *12*(1), 9-24.

Ben-Elia, E., & Shiftan, Y. (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A: Policy and Practice*, *44*(4), 249-264.

Bhat, C. R. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B: Methodological*, *79*, 50-77.

Bhat, C. R., & Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B: Methodological*, *67*, 68-85.

Bhat, C. R., Pinjari, A. R., Dubey, S. K., & Hamdi, A. S. (2016). On accommodating spatial interactions in a generalized heterogeneous data model (GHDM) of mixed types of dependent variables. *Transportation Research Part B: Methodological*, *94*, 240-263.

Biener, L., & Abrams, D. B. (1991). The Contemplation Ladder: validation of a measure of readiness to consider smoking cessation. *Health Psychology*, *10*(5), 360.

Blake, M. R., Lancsar, E., Peeters, A., & Backholer, K. (2018). The effect of sugar-sweetened beverage price increases and educational messages on beverage purchasing behavior among adults. *Appetite*, *126*, 156-162.

Blake, M. R., Lancsar, E., Peeters, A., & Backholer, K. (2019). Sugar-sweetened beverage price elasticities in a hypothetical convenience store. *Social Science and Medicine*, *225,* 98-107.

Bunch, D. S. (1991). Estimability in the multinomial probit model. *Transportation Research Part B: Methodological*, *25*(1), 1-12.

Büttner, O. B., Wieber, F., Schulz, A. M., Bayer, U. C., Florack, A., & Gollwitzer, P. M. (2014). Visual attention and goal pursuit: deliberative and implemental mindsets affect breadth of attention. *Personality and Social Psychology Bulletin*, *40*(10), 1248-1259.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861-898.

Chandon, P., Hutchinson, J. W., Bradlow, E. T., & Young, S. H. (2009). Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing*, *73*(6), 1-17.

Connors, R. D., Hess, S., & Daly, A. (2014). Analytic approximations for computing probit choice probabilities. *Transportmetrica A: Transport Science*, *10*(2), 119-139.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201.

Craig, P. (2008). A new reconstruction of multivariate normal orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 227-243.

Daziano, R. A. (2015). Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model. *Transportation Research Part B: Methodological*, *76*, 1-26.

Erdem, T., Swait, J., Broniarczyk, S., Chakravarti, D., Kapferer, J.-N., Keane, M., Roberts, J., Steenkamp, J.-B. E., & Zettelmeyer, F. (1999). Brand equity, consumer learning and choice. *Marketing Letters*, *10*(3), 301-318.

Fenko, A., Nicolaas, I., & Galetzka, M. (2018). Does attention to health labels predict a healthy food choice? An eye-tracking study. *Food Quality and Preference*, *69*, 57-65.

Franzese, R. J., Hays, J. C., & Schaffer, L. M. (2010). Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. *APSA 2010 Annual Meeting* https://ssrn.com/abstract=1643867

Gabaix, X., Laibson, D., Moloche, G., & Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *The American Economic Review*, *96*(4), 1043-1068.

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102.

Haider, H., & Frensch, P. A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied*, *5*(2), 129.

Hajivassiliou, V., McFadden, D., & Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics*, *72*(1), 85-134.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*(1-3), 43-64.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49-63.

Henderson, J. M., Williams, C. C., Castelhano, M. S., & Falk, R. J. (2003). Eye movements and picture processing during recognition. *Perception and Psychophysics*, *65*(5), 725-734.

Huse, O., Blake, M. R., Brooks, R., Corben, K., & Peeters, A. (2016). The effect on drink sales of removal of unhealthy drinks from display in a self-service café. *Public Health Nutrition*, *19*(17), 3142-3145.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. New York, NY: Cambridge University Press.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852-13857.

Krucien, N., Ryan, M., & Hermens, F. (2017). Visual attention in multi-attributes choices: What can eye-tracking tell us? *Journal of Economic Behavior & Organization*, *135,* 251-267.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311-1328.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*: Chapman and Hall/CRC.

Meißner, M., & Decker, R. (2010). Eye-tracking information processing in choice-based conjoint analysis. *International Journal of Market Research*, *52*(5), 593.

Meißner, M., Musalem, A., & Huber, J. (2016). Eye tracking reveals processes that enable conjoint choices to become increasingly efficient with practice. *Journal of Marketing Research*, *53*(1), 1-17.

Meyerding, S. G. (2018). Combining eye-tracking and choice-based conjoint analysis in a bottom-up experiment. *Journal of Neuroscience, Psychology, and Economics*, *11*(1), 28.

Mullett, T. L., & Stewart, N. (2016). Implications of visual attention phenomena for models of preferential choice. *Decision*, *3*(4), 231.

Orquin, J., & Scholderer, J. (2011). Attention to health cues on product packages. *Journal of Eyetracking, Visual Cognition and Emotion*, *1*(1), 59-63.

Orquin, J. L., Bagger, M. P., Lahm, E. S., Grunert, K. G., & Scholderer, J. (2019). The visual ecology of product packaging and its effects on consumer attention. *Journal of Business Research*, *111*, 187-195. https://doi.org/10.1016/j.jbusres.2019.01.043

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: a review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190-206.

Patil, P. N., Dubey, S. K., Pinjari, A. R., Cherchi, E., Daziano, R., & Bhat, C. R. (2017). Simulation evaluation of emerging estimation techniques for multinomial probit models. *Journal of Choice Modelling*, *23*, 9-20.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Pieters, R., & Zeelenberg, M. (2007). A theory of regret regulation 1.1. *Journal of Consumer Psychology*, *17*(1), 29-35.

Ran, T., Yue, C., & Rihn, A. (2015). Are grocery shoppers of households with weight-concerned members willing to pay more for nutrtional information on food? *Journal of Food Distribution Research*, *46*(3), 113-130.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: an eye-tracking study. *The American Economic Review*, *101*(2), 900-926.

Rose, J. M., Collins, A. T., Bliemer, M. C., & Hensher, D. A. (2009). *Ngene stated choice experiment design software*, (Version 1.1.2). Sydney, Australia: University of Sydney.

Ryan, M., Krucien, N., & Hermens, F. (2018). The eyes have it: Using eye tracking to inform information processing strategies in multi-attributes choices. *Health Economics*, *27*(4), 709-721.

Sándor, Z., & András, P. (2004). Alternative sampling methods for estimating multivariate normal probabilities. *Journal of Econometrics*, *120*(2), 207-234.

Schulte-Mecklenbeck, M., Kuehberger, A., & Johnson, J. G. (2019). *A handbook of process tracing methods*. New York, NY: Routledge.

Senftleben, U., Schoemann, M., Schwenke, D., Richter, S., Dshemuchadse, M., & Scherbaum, S. (2019). Choice perseveration in value-based decision making: The impact of inter-trial interval and mood. *Acta Psychologica*, *198*, 102876.

Simonson, I. (2008). Will I like a ‘medium’ pillow? another look at constructed and inherent preferences. *Journal of Consumer Psychology*, *18*, 157-171.

Singh, G. M., Micha, R., Khatibzadeh, S., Lim, S., Ezzati, M., & Mozaffarian, D. (2015). Estimated global, regional, and national disease burdens related to sugar-sweetened beverage consumption in 2010. *Circulation*, *132*(8), 639-666.

Souza, F. F. (2015). *Goal-based choice set formation*, PhD Thesis. Adelaide, Australia: University of South Australia.

Spinks, J., & Mortimer, D. (2015). Lost in the crowd? Using eye-tracking to investigate the effect of complexity on attribute non-attendance in discrete choice experiments. *BMC Medical Informatics and Decision Making*, *16*(1), 14.

Swait, J. D. (1984). *Probabilistic choice set generation in transportation demand models*. Cambridge, MA: Massachusetts Institute of Technology.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica, 135*(2), 77-99.

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences, 110*(40), E3858-E3867.

Train, K. (2000). *Halton sequences for mixed logit*: UC Berkeley: Department of Economics.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review, 79*(4), 281-299.

Tversky, A., & Thaler, R. H. (1990). Anomalies: preference reversals. *The Journal of Economic Perspectives, 4*(2), 201-211.

Valenzuela, A., Raghubir, P., & Mitakakis, C. (2013). Shelf space schemas: Myth or reality? *Journal of Business Research, 66*(7), 881-888.

van der Laan, L. N., Hooge, I. T., De Ridder, D. T., Viergever, M. A., & Smeets, P. A. (2015). Do you like what you see? The role of first fixation and total fixation duration in consumer choice. *Food Quality and Preference, 39*, 46-55.

van der Laan, L. N., Papies, E. K., Hooge, I. T., & Smeets, P. A. (2016). Goal-directed visual attention drives health goal priming: an eye-tracking experiment. *Health Psychology, 36*(1), 82-90.

Van Loo, E. J., Grebitus, C., Nayga Jr, R. M., Verbeke, W., & Roosen, J. (2018a). On the measurement of consumer preferences and food choice behavior: the relation between visual attention and choices. *Applied Economic Perspectives and Policy, 40*(4), 538-562.

Van Loo, E. J., Nayga Jr, R. M., Campbell, D., Seo, H.-S., & Verbeke, W. (2018b). Using eye tracking to account for attribute non-attendance in choice experiments. *European Review of Agricultural Economics*, *45*(3), 333-365.

Varin, C. (2008). On composite marginal likelihoods. *AStA- Advances in Statistical Analysis*, *92*(1), 1-28.

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*, 5-42.

Varin, C., & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, *92*(3), 519-528.

Vass, C., Rigby, D., Tate, K., Stewart, A., & Payne, K. (2018). An exploratory application of eye-tracking methods in a discrete choice experiment. *Medical Decision Making*, *38*(6), 658-672.

Yegoryan, N., Guhl, D., & Klapper, D. (2019). Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. *Journal of Business Research*, *111*, 290-304.

Table 1: Parameter estimates for visual attention (total fixation duration on) beverage j, task t, j=1,…,8, t=1,…,T

		Visual attention (fixation duration) on beverage j during task t (Y_{jt}) (t-statistic)						
Theoretical construct [a]	**Explanatory variables**	**Bottled water**	**Energy drink**	**Plain low-fat milk**	**Flavored milk**	**Soft drink (regular)**	**Soft drink (diet)**	**Fruit juice**
Alternative Specific Constant (ASC) (α_j)	Alternative Specific Constant (ASC)	.024 (4.5)	-.170 (-7.7)	-.117 (-9.8)	-.113 (-10.5)	-.198 (-11.3)	-.298 (-16.9)	-.319 (-11.66)
Habit (H_l) [b] (measured by automaticity, base=strongly disagree)	Disagree	-.022 (-2.2)	.084 (2.6)	.090 (3.6)	.084 (6.9)	.123 (2.3)	.096 (3.0)	.068 (2.6)
	Neutral	-.211 (-2.1)	.070 (4.1)	.013 (2.3)	-.021 (-3.0)	.070 (3.4)	-.046 (-2.1)	-.098 (-4.2)
	Agree and strongly agree	-.150 (-5.9)	-.036 (-5.5)	.013 (2.3)	.050 (2.7)	.070 (3.4)	-.046 (-2.1)	-.098 (-4.2)
Health goals (Ψ_k) **(**Intention to drink less SSBs; 1-10ordinal scale, 1=no thought of drinking less, 10=taking action to drink less (base: score 7-10)	Score 1	.114 (3.9)	.012 (2.0)	-.026 (-4.4)	-.037 (-5.5)	-.049 (-6.5)	NS	-.034 (-4.8)
	Score 2	.114 (3.9)	.012 (2.0)	-.026 (-4.4)	-.037 (-5.5)	-.049 (-6.5)	NS	-.034 (-4.8)
	Score 3	.114 (3.9)	.012 (2.0)	-.026 (-4.4)	-.037 (-5.5)	-.049 (-6.5)	NS	-.034 (-4.8)
	Score 4	.114 (3.9)	.012 (2.0)	-.026 (-4.4)	-.037 (-5.5)	-.049 (-6.5)	NS	-.034 (-4.8)
	Score 5	.114 (3.9)	.012 (2.0)	.098 (7.2)	.055 (4.8)	-.049 (-6.5)	.097 (7.9)	.118 (8.9)
	Score 6	.114 (3.9)	.012 (2.0)	.098 (7.2)	.055 (4.8)	-.049 (-6.5)	.097 (7.9)	NS
Learning [c]	Shorter-term ($d_{j,t-1}$): Same alternative chosen in the last choice task (Yes=1, No=0)	.309 (14.7)	.384 (6.4)	.453 (7.9)	.510 (9.4)	.364 (7.9)	.505 (9.5)	.552 (9.8)
Learning (ρ_j, autoregressive parameter)	Time spent on beverage j in previous task (t-1)	.586 (9.5)	.569 (8.0)	.686 (9.8)	.725 (7.3)	.679 (8.7)	.597 (10.8)	.642 (5.1)

NS, not significant. [a] Results for cognitive analysis time (visual attention time on choice experiment, excluding visual attention to alternative and attribute information) were not significant were therefore omitted from the final model and are not reported here. [b] Habit (automaticity): This variable was constructed to measure the automaticity in habit towards drinking SSBs (sugar-sweetened beverages) by taking the average of scores reported for following statements: I consume non-diet cordial, non-diet soft drinks, sports drinks, energy drinks, flavoured milk and fruit drink… (i) Automatically, (ii) Without having to consciously remember, (iii) Without thinking, and (iv) Before I realise I'm drinking it. Four questions on five-point Likert scales from strongly disagree (1) to strongly agree (5). Means were constructed from responses to each of the four items for analysis. [c] Longer-term learning ($D_{j,t}$) results not displayed as all findings non-significant.

Table 2: Parameter estimates for Multinomial Probit (MNP) choice model

		Utility of beverage j, task t (U_{jt}) (t-statistic)							
Theoretical Construct	**Explanatory Variables**	**Bottled water**	**Energy drink**	**Plain low-fat milk**	**Flavored milk**	**Soft drink (regular)**	**Soft drink (diet)**	**Fruit juice**	**No drink**
Alternative Specific Constant (ASC) α_j	Alternative Specific Constant (ASC)	NS	-.170 (-1.88)	-.169 (-2.66)	-.198 (-1.89)	.068 (2.34)	-.048 (-2.42)	.718 (4.51)	.696 (6.46)
Design variable (β_j)	Volume/Price Ratio (ml/AU$)	.523 (20.52)	.403 (12.75)	.220 (6.99)	.782 (8.14)	.598 (10.63)	.680 (12.02)	.713 (14.83)	NS
Shorter-term choice effect ($d_{j,t-1}$)	Same alternative chosen in the last choice task (Yes=1, No=0)	NS	NS	NS	NS	NS	NS	NS	.513 (8.88)
Longer-term choice effect ($D_{j,t,}$)	Cumulative sum of choice of the same alternative until last choice task	.080 (13.20)	.221 (14.20)	.193 (12.75)	.121 (3.50)	.118 (4.98)	.159 (8.34)	.043 (2.10)	.243 (15.15)
Visual attention (Y_{jt})	Natural logarithm of time spent on beverage j, task t	1.374 (15.37)	1.482 (9.83)	1.435 (10.99)	1.199 (11.80)	1.288 (13.85)	1.298 (11.05)	1.155 (13.01)	NS
	Autoregressive parameter value (on unobserved utility)[a]	.573 (7.01)							

NS, not significant. [a] Bottled water has the highest choice share, therefore we take this as the reference alternative.
Results for habit and goal parameters (direct effect of habit and goals on utility) were not statistically significant and were therefore omitted from the final model and are not reported here.

Table 3: Covariance matrix (Ω) parameter estimates

Utility of beverage j, task t (U_{jt})	**Correlation of visual attention across beverage alternatives (Ω_{ξ})**							**Correlation of stochastic utilities across beverage alternatives Ω_{η}**						
	Bottled water	Energy drink	Plain low-fat milk	Flavored milk	Soft drink (regular)	Soft drink (diet)	Fruit juice	Energy drink	Plain low-fat milk	Flavored milk	Soft drink (regular)	Soft drink (diet)	Fruit juice	No drink
Bottled water	.597 (8.51)													
Energy drink	NS	.560 (8.33)												
Plain low-fat milk	.277 (2.11)	.294 (2.95)	.471 (6.96)											
Flavored milk	NS	NS	.174 (7.16)	.477 (8.53)										
Soft drink (regular)	NS	.321 (10.99)	.169	NS	.650 (8.59)									
Soft drink (diet)	.352 (11.47)	NS	.034	-.121	0.353 (4.58)	.721 (6.94)								
Fruit juice	.253 (9.78)	NS	.205 (9.43)	.158 (4.40)	NS	.209 (6.41)	.762 (3.92)							
Energy drink	NS	NS	.025	.016 (1.85)	NS	.014	.222 (4.30)	1.000						
Plain low-fat milk	NS	NS	NS	-.020	NS	.031	0242 (4.53)	.580	.775 (8.22)					
Flavored milk	NS	NS	.070 (2.35)	.128 (2.96)	-.040 (-2.25)	-.021 (-1.15)	.409	.760 (3.32)	.634 (11.83)	1.087 (11.62)				
Soft drink (regular)	.029 (2.54)	NS	-.001	-.047 (-4.77)	.036 (2.13)	.071 (2.10)	.103 (3.56)	.547	.486	.570 (9.49)	.839 (9.28)			
Soft drink (diet)	NS	NS	NS	NS	.052 (2.89)	.040 (2.42)	.006	.504	.219 (2.15)	.425	.391	.805 (8.28)		
Fruit juice	-.008	NS	.013	.018	NS	.023	.228 (6.41)	.723 (8.52)	.597	.727	.588	.332 (5.86)	1.084 (11.33)	
No drink	.398 (13.38)	.135 (6.26)	.308 (7.64)	.112 (9.03)	.248 (10.61)	.418 (12.08)	.527 (12.22)	.619	.589	.641	.572	.441	.526	1.563

NS, not significant.

Table 4: Model fit statistics

Role of Visual Attention (Y_{jt})	Model	Number of parameters	Composite Marginal Likelihood Value	AR(1) parameter value (t-statistic)	Adjusted composite likelihood ratio (*p*-value comparison with AR(1)EJ model)[a]	Composite likelihood information criteria (CLIC)
Used as a preference driver (fully compensatory model)	Joint-AR(1)VE	139	-4943689.19	.016 (2.25)	.446	N/A
	Joint-AR(1)V	139	-4943351.02	-.052 (-2.13)	.475	N/A
	Joint-AR(1)E	137	-4942782.99	.573 (7.01)	N/A	-4942922.25
	Independent-AR(1)VE	135	-4945495.05	.112 (1.84)	.010	N/A
Used to capture screening behavior through penalty function [b] (two step approach)	Joint-AR(1)VEP	138	-4961851.03	.431 (5.42)	N/A	-4961982.10
	Joint-AR(1)EP	138	-4958239.42	.594 (13.13)	N/A	-4958377.42

[a] *p*-value calculation is based on 100 bootstrap samples. [b] Beta values for all penalty function times were positive- results available on request from authors.

Table 5: Model fit for in- and out-of- sample prediction

	Predicted Share								
Model	**Bottled water**	**Energy drink**	**Plain low-fat milk**	**Flavored milk**	**Soft drink (regular)**	**Soft drink (diet)**	**Fruit juice**	**No drink**	**Mean absolute error (MAE)**
In-Sample [a]									
Observed share	.27	.07	.06	.10	.11	.09	.20	.09	
Joint-AR(1)VE	.18	.12	.16	.11	.11	.07	.18	.07	.039
Joint-AR(1)V	.27	.10	.15	.08	.10	.13	.15	.03	.038
Joint-AR(1)E	.24	.09	.12	.15	.10	.09	.16	.05	.031
Independent-AR(1)E	.58	.06	.06	.05	.05	.07	.10	.04	.075
Joint-AR(1)VEP	.15	.04	.23	.18	.13	.10	.16	.01	.069
Joint-AR(1)EP	.21	.05	.11	.17	.11	.12	.21	.03	.038
Out-of-sample [b]									
Observed share	.20	.08	.13	.13	.11	.12	.20	.04	
Joint-AR(1)VE	.19	.12	.17	.11	.11	.07	.17	.08	.029
Joint-AR(1)V	.28	.09	.16	.08	.10	.12	.14	.03	.031
Joint-AR(1)E	.20	.05	.12	.16	.10	.12	.21	.03	.013
Independent-AR(1)E	.59	.06	.06	.05	.05	.07	.09	.04	.098
Joint-AR(1)VEP	.14	.04	.24	.18	.14	.10	.16	.01	.048
Joint-AR(1)EP	.25	.09	.13	.15	.10	.09	.14	.05	.024

[a] Sample size=2780 (139 individuals with 20 choice tasks)
[b] Sample size=206 (13 individuals with varying number of choice tasks)

Table 6: Average treatment effect (ATE) on probability of choosing a particular option due to 10% increase in total time spent looking at that option including attribute values (standard errors): comparison of independent and joint model performance

Alterative	Baseline observed choice share	ATE for 10% increase in fixation time Independent-AR(1)E	ATE for 10% increase in fixation time Joint-AR(1)E model [a]	*p*-value	True effect [b]	Spurious effect [c]
Bottled water	.27	.031 (.003)	.020 (.003)	.005	65%	35%
Energy drink	.07	.018 (.002)	.010 (.002)	.002	56%	44%
Plain low-fat milk	.06	.019 (.002)	.010 (.002)	.001	53%	47%
Flavored milk	.10	.018 (.002)	.011 (.002)	.007	61%	39%
Soft drink (regular)	.11	.018 (.002)	.011 (.002)	.007	61%	39%
Soft drink (diet)	.09	.016 (.002)	.010 (.002)	.017	63%	37%
Fruit juice	.20	.023 (.003)	.015 (.002)	.013	65%	35%
None	.09					

[a] ATE values are based on 500 model estimation repetitions.
[b] The true effect is the ratio of share estimations from the joint model/ independent model estimations.
[c] Additional percentage of share not accounted for by true effect.

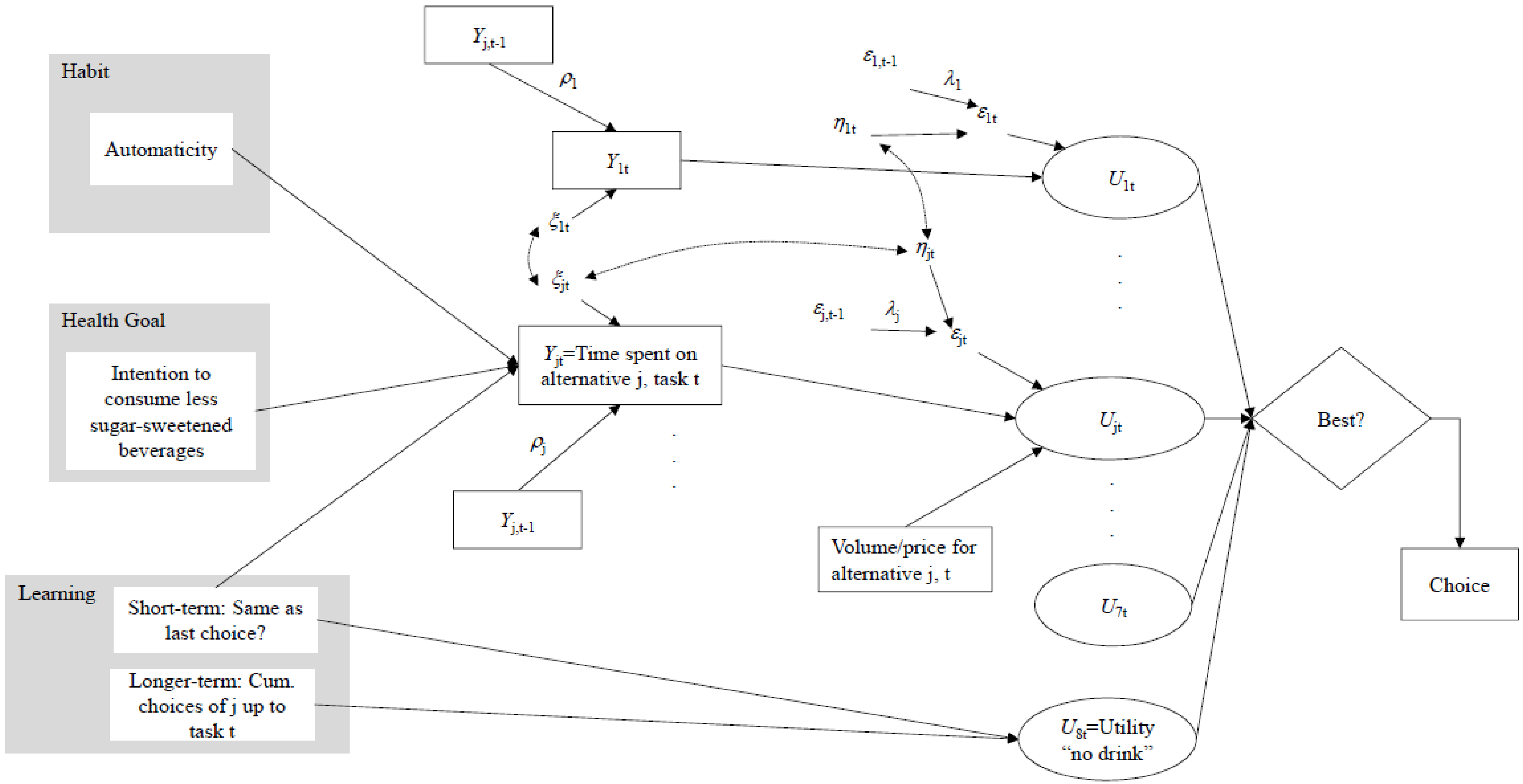


Fig. 1: Econometric model schematic

WEB APPENDIX A: DISCRETE CHOICE EXPERIMENT DETAILS

Experimental Design

In the labelled DCE, participants selected a beverage within a hypothetical convenience store setting. Each participant completed 20 choice tasks involving three SSB alternatives (energy drink, flavored milk, regular soft drink (soda)), four non-sugar-sweetened alternatives (non-SSBs: plain low-fat milk, fruit juice, diet soft drink, bottled water), and a "no drink" alternative (meaning that they would "consume no drink on this occasion"). Each beverage was described by alternative-specific prices and generic volume attributes which varied over four levels each. An orthogonal design was generated using Ngene software (Rose, Collins, Bliemer & Hensher 2009). An example choice task (**Figure A.1**) and list of attribute levels for each alternative (**Table A.1**) are given below. Prior to completing the choice tasks, half of participants were randomly exposed to a real-world educational message designed to discourage selection of SSBs. The other half did not see any message. After the 10^{th} choice task all participants were presented with a message reminding them to "consider their options carefully", to ameliorate potential fatigue effects. As described later, we tested for the impact of the education message in the analysis and found no significant effect on beverage choice, hence sub-samples were pooled and we used the full sample in the estimation results we present later.

Following the DCE, participants completed questions on stated attendance to attributes and alternatives as well as strength of SSB consumption habit. This included an 11-point scale of readiness to consider reducing SSB intake based on a validated tool to assess readiness to quit smoking (Biener & Abrams 1991) and the Self-Report Behavioral Automaticity Index, a 4-item measure of habit strength measured on a 5-point Likert scale with higher scores signifying a stronger habit (Gardner, Abraham, Lally & de Bruijn 2012).

Please read the information below carefully:

For this survey, imagine that you are now going into your local convenience store (e.g. newsagent, 7-Eleven, independent milk bar) with the intention to buy a pre-packaged drink (in a bottle, can or carton) for yourself to drink immediately. Please note that this does not include supermarkets or petrol stations, hospital, sports and recreation facilities etc. where you may have entered the store for another main purpose.

You will be presented with a number of individual shopping scenarios. In each scenario, you will be presented with 7 drink options, each drink will be described by its price and size (volume). The information describing price and volume will change between each task. Assume the displayed products are the only available options.

Please note that 'energy drink' refers to a drink specifically designed to give a short term 'energy' boost such as those with added taurine, guarana or caffeine. It does not include 'sports drinks'.

In each scenario, please indicate which **one option** you would choose. Either select the drink that you would buy OR select 'no beverage' if you would exit the convenience store without having purchased a drink in this situation, after already having entered the convenience store with the intention to buy a pre-packaged drink. This would mean you would not consume a drink on this occasion.

Please also treat each scenario as separate (i.e. as if you had not just made the previous choice).

Please note: there are no right or wrong answers, the researchers are interested in your individual preference among the options presented.

On the next page will be a practice scenario.

Fig. A.1 (part 1): Discrete Choice Experiment scenario explanation and sample choice scenario

You have gone into your local convenience store now (e.g. newsagent, 7-Eleven, independent milk bar) with the intention to buy a pre-packaged drink (in a bottle, can or carton) to drink immediately yourself. Select the option below that you would choose.

	Energy drink	Plain low-fat milk	Flavoured milk	Bottled water	Soft drink (regular)	Soft drink (diet)	Fruit juice	No drink
Price	$5.90	$5.00	$6.50	$1.00	$6.50	$6.50	$5.90	N/A
Volume (size)	200mL	200mL	200mL	600mL	200mL	200mL	200mL	N/A
Which would you choose?	○	○	○	○	○	○	○	○

Fig. A.1 (part 2): Discrete Choice Experiment scenario explanation and sample choice scenario

Reprinted from Appetite, Vol. 126, Blake MR, Lancsar E, Peeters A, Backholer K, The effect of sugar-sweetened beverage price increases and educational messages on beverage purchasing behavior among adults, 156-162, Copyright (2018), with permission from Elsevier.

Table A.1: Alternative attribute levels

Alternative	Experimental volumes tested	Experimental prices tested (AUD)
Energy drink	200mL, 330mL, 460mL, 600mL	$2.00, $3.30, $4.60, $5.90
Soft drink (regular)		$2.00, $3.50, $5.00, $6.50
Soft drink (diet)		$2.00, $3.50, $5.00, $6.50
Plain low-fat milk		$1.00, $2.30, $3.70, $5.00
Flavoured milk		$2.00, $3.50, $5.00, $6.50
Bottled water		$1.00, $2.30, $3.70, $5.00
Fruit juice		$2.00, $3.30, $4.60, $5.90

Implementation of Eye-Tracking Measurements

All participants completed the DCE in an eye-tracking laboratory at the study university. The task involved sitting and completing the DCE on a computer-screen. A discrete, web-cam like device tracked eye movements (Tobii Pro, 2011, Tobii TX300; Stockholm, Sweden). The choice tasks were presented through a web-browser using Tobii Studio version 3.2 (Tobii Pro, 2012, Stockholm, Sweden). Eye movements were recorded at 300 Hz on a screen resolution of 1920 x 1080 pixels. Minimum fixation duration was 60ms.

Participants were positioned with their head 64cm from the screen as per recommended Tobii T-series validity requirements. Participants' eye-movements were calibrated before the experiment using nine static calibration locations on the screen. Participants were eye-tracked during the entire survey, however only visual attention data corresponding to the DCE are analyzed here.

References

Biener, L., & Abrams, D. B. (1991). The Contemplation Ladder: validation of a measure of readiness to consider smoking cessation. *Health Psychology*, *10*(5), 360.

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102.

Rose, J. M., Collins, A. T., Bliemer, M. C., & Hensher, D. A. (2009). *Ngene stated choice experiment design software*, (Version 1.1.2). Sydney, Australia: University of Sydney.

WEB APPENDIX B: DETAILED METHODOLOGY

Our model has three components: continuous (visual attention duration), ordered (habit measures), and nominal (choice outcome) variables. We first describe the construction of each component separately and then bring them together using a covariance approach.

Visual Attention Model

Let $\tilde{t}$ be the index for task instance ($\tilde{t}$ = 1,2,…, $\tilde{T}$) and $\tilde{h}$ be the index for the continous outcome ($\tilde{h}$ = 1,2,…, $\widetilde{H}$). Then, we can write in the usual linear regression form:

(1) $\tilde{y}_{\tilde{h},\tilde{t}} = \tilde{\rho}_{\tilde{h}}\tilde{y}_{\tilde{h},\tilde{t}-1} + \gamma'_{\tilde{h}}x_{\tilde{h},\tilde{t}} + \xi_{\tilde{h}}$

Where $\tilde{\rho}_{\tilde{h}}$ is the autoregressive (AR-1) coefficient which ranges between -1 to 1, $x_{\tilde{h},\tilde{t}}$ is a $(k_{\tilde{h}} \times 1)$ vector of exogenous variables (including a constant), $\gamma_{\tilde{h}}$ is the corresponding $(k_{\tilde{h}} \times 1)$ vector of coefficients, and $\xi_{\tilde{h}}$ is a normally distributed error term. The autoregressive coefficient helps us capture the time-multiplier effect (i.e., the effect of previous time period on the current time period for both observed and unobserved variables). Now, stack all the $\widetilde{H}$ continuous outcomes for all task instances $\tilde{T}$ in a vector $\tilde{y} = (\tilde{y}_{1,1}, \tilde{y}_{2,1}, \dots, \tilde{y}_{\widetilde{H},1}, \dots, \tilde{y}_{\widetilde{H},\tilde{T}})$ $(\widetilde{H}\tilde{T} \times 1)$, autoregressive coefficient $\tilde{\rho}_{\tilde{h}}$ for all the $\widetilde{H}$ continuous outcomes in a vector $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, \tilde{\rho}_2, \dots, \tilde{\rho}_{\widetilde{H}})$ of size $(\widetilde{H} \times 1)$, exogenous variable's coefficients in a matrix $\gamma = (\gamma'_1, \gamma'_2, \dots, \gamma'_{\widetilde{H}})$ of size $(\widetilde{H} \times k_{\tilde{h}})$, exogenous variables in a matrix $x_{\widetilde{H},\tilde{T}} = (x'_{1,1}, x'_{2,1}, \dots, x'_{\widetilde{H},1} \dots, x'_{\widetilde{H},\tilde{T}})$ of size $(\widetilde{H}\tilde{T} \times k_{\tilde{h}})$ and all the error terms in $(\xi = \xi_1, \xi_2, \dots, \xi_{\widetilde{H}})$ of size $(\widetilde{H} \times 1)$. Where (,…) inside the bracket refers to placement of next variable in the next row. Also, let Ξ be the covariance matrix of ξ.

Now, to write the equation (1) in the matrix form, define the following matrices:

construct a matrix $\mathbf{F}_{\widetilde{H}\tilde{T}}$ of size $(\widetilde{H}\tilde{T} \times \widetilde{H}\tilde{T})$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of matrix $\mathbf{F}_{\widetilde{H}\tilde{T}}$.

for j = 2 to $\tilde{T}$

for i = 1 to $\widetilde{H}$

$$\mathbf{F}_{\widetilde{H}\tilde{T}}\left[(j-1) * \ \widetilde{H}+i,(j-1) * \widetilde{H}+i\right] = \ \tilde{\rho}[i,1]$$

end

end

For example: a $F_{\widetilde{H}\tilde{T}}$ matrix with $\widetilde{H}=2$ and $\tilde{T}=3$ will take the following form:

$$\mathbf{F}_{\widetilde{H}\tilde{T}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \tilde{\rho}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \tilde{\rho}_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{\rho}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{\rho}_2 & 0 & 0 \end{bmatrix}$$

Also, construct a matrix $\mathbf{I}_{\widetilde{H}\tilde{T}}$ of size ($\widetilde{H}\tilde{T} \ \times \widetilde{H}\tilde{T}$) with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix $\mathbf{I}_{\widetilde{H}\tilde{T}}$

for j = 2 to $\tilde{T}$

for i = 1 to $\widetilde{H}$

$$I_{\widetilde{H}\tilde{T}}\left[(j-1) * \ \widetilde{H}+i,(j-2) * \widetilde{H}+i\right] = \ 1$$

end

end

With this, equation (1) may be written in the matrix form as follows:

(2) $\tilde{y}$=S*[sumc[$(\tilde{\gamma}.* \, x_{\widetilde{H}\tilde{T}})'$] $+ \tilde{\xi}$]

where $\tilde{\gamma}$=ones($\tilde{T}$, 1).*.γ, $\tilde{\xi} = \text{ones}(\tilde{T},1).*.\xi$, ".*." refers to Kronecker product, ".*" refers to element by element multiplication, the operator sumc(.) returns the sum of columns of matrix in a column vector, ones($\tilde{T}$, 1) indicates a vector of size $\tilde{T}$ whose all the elements are filled with a value of "1", $\mathbf{1}_{\tilde{H}\tilde{T}}$ refers to an identity matrix of size $\tilde{H}\tilde{T}$ and $\boldsymbol{S} = [\mathbf{1}_{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{T}}} - (\boldsymbol{F}_{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{T}}}.*.\boldsymbol{I}_{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{T}}})]^{-1}$ of size $(\tilde{H}\tilde{T} \times \tilde{H}\tilde{T})$.

From equation (2), it can be observed that y ̃ is distributed normally with mean S*[sumc[(γ ̃.*x_(H ̃T ̃))'] and covariance $\mathbf{S}*[\mathbf{1}_{\tilde{\boldsymbol{T}}}.*.\,\Xi] * \boldsymbol{S}'$ [2]. Also, to maintain the bound on the autoregressive parameter vector $\tilde{\boldsymbol{\rho}}$, we parametrize the parameter as $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}_{\boldsymbol{p}}/[\mathbf{1} + (\tilde{\boldsymbol{\rho}}_{\boldsymbol{p}})^{\mathbf{2}}]^{\mathbf{0.5}}$. Where $\tilde{\rho}_p$ is the value passed to the optimization module.

Habit and Goal Variable Model

Strength of habit and goals were considered on an ordinal scale. Let $\ddot{t}$ be the index for task instance ($\ddot{t}$ = 1,2,…,$\ddot{T}$) and $\ddot{n}$ be the index for the ordinal outcome ($\ddot{n}$ = 1,2,…,$\ddot{N}$). Also, let $J_{\ddot{n}}$ (>1) be the number of categories for the $\ddot{n}^{th}$ordinal outcome and the correponding index be $j_{\ddot{n}} = (1,2, \ldots, J_{\ddot{n}})$. [3]Let $\ddot{y}^{*}_{\ddot{n},\ddot{t}}$ be the underlying latent variable. Then in the usual ordered-response formulation, we may write:

---

[2] In a time-series based regression such as the one described here, the dependence between a particular continuous variable's task instances or time periods is generated through the autoregressive parameter and the dependence across continuous variables is captured through the covariance matrix Ξ. This allows the analyst to exclude random taste heterogeneity in the model. Our experience with the model suggests that recovery of random parameters in such a highly non-linear model is relatively difficult. Therefore, we suggest the inclusion of either autoregressive parameters or random taste parameters in the model depending upon the analyst's requirement. Random taste parameters can be included in a straighforward manner as follows: let $\boldsymbol{\Omega}$ be a $(k_{\tilde{h}} \times k_{\tilde{h}})$ covariance matrix of exogenous variables. Then, stack the exogenous variables in a matrix of size ($\tilde{H}\tilde{T} \times k_{\tilde{h}}\tilde{T}$ as follows:

$$X_{\tilde{H}\tilde{T}} = \begin{bmatrix} x'_{1,1} & 0 & 0 & 0 \\ x'_{2,1} & 0 & 0 & 0 \\ x'_{\tilde{H},1} & 0 & 0 & 0 \\ 0 & \vdots & 0 & 0 \\ 0 & \vdots & 0 & \vdots \\ 0 & 0 & \cdots & x'_{\tilde{H}\tilde{T}} \end{bmatrix}$$

With this $\tilde{\boldsymbol{y}} \sim \boldsymbol{N}[\mathbf{S} * [\mathbf{sumc}[(\tilde{\boldsymbol{\gamma}}.*\,\boldsymbol{x}_{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{T}}})'], \mathbf{S} * [\mathbf{1}_{\tilde{\boldsymbol{T}}}.*.\,\Xi + \boldsymbol{X}_{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{T}}} * (\mathbf{1}_{\tilde{\boldsymbol{T}}}.*.\,\boldsymbol{\Omega}) * \boldsymbol{X}'_{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{T}}}] * \boldsymbol{S}']$

[3] The requirement of number of categories to be greater than 1 instead of 2 enables us to model binary outcomes as ordinal outcomes with no additional thresholds being estimated.

(3) $\ddot{y}^{*}_{\ddot{n},\ddot{t}} = \delta'_{\ddot{n}} x_{\ddot{n},\ddot{t}} + \zeta_{\ddot{n}}$, and $\Psi_{\ddot{n},a_{\ddot{n},\ddot{t}}-1} < \ddot{y}^{*}_{\ddot{n},\ddot{t}} < \Psi_{\ddot{n},a_{\ddot{n},\ddot{t}}}$, if $\ddot{y}_{\ddot{n},\ddot{t}}=a_{\ddot{n},\ddot{t}}$

where $x_{\ddot{n},\ddot{t}}$ is a $(k_{\ddot{n}} \times 1)$ vector of exogenous variables (including constant)[4] , $\delta_{\ddot{n}}$ is the corresponding $(k_{\ddot{n}} \times 1)$ vector of parameters, $a_{\ddot{n},\ddot{t}}$ is the observed outcome category at time period $\ddot{t}$ for the $\ddot{n}^{th}$ ordinal variable, and $\zeta_{\ddot{n}}$, is a standard normal error term[5]. Further, the thresholds for the ordinal outcome should be in ascending order (i.e.,

$\Psi_{\ddot{n},0} < \Psi_{\ddot{n},1} < \cdots < \Psi_{\ddot{n},J_{\ddot{n}}-1} < \Psi_{\ddot{n},J_{\ddot{n}}}$; $\Psi_{\ddot{n},0} = -\infty, \Psi_{\ddot{n},1}$=0, and $\Psi_{\ddot{n},J_{\ddot{n}}} = \infty$).

Now, stack the threshold elements as follows:

$\Psi_{\ddot{n}} = (\Psi_{\ddot{n},0}, \Psi_{\ddot{n},1}, \dots, \Psi_{\ddot{n},J_{\ddot{n}}})$ $[(J_{\ddot{n}} + 1) \times 1]$ vector,

$\Psi_{\ddot{t}} = (\Psi'_1, \Psi'_2, \dots, \Psi'_{\ddot{N}})'$ $[\ddot{N}(J_{\ddot{n}} + 1) \times 1]$ vector,

$\Psi_{low} = (\Psi_{1,a_{1,1}-1}, \Psi_{1,a_{2,1}-1}, \dots, \Psi_{1,a_{\ddot{N},1}-1}, \dots, \Psi_{1,a_{\ddot{N},\ddot{T}}-1})$ $[\ddot{N}\ddot{T} \times 1]$ vector, and

$\Psi_{up} = (\Psi_{1,a_{1,1}}, \Psi_{1,a_{2,1}}, \dots, \Psi_{1,a_{\ddot{N},1}}, \dots, \Psi_{1,a_{\ddot{N},\ddot{T}}})[\ddot{N}\ddot{T} \times 1]$ vector [6].

Further, stack the $\ddot{N}\ddot{T}$ underlying latent variables in a $(\ddot{N}\ddot{T} \times 1)$ vector $\ddot{\boldsymbol{y}}^{*} = (\ddot{y}^{*}_{1,1}, \ddot{y}^{*}_{2,1}, \dots, \ddot{y}^{*}_{\ddot{N},1}, \dots, \ddot{y}^{*}_{\ddot{N},\ddot{T}})$, exogenous variables in a matrix $\boldsymbol{x}_{\ddot{N},\ddot{T}} = (\boldsymbol{x}'_{1,1}, \boldsymbol{x}'_{2,1}, \dots, \boldsymbol{x}'_{\ddot{N},1}, \dots, \boldsymbol{x}'_{\ddot{N},\ddot{T}})$ of size $(\ddot{N}\ddot{T} \times k_{\ddot{n}})$, exogenous variables' coefficients in a matrix $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2, \dots, \boldsymbol{\delta}'_{\ddot{N}})$ of size $(\ddot{N} \times k_{\ddot{n}})$, and all the error terms in $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \dots, \zeta_{\ddot{N}})$ of size $(\ddot{N} \times 1)$. Also, let Γ be the correlation matrix of ζ. Then, we may write, equation (3) in the matrix form as follows:

(4) $\ddot{y}^{*} = \text{sumc}\left[\left(\ddot{\boldsymbol{\delta}} .* \boldsymbol{x}_{\ddot{N},\ddot{T}}\right)'\right] + \boldsymbol{\zeta}, \boldsymbol{\Psi_{low}} < \ddot{\boldsymbol{y}}^{*} < \boldsymbol{\Psi_{up}}$ [7]

where $\boldsymbol{\delta} = ones(\ddot{T}, 1).*.\boldsymbol{\delta}$ and $\ddot{\boldsymbol{\zeta}} = ones(\ddot{T}, 1).*.\boldsymbol{\zeta}$.

---

[4] We fix the second threshold to a value of zero and thus estimate the constant for every ordinal outcome.

[5] The normalization on the error term is needed for identification, as in the usual ordered-response model; see McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*(1), 103-120..

[6] Here for ease in notation, we assume that all the ordinal outcomes have same number of categories. However, this may not be the case. In situations with different number of categories, one can fill the remaining/extra cells with zeros.

[7] If the ordinal outcomes are observed for more than one time period, then one would be tempted to include random-taste parameters in order to capture the dependence across time-periods. Similar to the

Choice Model

Let t be the index for choice occasion $(t = 1,2,\ldots,T)$, i be the index for nominal outcome $(i = 1,2,\ldots,I)$, and k be the index for number of alternatives per nominal outcome $(k = 1,2,\ldots,K)$[8]. Then, we can write the utility of alternative k from the i^{th} nominal variable in the time period t as:

(5) $U_{i_k t} = \boldsymbol{\beta}_i' \boldsymbol{x}_{\boldsymbol{i_k t}} + \boldsymbol{\varepsilon}_{\boldsymbol{i_k}}$

where $x_{i_k t}$ is a (g_i ×1) vector of exogenous variables at choice occasion t, β_i is the corresponding (g_i ×1) vector of coefficients, and ε_{i_k} is a normally distributed error term (all the notations correspond to the nominal outcome i). Now, define the following notations:

I_k (total number of alternatives) = $\sum_{t=1}^{I} \boldsymbol{i_K}$,

U_{it} = $(U_{1t}, U_{2t}, \ldots, U_{i_K t})[(i_K \times 1)]$ vector, U_t = $(U_{1t}, U_{2t}, \ldots, U_{It})[(I_K \times 1)]$ vector,

$\boldsymbol{U}$ = $(\boldsymbol{U_1}, \boldsymbol{U_2}, \ldots, \boldsymbol{U_T})[(I_K \times 1)]$ vector, $\boldsymbol{\beta}$=$(\boldsymbol{\beta}'_{\boldsymbol{11}}, \boldsymbol{\beta}'_{\boldsymbol{12}}, \ldots, \boldsymbol{\beta}'_{\boldsymbol{1K}}, \ldots, \boldsymbol{\beta}'_{\boldsymbol{IK}})[(I_K \times g_i)]$ vector,

x=$(x'_{11t}, x'_{12t}, \ldots, x'_{1Kt}, \ldots, x'_{IKt}[(I_K \times g_i)]$ matrix, x=$(x_1, x_2, \ldots, x_T)[(TI_K \times g_i)]$ matrix,

ε_i = $(\varepsilon_{i_1}, \varepsilon_{i_2}, \ldots, \varepsilon_{i_K})[(i_k \times 1)]$ vector, $\boldsymbol{\varepsilon_t}$ =$(\varepsilon_{1_1}, \varepsilon_{1_2}, \ldots, \varepsilon_{1_k}, \ldots, \varepsilon_{I_k})$ $[(I_K \times 1)]$ vector,

$\tilde{\beta} = [\text{ones}(T,1).*.\beta][(TI_K \times g_i)]$ matrix, and ε = [ones(T,1).*. ε_t]$[(TI_K \times 1)]$ vector.

Also, let $\boldsymbol{\Lambda}_i$be the covariance matrix of ε_i. Then, we may write, equation (5) in the matrix form as follows:

---

continuous variable model, the incorporation of random-taste parameter is straightforward. Let $\boldsymbol{\Psi}$ be a $(k_{\ddot{n}} \times k_{\ddot{n}})$, covariance matrix of exogenous variables. Then, stack the exogenous variables in a matrix of size $(\ddot{N}\ddot{T} \times k_{\ddot{n}}\ddot{T})$ as follows: $X_{\ddot{N}\ddot{T}} = \begin{bmatrix} x'_{1,1} & 0 & 0 & 0 \\ x'_{2,1} & 0 & 0 & 0 \\ x'_{\ddot{N},1} & 0 & 0 & 0 \\ 0 & \vdots & 0 & 0 \\ 0 & \vdots & 0 & \vdots \\ 0 & 0 & \ldots & x'_{\ddot{N},\ddot{T}} \end{bmatrix}$

With this the covariance matrix for $\ddot{\boldsymbol{y}}^*$ becomes $[1_{\ddot{T}} .*. \Gamma + X_{\ddot{N}\ddot{T}} * (1_{\ddot{T}}.*. \Psi) * X'_{\ddot{N}\ddot{T}}]$

[8] We supress the index for the individual participant (i) for ease in presentation as it is a non-spatial model.

(6) $\boldsymbol{U} = \mathbf{sumc}\left[(\tilde{\boldsymbol{\beta}}.*\boldsymbol{x})'\right] + \boldsymbol{\varepsilon}$

With this, we may write the distribution of $\boldsymbol{U}$ as

$\boldsymbol{U} \sim N_{(TI_K \times TI_K)}\left[\text{sumc}\left[\left(\tilde{\beta}.*x\right)'\right],\ \mathbf{1}_T.*.\boldsymbol{\Lambda}\right]$. Where,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda_1} & \boldsymbol{\Lambda_{12}} & \boldsymbol{\Lambda_{1,I-1}} & \boldsymbol{\Lambda_{1,I}} \\ \boldsymbol{\Lambda'_{12}} & \boldsymbol{\Lambda_2} & \boldsymbol{\Lambda_{2,I-1}} & \boldsymbol{\Lambda_{2,I}} \\ \boldsymbol{\Lambda'_{1,I-1}} & \boldsymbol{\Lambda'_{2,I-1}} & \ddots & \boldsymbol{\Lambda_{I-1,I}} \\ \boldsymbol{\Lambda'_{1,I}} & \boldsymbol{\Lambda'_{2,I}} & \boldsymbol{\Lambda'_{I-1,I}} & \boldsymbol{\Lambda_I} \end{bmatrix}, \text{ and}$$

In the $\boldsymbol{\Lambda}$ matrix, the off-diagonal elements capture dependencies across nominal variables through correlation in unobserved variables[9].

Since only the differences in utility matter, only the difference of error-terms are identifiable and not the actual error terms after performing the normalization to fix the scale of

[9] This is not to say that this is the only way to capture dependencies across nominal variables. Another way to capture dependency may be achieved by random-taste parameter. However, this would require the analyst to have a common exogenous variable in all the nominal variables and in all the alternatives. This could be rather difficult given the differential impact of the same exogenous variable on different choice dimensions. On the other hand, one is free to incorporate random-taste parameters at the nominal variable level (with full or no correlation) with no cross-correlation across nominal variables. It could be incorporated as follows: Let Σ_i be the ($i_G \times i_G$) covariance matrix of exogenous variables for the i^{th} nominal variable. Where $G = \sum_{r=1}^{K} i_r$ is the total number of exogenous variables in the i^{th} nominal variable. Then, stack the exogenous variables for all the nominal variables in a matrix of size ($I_K\text{T} \times TG$) and all the random-taste parameter matrices into a $\boldsymbol{\Sigma}$ matrix as follows:

$$X_{I_K T} = \begin{bmatrix} x'_{1,1} & 0 & 0 & 0 \\ x'_{2,1} & 0 & 0 & 0 \\ x'_{1_k,1} & 0 & 0 & 0 \\ 0 & x'_{I,1} & 0 & 0 \\ 0 & x'_{I,1} & 0 & 0 \\ 0 & x'_{I_k,1} & 0 & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & x'_{I-1_k,T-1} & 0 \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & x'_{I_k,T} \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_I \end{bmatrix}$$

With this, we may write the distribution of $\boldsymbol{U}$ as
$\boldsymbol{U} \sim N_{I_K \times I_K}\left[sumc[(\tilde{\beta}.*x)'], [1_T.*.\Lambda + X_{1_K T} * \left(1_T.*.\Sigma * X'_{I_K T}\right)]\right]$.

utility. Therefore, we normalize the top diagonal element to 1 for estimation purposes (Keane 1992). However, all the differenced error matrices must originate from the same un-differenced error matrix. To do so, append the matrices Λ_i by adding a row and column of zeros on the top (Sidharthan & Bhat 2012) i.e., $\Lambda_i = \begin{bmatrix} 0 & 0_{1,i_K-1} \\ 0_{i_K-1,1} & \Lambda_i \end{bmatrix}$ or multiply the matrix **Λ** with a matrix **D** (i.e., expanded differenced matrix **DΛ** for all the nominal variables) constructed as follows:

Define a matrix **D** of size$[(I_K) \times (I_K - I)]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **D**.

```
for m=1 to I
        if(m==1)
                st_row =2
                end_row=m_K
                st_col =1
                end_col= m_K − 1
        else
                st_row =[∑_{n=1}^{m−1} n_K] + 2
                end_row=[∑_{n=1}^{m} n_K]
                st_col =[∑_{n=1}^{m−1}(n_K − 1)] +1
                end_col= [∑_{n=1}^{m}(n_K − 1)]
        end
        D[st_row: end_row, st_col:end_col]= 1_{m_K−1}
end
```

Now, similar to the continuous variable model, we introduce the AR-1 structure in the unobserved part of the utility as follows:

$$\varepsilon_{i_k t} = \lambda_i \varepsilon_{i_k t-1} + \eta_{i_k}$$

where λ_i is the autoregressive coefficient for the i^{th} nominal variable and $\eta_{i_k}$is the time-independent component of the error-term. That is, η'_{i_k} can be correlated for a nominal variable in a given time period, but are independent across time-periods. With this, we may re-write the equation (5) as follows with all the notations as above:

(7) $U_{i_k t} = \boldsymbol{\beta}'_i x_{i_k t} + \varepsilon_{i_k t}$

Now, stack the time-independent error terms and the nominal variable specific AR coefficients as follows:

$\boldsymbol{\eta_i} = \left(\eta_{i_1}, \eta_{i_2}, \dots, \eta_{i_K}\right)[(\mathrm{i}_k \times 1)]$ vector, $\boldsymbol{\eta_t} = (\eta_{1_1}, \eta_{1_2}, \dots, \eta_{1_k}, \dots, \eta_{1_K})[(\mathrm{I}_K \times 1)]$ vector,

$\boldsymbol{\eta} \;\; = [\mathrm{ones}(T, 1).*.\boldsymbol{\eta_t}][(TI_K \times 1)]$ vector, and $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_I)[(I \times 1)]$ vector.

With this, we assume that Λ_i is the covariance matrix of $\boldsymbol{\eta_i}$[10]. Now, define the additional matrices in order to write equation (7) in the matrix form:

Define a matrix **R** of size $[(TI_K) \times (TI_K)]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **R**.

```
for m=2 to T
    for n=1 to I
        if(n==1)
            for j =1 to n_K
                row = (m-1) * I_K + j
                col = (m-2) * I_K + j
                R[row,col]=1
```

[10] Here we use the same notation for the covariance matrix of $\boldsymbol{\eta_i}$ as $\boldsymbol{\varepsilon_i}$ to avoid redundancy. To be precise, one can motivate the model directly by incorporating AR-1 structure, avoiding the need for redundancy.

```
                end
            else
                for j =1 to n_K
                    row = (m-1) * I_K + [∑_{r=1}^{n-1} r_K] + j
                    col = (m-2) * I_K + [∑_{r=1}^{n-1} r_K] + j
                    R[row,col]=1
                end
            end
        end
    end
```

Next, construct a matrix $\boldsymbol{F_{I_K T}}$ of size $(TI_K \times TI_K)$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of matrix $\boldsymbol{F_{I_K T}}$.

```
    for m=2 to T
        for n=1 to I
            if(n==1)
                for j =1 to n_K
                    row = (m-1) * I_K + j
                    col = (m-2) * I_K + j
                    F_{I_K T}[row,col]=λ[n.1]
                end
            else
                for j =1 to n_K
                    row = (m-1) * I_K + [∑_{r=1}^{n-1} r_K] + j
                    col = (m-2) * I_K + [∑_{r=1}^{n-1} r_K] + j
```

F_{I_KT}**[row,col]=** $\boldsymbol{\lambda}[\boldsymbol{n}.\mathbf{1}]$

end

end

end

end

With this, equation (7) can be written in the matrix form as follows:

(8) $\boldsymbol{U} = \text{sumc}\left[(\tilde{\beta}.* x)'\right] + \boldsymbol{C_\eta}$

where $\mathbf{C} = [\mathbf{1}_{I_KT}$ - $(\mathbf{F}_{\boldsymbol{I_KT}}.*|.\ \mathbf{R}_{I_KT}\ )]^{-1}$ of size $(TI_K \times TI_K)$.

From equation (8), it is easy to observe that U is distributed normally with mean $sumc\left[(\tilde{\beta}.* x')\right]$ and covariance $\mathbf{C} * [\mathbf{1}_{\boldsymbol{T}}.*.\boldsymbol{D\Lambda}] * \mathbf{C}'$. Also, to maintain the bound on autoregressive parameter vector $\boldsymbol{\lambda}$, we parametrize the parameter as $\boldsymbol{\lambda} = \boldsymbol{\lambda_\rho}/[\mathbf{1} + (\boldsymbol{\lambda_\rho})^{\mathbf{2}}]^{\mathbf{0.5}}$, where $\boldsymbol{\lambda_\rho}$ is the value passed to the optimization module.

Joint Model Estimation

Now, we bring the individual components of the model together to form a joint model followed by model estimation approach. To write the joint model in a matrix form, define the following vector and matrices:

$Y_tU_t = \left(\tilde{y}_{1,t}, \tilde{y}_{2,t}, \dots, \tilde{y}_{\tilde{H},t}, \ddot{y}^*_{1,t}, \ddot{y}^*_{2,t}, \dots, \ddot{y}^*_{\ddot{N},t}, U_t\right)[(\tilde{H} + \ddot{N} + I_K) \times 1]$ vector, 1

$\boldsymbol{YU} = [(\boldsymbol{Y_1U_1}), (\boldsymbol{Y_2U_2}), \dots, (\boldsymbol{Y_TU_T})]'\left[T * \left(\tilde{H} + \ddot{N} + I_K\right) \times 1\right]$ vector, [11]

[11] The assumption here is that $\tilde{T} = \ddot{T} = T$. However, this need not be the case. If $\tilde{T} \neq \ddot{T} \neq T$, we assume that $T \geq \tilde{T}\ \&\ T \geq \ddot{T}$ given the focus of discrete choice models to model the choice outcome. Later we provide a design matrix which can be multiplied with the vector ***YU*** to extract the relevant components. In the meantime, all the missing values can be replaced by zero. Thus, from now on we assume $T \geq \tilde{T}\ \&\ T \geq \ddot{T}$ and thus all the matrices/vector will be created to accommodate the highest dimension *T*.

$X_t = \left(x'_{1,t}, x'_{2,t}, \dots, x'_{\widetilde{H},t}, x'_{1,t}, x'_{2,t}, \dots, x'_{\ddot{N},t}, x'_t\right)\left[\left(\widetilde{H} + \ddot{N} + I_K\right) \times \max\left(k_{\widetilde{h}}, k_{\ddot{n},} \mathrm{g}_i\right)\right]$

matrix,

$\boldsymbol{X} = \left(\boldsymbol{X_1}\boldsymbol{X_2}, \dots, \boldsymbol{X_T}\right)\left[T * \left(\widetilde{H} + \ddot{N} + I_K\right) \times \max\ \left(k_{\widetilde{h}}, k_{\ddot{n},} \mathrm{g}_i\right)\right]$ matrix,

$\overleftrightarrow{\boldsymbol{\beta}} = \left(\boldsymbol{\gamma}', \boldsymbol{\delta}', \boldsymbol{\beta}'\right)\left[\left(\widetilde{H} + \ddot{N} + I_K\right) \times \max\ \left(k_{\widetilde{h}}, k_{\ddot{n},} \mathrm{g}_i\right)\right]$ matrix,

$\overleftrightarrow{\beta} = ones(T, 1).*.\overleftrightarrow{\beta}\left[T * \left(\widetilde{H} + \ddot{N} + I_K\right) \times \max\ \left(k_{\widetilde{h}}, k_{\ddot{n},} \mathrm{g}_i\right)\right]$ matrix.

Define a matrix **D_Mat** of size $\left[\left(\widetilde{H} + \ddot{N} + I_K\right) \times \left(\widetilde{H} + \ddot{N} + I_K - I\right)\right]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **D_Mat**.

D_Mat $[1{:}\ \widetilde{H} + \ddot{N}, 1{:}\ \widetilde{H} + \ddot{N}] = \mathbf{1}_{\widetilde{H}+\ddot{N}}$

for m =1 to I

if(m==1)

st_row = $\widetilde{H} + \ddot{N} + 2$

end_row = $\widetilde{H} + \ddot{N} + m_K$

st_col = $\widetilde{H} + \ddot{N} + 1$

end_col = $\widetilde{H} + \ddot{N} + m_K - 1$

else

st_row = $\widetilde{H} + \ddot{N} + [\sum_{n=1}^{m-1} n_K] + 2$

end_row = $\widetilde{H} + \ddot{N} + [\sum_{n=1}^{m} n_K]$

st_col = $\widetilde{H} + \ddot{N} + [\sum_{n=1}^{m-1}(n_K - 1)] + 1$

end_col = $\widetilde{H} + \ddot{N} + [\sum_{n=1}^{m}(n_K - 1)]$

end

D_Mat[st_row: end_row, st_col: end_col]= $\mathbf{1}_{\boldsymbol{m_K-1}}$

end

Construct a matrix **Cap_RI** of size $[T(\widetilde{H} + \ddot{N} + I_K) \times T(\widetilde{H} + \ddot{N} + I_K)]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **Cap_RI**.

for *m* =2 to T

for *n* =1 to $\widetilde{H}$

row = $(m-1) * (\widetilde{H} + \ddot{N} + I_K) + n$

col = $(m-2) * (\widetilde{H} + \ddot{N} + I_K) + n$

Cap_RI[row,col]=1

end

end

for *m* =2 to T

for *n* =1 to I

if(n==1)

for $j = 1$ to n_K

row = $(m-1) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + j$

col = $(m-2) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + j$

Cap_RI[row,col]=1

end

else

for $j = 1$ to n_K

row = $(m-1) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + [\sum_{r=1}^{n-1} r_K] + j$

col = $(m-2) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + [\sum_{r=1}^{n-1} r_K] + j$

Cap_RI[row,col]=1

end

end

end

end

Finally, construct two matrices **I_Mean** and **I_Error** of size $[T(\widetilde{H}+\ddot{N}+I_K)\times T(\widetilde{H}+\ddot{N}+I_K)]$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **I_Mean** and **I_Error**.

for m =2 to T

for n =1 to $\widetilde{H}$

if(m==1)

row = $(m-1)*(\widetilde{H}+\ddot{N}+I_K)+n$

col = $(m-2)*(\widetilde{H}+\ddot{N}+I_K)+n$

I_Mean[row,col]= $\widetilde{\boldsymbol{\rho}}[\boldsymbol{i},\mathbf{1}]$

I_Error[row,col]= $\widetilde{\boldsymbol{\rho}}[\boldsymbol{i},\mathbf{1}]$

end

end

for m =2 to T

for n =1 to I

if(n==1)

for $j=1$ to n_K

row = $(m-1)*(\widetilde{H}+\ddot{N}+I_K)+(\widetilde{H}+\ddot{N})+\ j$

col = $(m-2)*(\widetilde{H}+\ddot{N}+I_K)+(\widetilde{H}+\ddot{N})+\ j$

I_Error[row,col]= $\boldsymbol{\lambda}[\boldsymbol{n},\mathbf{1}]$

end

else

for $j=1$ to n_K

row = $(m-1)*(\widetilde{H}+\ddot{N}+I_K)+(\widetilde{H}+\ddot{N})+[\sum_{r=1}^{n-1} r_K]+\ j$

col = $(m-2)*(\widetilde{H}+\ddot{N}+I_K)+(\widetilde{H}+\ddot{N})+[\sum_{r=1}^{n-1} r_K]+\ j$

I_Error[row,col]= $\boldsymbol{\lambda}[\boldsymbol{n},\mathbf{1}]$

end

end

end

end

Also, collect all the error-covariance matrices as follows:

$$\overleftrightarrow{\boldsymbol{\Sigma}}=\begin{bmatrix}\Xi & \text{Cov}(\boldsymbol{\Xi};\boldsymbol{\Gamma})' & \text{Cov}(\boldsymbol{\Xi};\boldsymbol{\Lambda})' \\ \text{Cov}(\boldsymbol{\Xi};\boldsymbol{\Gamma}) & \boldsymbol{\Gamma} & \text{Cov}(\boldsymbol{\Gamma};\boldsymbol{\Lambda})' \\ \text{Cov}(\boldsymbol{\Xi};\boldsymbol{\Lambda}) & \text{Cov}(\boldsymbol{\Gamma};\boldsymbol{\Lambda}) & \boldsymbol{\Lambda}\end{bmatrix}[(\widetilde{H}+\ddot{N}+I_K)\times(\widetilde{H}+\ddot{N}+I_K)]$$

where off-diagonal elements capture the dependence across different type of variables (continuous, ordered, and nominal variables).

With this, we can write the distribution of joint model as follows:

$\boldsymbol{YU}\sim MVN(B_{T*(\widetilde{H}+\ddot{N}+I_K)},\Theta_{T*(\widetilde{H}+\ddot{N}+I_K)\times T*(\widetilde{H}+\ddot{N}+I_K)})$,

where $\boldsymbol{B}=\textbf{F\_Mean}*\textbf{sumc}\left[\left(\overleftrightarrow{\boldsymbol{\beta}}.*\boldsymbol{X}\right)'\right]$,

$\boldsymbol{\Theta}=\textbf{F\_Error}*\left[\mathbf{1}_{\boldsymbol{T}}.*.\left(\boldsymbol{D}_{\boldsymbol{MAT}}*\overleftrightarrow{\boldsymbol{\Sigma}}\right)\right]*\textbf{F\_Error}'$,

$\textbf{F\_Mean}=[\mathbf{1}_{\boldsymbol{T}(\widetilde{\boldsymbol{H}}+\ddot{\boldsymbol{N}}+\boldsymbol{I_K})}-\textbf{I\_Mean}.*.\textbf{Cap\_RI})]^{-\mathbf{1}}$, and

$\textbf{F\_Error}=[\mathbf{1}_{\boldsymbol{T}(\widetilde{\boldsymbol{H}}+\ddot{\boldsymbol{N}}+\boldsymbol{I_K})}-\textbf{I\_Error}.*.\textbf{Cap\_RI})]^{-\mathbf{1}}$

Next, to estimate the model, we take the utility difference between the chosen alternative (i_{m_k}) and non-chosen alternatives for all the nominal variables. To perform utility difference, construct a matrix **M_mat** of size $[T(\widetilde{H}+\ddot{N}+I_K-I)\times T(\widetilde{H}+\ddot{N}+I_K)]$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **M_mat**.

For m=1 to T

M=zeros$\left((\widetilde{H}+\ddot{N}+I_K-I\right), (\widetilde{H}+\ddot{N}+I_K))$

$\mathbf{M}\left[\mathbf{1}:\widetilde{H}+\ddot{N}\right), 1:\left(\widetilde{H}+\ddot{N}\right)] = \mathbf{1}_{(\widetilde{H}+\ddot{N})}$

for n=1 to I

Iden_mat=$\mathbf{1}_{n_K-1}$

O_neg =-1*ones(n_K-1,1)

if($n_{m_K} == 1$)

temp_mat =O_neg ~ Iden_mat

else if($n_{m_K} == n_K$)

temp_mat=Iden_mat[.,1: $n_{m_K}1]$~O_neg~Iden_mat[., $n_{m_K}: n_K - 1]$

end

if(n==1)

row1 = $\left(\widetilde{H}+\ddot{N}\right)+1$

row2 = $\left(\widetilde{H}+\ddot{N}\right)+n_K-1$

col1 = $\left(\widetilde{H}+\ddot{N}\right)+1$

col2 = $\left(\widetilde{H}+\ddot{N}\right)+n_K$

else

row1 = $\left(\widetilde{H}+\ddot{N}\right)+(\sum_{j=1}^{n-1}(j_K-1))+1$

row2 = $\left(\widetilde{H}+\ddot{N}\right)+(\sum_{j=1}^{n}(j_K-1))+1$

col1 = $\left(\widetilde{H}+\ddot{N}\right)+(\sum_{j=1}^{n-1}(j_K))+1$

col2 = $\left(\widetilde{H} + \ddot{N}\right) + (\sum_{j=1}^{n-1}(j_K))+n_K$

end

M[row1:row2,col1:col2]=temp_mat

end

s_row1 = $(m-1) * \left(\widetilde{H} + \ddot{N} + I_K - I\right) + 1$

s_row2 = $(m) * \left(\widetilde{H} + \ddot{N} + I_K - I\right)$

s_col1 = $(m-1) * \left(\widetilde{H} + \ddot{N} + I_K\right) + 1$

s_col2 = $(m) * \left(\widetilde{H} + \ddot{N} + I_K\right)$

M_mat[s_row1:s_row2,s_col1:s_col2]=**M**

end

where "~" refers to horizontal concatenation.

With this we may write the distribution of $\bar{\boldsymbol{Y}}\bar{\boldsymbol{U}}$ (same as $\boldsymbol{YU}$ but with utility difference w.r.t the chosen alternative for all the nominal variables) as $\bar{Y}\bar{U} \sim MVN_{T*(\widetilde{H}+\ddot{N}+I_K-I)}\left(\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\Theta}}\right)$ where $\widetilde{\boldsymbol{B}}$ =**M_mat***$\boldsymbol{B}$**, and** $\widetilde{\boldsymbol{\Theta}} = \mathbf{M\_mat} * \widetilde{\boldsymbol{\Theta}} * \mathbf{M\_mat}'$.

Next, we define a matrix to re-arrange the elements of mean and covariance matrix of $\bar{\boldsymbol{Y}}\bar{\boldsymbol{U}}$ in the following order: continuous, ordered, and nominal. This makes it easy to find the conditional distribution of non-continuous variables in a matrix format. To do so, define a matrix **R_mat** of size $[T\left(\widetilde{H} + \ddot{N} + I_K - I\right) \times T\left(\widetilde{H} + \ddot{N} + I_K - I\right)]$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **R_mat.**

---------------For continuous variables---

For *m* =1 to *T*

row1=(*m*-1) * $\widetilde{H}$+1

row2=(*m*) *$\widetilde{H}$

col1=(*m*-1) * $\left(\widetilde{H} + \ddot{N} + I_K - I\right) + 1$

col2=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\widetilde{H}$

R_mat[row1: row2, col1:col2]= $1_{\widetilde{H}}$

end

-----------------For ordered variables---

For m =1 to T

row1=$\widetilde{H}T$+ (m-1) * $\ddot{N}$+1

row2=$\widetilde{H}T$+ (m) * $\ddot{N}$+1

col1=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\widetilde{H}+1$

col2=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\widetilde{H}+\ddot{N}$

R_mat[row1: row2, col1:col2]= $\mathbf{1}_{\ddot{N}}$

end

-----------------------For nominal variables--

For m =1 to T

row1=$(\widetilde{H}+\ddot{N})T$ + (m-1) * (I_K-I) +1

row2=$(\widetilde{H}+\ddot{N})T$ + (m) * (I_K-I) +1

col1=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\left(\widetilde{H}+\ddot{N}\right)+1$

col2=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\left(\widetilde{H}+\ddot{N}\right)+(I_K-I)$

R_mat[row1: row2, col1:col2]= $1_{(I_K-I)}$

end

With this, we may write:

$\overline{\boldsymbol{Y}}\overline{\boldsymbol{U}}\sim MVN_{T*(\widetilde{H}+\ddot{N}+I_K-I)}\left(\ddot{\boldsymbol{B}},\ddot{\boldsymbol{\Theta}}\right)$

where $\widetilde{\boldsymbol{B}}$ =**R_mat***$\widetilde{\boldsymbol{B}}$, and $\overleftrightarrow{\boldsymbol{\Theta}}=\mathbf{R\_mat}*\widetilde{\boldsymbol{\Theta}}*\mathbf{R\_mat}'$.

Next, to account for un-balanced panel data structure, we define a matrix **RM_mat** of size $[\tilde{T}\tilde{H} + \ddot{T}\ddot{N} + T(I_K - I) \times T(\tilde{H} + \ddot{N} + I_K - I)]$ with all the cells filled with zeros. It will allow us to collect the relevant elements from the vector $\overleftrightarrow{B}$ and matrix $\overleftrightarrow{\Theta}$. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **RM_mat**.

---------------For continuous variables---

For m =1 to $\tilde{T}$

row1=(m-1) * $\tilde{H}$+1

row2=(m) *$\tilde{H}$

col1=(m-1) * $\tilde{H} + 1$

col2=(m) * $\tilde{H}$

R_mat[row1: row2, col1:col2]= $\mathbf{1}_{\tilde{H}}$

end

-----------------For ordered variables---

For m =1 to $\ddot{T}$

row1=$\tilde{H}\tilde{T}$+ (m-1) * $\ddot{N}$+1

row2=$\tilde{H}\tilde{T}$+ (m) * $\ddot{N}$

col1=$\tilde{H}T$ + (m-1) * $\ddot{N} + 1$

col2=$\tilde{H}T$ + (m) * $\ddot{N}$

R_mat[row1: row2, col1:col2]= $1_{\ddot{N}}$

end

-----------------------For nominal variables--

For m =1 to T

row1=$\tilde{H}\tilde{T} + \ddot{N}\ddot{T}$ + (m-1) * $(I_K - I)$ +1

row2=$\tilde{H}\tilde{T} + \ddot{N}\ddot{T}$ + (m) * $(I_K - I)$ +1 col1= $\left(\widetilde{H}+\ddot{N}\right)T+(m-1)*(I_K-I)\right)+1$

col2=$\left(\widetilde{H}+\ddot{N}\right)T+(m)*(I_K-I)\right)$

R_mat[row1: row2, col1:col2]= $\mathbf{1}_{(I_K-I)}$

end

Now we may write:

$\overline{\boldsymbol{Y}}\overline{\boldsymbol{U}} \sim MVN_{\widetilde{H}\widetilde{T}+\ddot{N}\ddot{T}+T(I_K-I)}(\overleftrightarrow{\boldsymbol{B}},\overleftrightarrow{\boldsymbol{\Theta}})$

where $\overleftrightarrow{\boldsymbol{B}}=$ **RM_mat*** $\overleftrightarrow{\boldsymbol{B}}$, and $\overleftrightarrow{\boldsymbol{\Theta}}=\textbf{RM\_mat}*\overleftrightarrow{\boldsymbol{\Theta}}*\textbf{RM\_mat}'$.

Next, partition the $\overleftrightarrow{B}$ and $\overleftrightarrow{\boldsymbol{\Theta}}$ into the continuous and non-continuous variables as follows:

$$\overleftrightarrow{B}=\begin{bmatrix}\overleftrightarrow{B}_{\widetilde{H}}\\ \overleftrightarrow{B}_{\ddot{N}\overline{U}}\end{bmatrix}\begin{bmatrix}\widetilde{H}\widetilde{T}\times 1\\ \ddot{N}\ddot{T}+T(I_K-I)\times 1\end{bmatrix}, \text{and } \overleftrightarrow{\boldsymbol{\Theta}}=\begin{bmatrix}\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H}} & \overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H},\ddot{N}\overline{U}}\\ \overleftrightarrow{\boldsymbol{\Theta}}'_{\widetilde{H},\ddot{N}\overline{U}} & \overleftrightarrow{\boldsymbol{\Theta}}_{\ddot{N}\overline{U}}\end{bmatrix}.$$

With this, the conditional distribution of non-continuous variables can be written as:

$$\overline{\boldsymbol{B}}_{\ddot{N}\overline{U}}=\overleftrightarrow{\boldsymbol{B}}_{\ddot{N}\overline{U}}+\overleftrightarrow{\boldsymbol{\Theta}}'_{\widetilde{H},\ddot{N}\overline{U}}(\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H}})^{-1}(\tilde{\boldsymbol{y}}\,[1{:}\,\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{T}}]-\overleftrightarrow{\boldsymbol{B}}_{\widetilde{H}}),$$

$$\overline{\boldsymbol{\Theta}}_{\ddot{N}\overline{U}}=\overleftrightarrow{\boldsymbol{\Theta}}_{\ddot{N}\overline{U}}-\overleftrightarrow{\boldsymbol{\Theta}}'_{\widetilde{H},\ddot{N}\overline{U}}(\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H}})^{-1}\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H},\ddot{N}\overline{U}}.$$

Also, append the threshold vectors as follows:

$\overline{\Psi}_{low}$=$[(\Psi_{low}[1{:}\,\ddot{N}\ddot{T}])',(-\infty_{T(I_K-I)})']\ \left[\left(\ddot{N}\ddot{T}+T(I_K-I)\right)\times 1\right]$ vector, and,

$\overline{\Psi}_{up}$=$[(\Psi_{up}[1{:}\,\ddot{N}\ddot{T}])',(0_{T(I_K-I)})']'\ \left[\left(\ddot{N}\ddot{T}+T(I_K-I)\right)\times 1\right]$ vector. Where $-\infty_{T(I_K-I)})$ and $0_{T(I_K-I)}$ are column vectors of size $T(I_K-I)$with all the cells filled with a value of “-∞” and “0” respectively.

Then the likelihood function may be written as:

$$L(\theta)=f_{\widetilde{H}\widetilde{T}}(\tilde{y}\,[1{:}\,\widetilde{H}\widetilde{T}]\overleftrightarrow{B}_{\widetilde{H}},\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H}})\times\int_{\overline{\Psi}_{low}}^{\overline{\Psi}_{up}}f_{\ddot{N}\ddot{T}+T(I_K-I)}(\boldsymbol{r}|\,\overline{\boldsymbol{B}}_{\ddot{N}\overline{U}},\overleftrightarrow{\boldsymbol{\Theta}}_{\ddot{N}\overline{U}})dr \qquad (9)$$

where $\boldsymbol{\theta} = [\boldsymbol{\gamma}', \boldsymbol{\delta}', \boldsymbol{\beta}', \boldsymbol{\rho}', \boldsymbol{\lambda}', (\boldsymbol{Vech}(\overleftrightarrow{\boldsymbol{\Sigma}}))']$ and Vech (.) operator vectorizes the unique element of a matrix.

The likelihood function involves computation of a $\widetilde{H}\widetilde{T}$ dimensional multi-variate normal probability density (MVNPD) function and $\ddot{N}\dddot{T} + T(I_K - I)$ dimensional multi-variate normal cumulative density (MVNCD) function. While the MVNPD function has a closed form expression, increase in dimensionality can lead to calculation of numerical value very close to zero and thus causing issues during estimation[12]. On the other hand, the computation of a MVNCD function is a well-known challenge in the literature (Genz 1992; Heiss 2010; Connors et al. 2014). Even the powerful GHK simulator armed with sophisticated quasi-random sequences can calculate the value accurately only up to a limited number of dimensions (Sándor & András 2004)[13]. At the same time, it is well known and established that any simulation-based method loses its accuracy with increases in dimension due to simulation noise, not to mention the unreasonable computation time (Train 2000; Bhat 2003; Craig 2008). For example: the analysis section of the paper has 8 continuous variables with 20 time periods, 5 ordinal variables with 1 time period, and 1 nominal variable with 8 alternatives and 20 choice occasions. In the maximum likelihood (ML) approach, this translates to a computation of a 160 dimensional MVNPD function and a 145 dimensional MVNCD function. Therefore it may be quite challenging to solve equation (9) using ML approach.

While one can use Bayesian approach to solve such a complicated likelihood function involving a series of draws from conditional distribution, a review of literature involving Probit kernel shows that Bayesian approach has not performed as expected in terms of recovering

---

[12] Consider a situation where there are 20 continuous dependent variables. Now, estimate a uni-variate regression for each of the 20 continuous variables which may include parameters apart from a constant. Now, if one wish to estimate a joint model for all the 20 continuous variables, even with a good starting value (obtained from uni-variate regression), the MVNPD value may be very close to zero (numerically).

[13] The assumption is that the number of draws are finite (less than 1000) to maintain reasonable estimation time.

parameters and their standard errors (Franzese et al. 2010; Patil et al. 2017). On the other hand, there have been few studies (Daziano 2015; Zhou et al. 2016) which have found the performance of Bayesian approach to be quite good. However, these studies did not compare the performance of Bayesian approach against ML or other approaches. This is not to say that the Bayesian approach may not work. A comprehensive evaluation of the present model using Bayesian approach is outside the scope of the paper and we leave this for future explorations. Therefore, we use a composite marginal likelihood (CML) approach which has been established in the last decade as one of the powerful approach for solving likelihood functions with high dimensional integrals. A comprehensive discussion on the CML approach is outside the scope of this paper and readers are refer to Varin & Vidoni 2005; Varin 2008; Varin, Reid & Firth 2011 for a detailed discussion on CML and see Bhat & Dubey (2014) for its application in the context of discrete choice models. Further Bhat and colleagues have performed extensive simulation using CML approach for complex econometric models and have observed highly accurate results (Paleti & Bhat 2013; Bhat & Dubey 2014; Bhat 2015; Bhat et al. 2016).

Composite Marginal Likelihood Approach

The likelihood function can be written as follows using the CML approach:

$$L_{CML}(\theta) = \left(\prod_{h=1}^{\tilde{H}\tilde{T}-1} \prod_{h'=h+1}^{\tilde{H}\tilde{T}} f_2(\tilde{y}_{hh'} | \overleftrightarrow{B}_{hh'}, \overleftrightarrow{\Theta}_{hh'}) \right) \times$$

$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}-1} \prod_{r'=r+1}^{\ddot{N}\ddot{T}} \Pr(\ddot{y}_n = a_n, \ddot{y}_{n'} = a_{n'}) \right) \times$$

$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}} \prod_{t=1}^{T} \prod_{r'=1}^{I} \Pr\left(\ddot{y}_n = a_n, i_{r' t} = i_{r'_{m^{k'}} t} \right) \right) \times$$

$$\left(\prod_{r=1}^{TI-1} \prod_{r'=r+1}^{TI} \Pr(i_r = i_{r_{m_k}}, i_{r'} = i_{r'_{m_k}}) \right) \qquad (10)$$

In the above CML expression, the first expression corresponds to the pairing of two continuous variables at a time reducing the dimension of MVNPD function from $\tilde{H}\tilde{T}$ to a maximum of 2 avoiding any numerical issues in calculation of MVNPD function due to high dimensionality[14]. The second expression corresponds to the pairing of two ordinal variables reducing the dimensionality of integration to 2 from $\ddot{N}\ddot{T}$. The third expression corresponds to the pairing between an ordinal and nominal variable with a maximum dimensionality of integration equal to max($i_K \forall I$). Finally the fourth expression corresponds to the pairing between nominal variables with a highest dimensionality of integration being equal to 2*max($i_K \forall I$).

To explicitly write out the equation (10) in terms of MVNPD and MVNCD functions, we define a set of selection matrices: (1) construct a selection matrix **D_HH** of size [2× $\tilde{H}\tilde{T}$] with all the cells filled with zeros. Now, place a value of '1' in 1st row and h^{th} column and in 2nd row and $h'^{\,th}$ column. This matrix is designed to collect relevant elements for pairing between continuous variables within and across time-periods, (2) define a selection matrix **D_NI** of size $[i_K \times (\ddot{N}\ddot{T} + (I_K - I)T]$ with all the cells filled with zeros. This matrix is designed to collect relevant elements for pairing between ordered and nominal variables. Now, place a value of '1' in the 1st row and r^{th} column. Next if $r' = 1$, then place an identity matrix of size $r'_K - 1$ in the rows 2 to r'_K and columns $\ddot{N}\ddot{T} + (t-1)(I_K - I) + 1$ to $\ddot{N}\ddot{T} + (t-1)(I_K - I) + r'_K - 1$, otherwise place an identity matrix of size $r'_K - 1$ in the rows 2 to r'_K and columns $\ddot{N}\ddot{T} + (t-1)(I_K - I) + \left(\sum_{j=1}^{r'-1}(j_K - 1)\right)$ +1 to $\ddot{N}\ddot{T} + (t-1)(I_K - I) + \left(\sum_{j=1}^{r'-1}(j_K - 1)\right)$, and (3) define two variables as follows: alt_1 = $r - (\text{ceil}(r/1) - 1) * I$ and alt_2 = $r' - (\text{ceil}(r/I) - 1) * I$. Where **ceil(.)** operator rounds the value in parenthesis to next

---

[14] For all the pairings, different continuous variables in the same time-period and all continuous variables across time-periods are used. This also holds for all pairing between ordinal and ordinal, and nominal and nominal variables.

largest integer. Now, construct a selection matrix **D_II** of size $[(r_{alt_1} + r'_{alt_2} - 2) \times (\ddot{\ddot{N}}\ddot{\ddot{T}} + (I_K - I))\mathrm{T}]$ with all the cells filled with zeros. This matrix is designed to collect relevant elements for pairing between nominal variables within and across time-periods. Now, use the pseudocode provided below to fill-up the cells of **D_II** matrix.

if (alt_1 == 1)

row1=1

row2=$r_{alt_1} - 1$

col1 =$\ddot{\ddot{N}}\ddot{\ddot{T}} + (\text{ceil}(r/1) - 1) * (I_K - I) + I$

col2 =$\ddot{\ddot{N}}\ddot{\ddot{T}} + (\text{ceil}(r/1) - 1) * (I_K - I) + r_{alt_1} - 1$

else

row1=1

row2=$r_{alt_1} - 1$

col1 =$\ddot{\ddot{N}}\ddot{\ddot{T}} + (\text{ceil}(r/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt1-1}(j_K - 1)\right) + 1$

col2 =$\ddot{\ddot{N}}\ddot{\ddot{T}} + (\text{ceil}(r/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt1}(j_K - 1)\right)$

end

D_II[row1:row2,col1:col2]=$1_{i_{alt\_1}-1}$

if(alt_2==1)

row1=r_{alt_1}

row2=$r_{alt\_1} + r'_{alt\_2} - 2$

col1 =$\ddot{\ddot{N}}\ddot{\ddot{T}} + (\text{ceil}(r'/1) - 1) * (I_K - I) + 1$

col2 =$\ddot{\ddot{N}}\ddot{\ddot{T}} + (\text{ceil}(r'/1) - 1) * (I_K - I) + r'_{alt\_2} - 1$

else

row1=r_{alt_1}

row2=$r_{alt_1} + r'_{alt_2} - 2$

col1 = $\ddot{N}\ddot{T} + (\text{ceil}(r'/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt\_2-1}(j_K - 1)\right) + 1$

col2 = $\ddot{N}\ddot{T} + (\text{ceil}(r'/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt\_2}(j_K - 1)\right)$

end

D_II[row1:row2,col1:col2] $1_{i_{alt\_2}-1}$

With the selection matrices defined, now we define the appropriate mean vector and covariance matrix for pairing of dependent variables. Define the following vectors and matrices:

$$\widehat{\boldsymbol{B}}_{\boldsymbol{hh'}} = \mathbf{D\_HH} * \overleftrightarrow{\boldsymbol{B}}_{\boldsymbol{hh'}}, \widehat{\boldsymbol{\Theta}}_{\boldsymbol{hh'}} = \mathbf{D\_HH} * \overleftrightarrow{\boldsymbol{\Theta}}_{\boldsymbol{hh'}},* \mathbf{D\_HH}',\ \widehat{\boldsymbol{y}}_{\boldsymbol{hh'}} = \mathbf{D\_HH} * \widetilde{\boldsymbol{y}}_{\boldsymbol{hh'}},$$

$$v_{r,low} = \frac{[\Psi_{low}]_r - [\bar{\mathrm{B}}_{\ddot{N}\bar{U}}]_r}{\sqrt{[\bar{\Theta}_{\ddot{N}\bar{U}}]_{rr}}}, v_{r,up} = \frac{[\Psi_{up}]_r - [\bar{\mathrm{B}}_{\ddot{N}\bar{U}}]_r}{\sqrt{[\bar{\Theta}_{\ddot{N}\bar{U}}]_{rr}}}, v_{r',low} = \frac{[\Psi_{low}]_{r'} - [\bar{\mathrm{B}}_{\ddot{N}\bar{U}}]_{r'}}{\sqrt{[\bar{\Theta}_{\ddot{N}\bar{U}}]_{r'r'}}},$$

$$v_{r',up} = \frac{[\Psi_{up}]_{r'} - [\bar{\mathrm{B}}_{\ddot{N}\bar{U}}]_{r'}}{\sqrt{[\bar{\Theta}_{\ddot{N}\bar{U}}]_{r'r'}}}, \vartheta_{rr} = \frac{[\bar{\Theta}_{\ddot{N}\bar{U}}]_{rr'}}{\sqrt{[\bar{\Theta}_{\ddot{N}\bar{U}}]_{rr}}\sqrt{[\bar{\Theta}_{\ddot{N}\bar{U}}]_{r'r'}}}, \widehat{B}_{rr'} = \mathrm{D\_NI} * \bar{B}_{\ddot{N}\bar{U}},$$

$$\widehat{\Theta}_{rr'} = \mathrm{D\_NI} * \bar{\Theta}_{\ddot{N}\bar{U}},* \mathrm{D\_NI}', \widehat{\Psi}_{rr',low} = \mathrm{D\_NI} * \Psi_{low},$$

$$\widehat{\Psi}_{rr',low}[2{:}\,rows(\widehat{\Psi}_{rr',low})] = \text{zeros}(\text{rows}(\widehat{\Psi}_{rr',low}), 1),$$

$$\widehat{\Psi}_{rr',up} = \mathbf{D\_NI} * \Psi_{up}, \breve{B}_{rr'} = \mathbf{D\_II} * \bar{B}_{\ddot{N}\bar{U}}, \text{and } \breve{\Theta}_{rr'} = \mathbf{D\_II} * \bar{\Theta}_{\ddot{N}\bar{U}},* \mathbf{D\_II}'$$

With the help of above defined notations, we may write the equation (10) in the explicit form as follows:

$$L_{CML}(\theta) = \left(\prod_{h=1}^{\widetilde{H}\widetilde{T}-1} \prod_{h'=h+1}^{\widetilde{H}\widetilde{T}} \phi_2\,(\hat{y}_{hh'}|\widehat{B}_{hh'}, \widehat{\Theta}_{hh'})\right) \times$$

$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}-1} \prod_{r'=r+1}^{\ddot{N}\ddot{T}} \begin{bmatrix} \Phi_2(v_{r,up}, v_{r',up}, \vartheta_{rr'}) - \Phi_2(v_{r,up}, v_{r',low}, \vartheta_{rr'}) \\ -\Phi_2(v_{r,low}, v_{r',up}, \vartheta_{rr'}) + \Phi_2(v_{r,low}, v_{r',low}, \vartheta_{rr'}) \end{bmatrix} \right) \times$$

$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}} \prod_{t=1}^{T} \prod_{r'=1}^{I} [\Phi_{r_K'}[(\hat{\psi}_{rr',up} - \widehat{B}_{rr'}); \widehat{\Theta}_{rr'}] - \Phi_{r_K'}\,[(\hat{\psi}_{rr',low} - \widehat{B}_{rr'}); \widehat{\Theta}_{rr'}]]\right) \times$$

$$\left(\prod_{r=1}^{TI-1} \prod_{r'=r+1}^{TI} [\Phi_{r_K + r_K' - 2}[\breve{B}_{rr'}; \breve{\Theta}_{rr'}]]\right) \qquad (11)$$

where $\phi_r(.)$ and $\Phi_r(.)$ represents a MVNPD and MVNCD function of dimension r, respectively. The parameters Θ are obtained by maximizing the $\log[L_{CML}(\theta)]$. Further, unlike the ML approach, in the CML approach, the equivalence between the inverse of Hessian matrix $H(\theta)\left[-\frac{\partial^2 L_{CML}(\theta)}{\partial\theta*\partial\theta'}\right]^{-1}$ and the information matrix $I(\theta)$ $\left[\left(\frac{\partial L_{CML}(\theta)}{\partial\theta}\right)\times\left(\frac{\partial L_{CML}(\theta)}{\partial\theta}\right)'\right]$does not exist and therefore the standard errors are calculated using the inverse of sandwich matrix $G(\theta)^{-1} = H(\theta)^{-1}I\,(\theta)\,\mathrm{H}(\theta)^{-1}$. Now that the dimension of MVNCD function has been reduced to a computationally acceptable range, one may use the Geweke- Hajivassiliou-Keane (GHK) simulator (Hajivassiliou et al. 1996) with quasi-random sequences or Bhat's Maximum Approximate Composite Marginal Likelihood (MACML) approach (Bhat 2011). While the GHK simulator is a simulation based estimator, the MACML is an analytic approximation and thus is computationally faster than the GHK. However, based on extensive testing of both methods, we have found that the MACML approach is a good method up to a dimension of 8-10. Its performance starts to degrade rather quickly beyond a dimension of 12 in comparison with the GHK simulator[15]. In our empirical analysis, the highest dimension of integral is 14 and thus based on equation (11) we use GHK simulator with 200 Halton-draws for the estimation of MVNCD function[16]. Finally, since the standard errors are calculated using sandwich estimator, one will need to calculate the Hessian matrix numerically or analytically. However, unlike logit kernel where the Hessian matrix can be computed numerically using central difference method, the same is not true for the Probit kernel due to relatively slow rate of change of MVNCD function in comparison to the exponential function[17]. To avoid any such

---

[15] The simulation design corresponds to a multinomial Probit model estimation for both cross-section and panel data structure with 5 alternatives, 5 choice occasions, and 5 random parameters with full and no cross-correlation.

[16] In our simulation experiments, we found that the 200 Halton draws are sufficient up to 20 dimensions.

[17] Most software (except "R" software) fails to calculate the Hessian matrix for the models built on Probit kernel. The "R" software uses Richardson extrapolation method for calculating the Hessian matrix which ensures the computation of a positive definite Hessian matrix, but its accuracy is low in most of the cases.

issues, we analytically calculated the first and second order derivatives of the CML function involving MVNCD function.

Positive Definiteness of Covariance Matrices

To maintain the positive definiteness of the error covariance and random taste parameter covariance matrices, we work with the Cholesky decomposition of the matrices during estimation. i.e., if we are working with the full joint model, we pass the lower triangular Cholesky decomposition of the matrix $\overleftrightarrow{\boldsymbol{\Sigma}}$. Also, since the error covariance matrix for ordered variables are restricted to be a correlation matrix along with the first row of each of the nominal variables, we need to ensure that the during estimation, proper restrictions are maintained. Therefore, for all the rows of the matrix $\overleftrightarrow{\boldsymbol{\Sigma}}$ where the diagonal element is constrained to be 1, parametrize such rows of the lower triangular Cholesky decomposition of matrix $\overleftrightarrow{\boldsymbol{\Sigma}}$ as follows:

Let $\boldsymbol{LL}' = \ddot{\boldsymbol{\Sigma}}$**,** where $\boldsymbol{L}$ is the lower triangular Cholesky matrix. Then, for each of the required rows do the following: Let $a_i = [1 + L[i, 1{:}\, i-1].\,{}^\wedge 2]^2$ where i refers to the row number and the operator ".^" refers to element by element exponentiation. Then parametrize all non-diagonal elements of the i^{th} row as $L[i,r] = \frac{L[i,r]}{a_i} \forall r = 1\ to\ i - 1$ and the diagonal element as $L[i,i] = \frac{1}{a_i}$.

The same strategy can be used if one wishes to estimate the models independently. In this case just work with Cholesky decomposition of matrices $\boldsymbol{\Xi}$, $\boldsymbol{\Omega}$, $\boldsymbol{\psi}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$.

The above described model treats the visual attention data as a means to drive the preferences. The continuous model component of the system models the visual attention in terms of time spent on various alternatives, including its labels, which is then used as an explanatory variable in the choice model component). On the other hand, to test the hypothesis that habits, goals, and constraints work as a screening mechanism, we use the visual attention as an explanatory variable in the choice model but passed as a penalty. That is, we add a penalty term to the utility equation on each alternative which may be a function of individuals' habits and time-spent on alternatives.

$$U_{alt} = V_{alt} + \ln[1/1 + \exp(\mu_{alt})] + \xi_{alt}$$

Where U_{alt} is the utility of the alternative, V_{alt} is the deterministic component of the utility, ξ_{alt} is the normally distributed error term, and μ_{alt} is the penalty function. Further μ_{alt} = $\mu_{alt} = \text{f(individuals'havits, time spent on the alternative)}$. The first parametrization $[1/1 + \exp(\mu_{alt})]$ ensures that the value in the square bracket is bounded between 0 and 1 so that the natural logarithm of the function is bounded between -∞ and 0. This way, an alternative becomes unavailable or gets pushed out from the consideration set as soon as the expression $ln[1/1 + \exp(\mu_{alt})]$ takes a value of -∞. Please note that there is no stochastic component in the penalty function. Adding the stochastic component creates additional computational challenges in the realm of Probit kernel.

References

Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological*, *37*(9), 837-855.

Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, *45*(7), 923-939.

Bhat, C. R. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B: Methodological*, *79*, 50-77.

Bhat, C. R., & Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B: Methodological*, *67*, 68-85.

Bhat, C. R., Pinjari, A. R., Dubey, S. K., & Hamdi, A. S. (2016). On accommodating spatial interactions in a generalized heterogeneous data model (GHDM) of mixed types of dependent variables. *Transportation Research Part B: Methodological*, *94*, 240-263.

Connors, R. D., Hess, S., & Daly, A. (2014). Analytic approximations for computing probit choice probabilities. *Transportmetrica A: Transport Science*, *10*(2), 119-139.

Craig, P. (2008). A new reconstruction of multivariate normal orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 227-243.

Daziano, R. A. (2015). Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model. *Transportation Research Part B: Methodological*, *76*, 1-26.

Franzese, R. J., Hays, J. C., & Schaffer, L. M. (2010). Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. *APSA 2010 Annual Meeting* https://ssrn.com/abstract=1643867

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, *1*(2), 141-149.

Hajivassiliou, V., McFadden, D., & Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics*, *72*(1), 85-134.

Heiss, F. (2010). The panel probit model: adaptive integration on sparse grids. In Greene, W., & Hill, R. C. (Eds.), *Maximum simulated likelihood methods and applications* (pp. 41-64). Bingley, UK: Emerald Group Publishing Limited.

Keane, M. P. (1992). A note on identification in the multinomial probit model. *Journal of Business & Economic Statistics*, *10*(2), 193-200.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*(1), 103-120.

Paleti, R., & Bhat, C. R. (2013). The composite marginal likelihood (CML) estimation of panel ordered-response models. *Journal of Choice Modelling*, *7*, 24-43.

Patil, P. N., Dubey, S. K., Pinjari, A. R., Cherchi, E., Daziano, R., & Bhat, C. R. (2017). Simulation evaluation of emerging estimation techniques for multinomial probit models. *Journal of Choice Modelling*, *23*, 9-20.

Sándor, Z., & András, P. (2004). Alternative sampling methods for estimating multivariate normal probabilities. *Journal of Econometrics*, *120*(2), 207-234.

Sidharthan, R., & Bhat, C. R. (2012). Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis*, *44*(4), 321-349.

Train, K. (2000). *Halton sequences for mixed logit*: UC Berkeley: Department of Economics.

Varin, C. (2008). On composite marginal likelihoods. *AStA- Advances in Statistical Analysis*, *92*(1), 1-28.

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*, 5-42.

Varin, C., & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, *92*(3), 519-528.

Zhou, Y., Wang, X., & Holguín-Veras, J. (2016). Discrete choice with spatial correlation: A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM). *Transportation Research Part B: Methodological*, *94*, 440-455.

WEB APPENDIX C: PARTICIPANT FLOW DIAGRAM

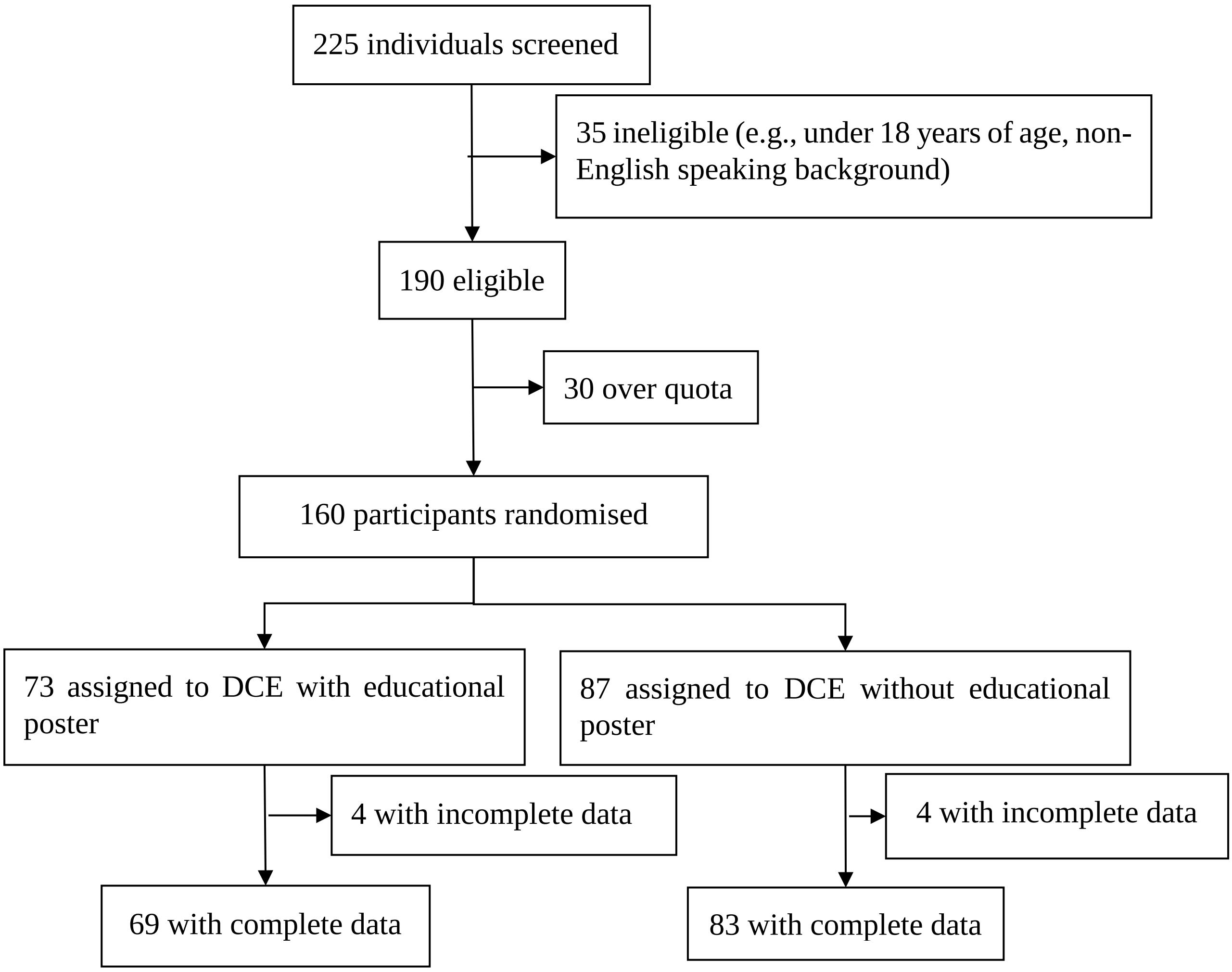


Fig. C.1. Flow chart of participants included/ excluded from eye-tracked discrete choice experiment (DCE)

Eye-tracking data were not captured for eight participants, due to technical errors, and therefore were excluded from this analysis. Of the remaining 152 participants, eye-tracking data were detected for some but not all of the 20 choice tasks for 13 individuals. These individuals were therefore excluded from the main analysis but used to test out-of-sample predictive power.

WEB APPENDIX D: DEMOGRAPHIC CHARACTERISTICS OF EYE-TRACKING STUDY PARTICIPANTS

Table D.1: Demographic characteristics of eye-tracking study participants in main estimation (n=139)

Characteristic	n (%)	Australian population
Females [a]		
18-35 years	41 (51.3%)	32.2%
36-59 years	32 (40.0%)	40.4%
60 years and over	7 (8.8%)	27.4%
Males [a]		
18-35 years	28 (47.5%)	34.0%
36-59 years	21 (35.6%)	40.7%
60 years and over	10 (17.0%)	25.3%
Equivalised household income quintile [b]		
Q1 (lowest income)	39 (28.1%)	20%
Q2	31 (22.3%)	20%
Q3	23 (16.6%)	20%
Q4	29 (20.9%)	20%
Q5 (highest income)	17 (12.2%)	20%
Highest educational attainment [c]		
Year 11 or below	4 (2.9%)	26%
Year 12 or equivalent	21 (15.1%)	18%
TAFE or Certificate, diploma	18 (13.0%)	21%
Undergraduate university	54 (38.9%)	29%
Postgraduate university	42 (30.2%)	6%
Body Mass Index (BMI) [d]		
<25kg/m^2 (normal or underweight)	78 (56.9%)	37.2%
25 to 30 kg/m^2 (overweight)	42 (30.7%)	35.3%
>30 kg/m^2 (obese)	17 (12.4%)	27.5%
SSB purchase frequency from convenience store in the past month [e]		
On about half of days or more	76 (34.7%)	-
A few times	56 (40.3%)	-
Never	7 (5.0%)	-

n=139 eye-tracking participants in main analysis (from total sample of 160). National statistics derived from: [a] Australian Bureau of Statistics (ABS) (2011), "Australian Demographic Statistics, Jun 2016, 'Table 1. Population Change, Summary - Australia ('000)', data cube: Excel spreadsheet, cat no. 3101.0," Available at: http://www.abs.gov.au/ ; [b] ABS (2013), "Household income and income distribution Australia." Available at: http://www.abs.gov.au/.[c] ABS (2016), "Education and Work, Australia, May 2016" Available at: http://www.abs.gov.au/; [d] BMI missing for 2 participants. ABS (2015)"National Health Survey: First Results, 2014-15, cat no. 4364.0.55.001," Available at http://www.abs.gov.au/. [e] 'Regular SSB (sugar-sweetened beverage) consumers' were defined as those who reported consumption of a SSB purchased from a convenience store at least a few times in the past month

WEB APPENDIX E: DESCRIPTIVE STATISTICS OF VISUAL ATTENDANCE

There was a significant correlation between fixation duration examining relevant choice set information with fixation duration out of choice set (R^2= 0.92, *p*-value <0.001). Based on this, the analyses below used ratio of time spent in and out of consideration set rather than absolute duration, unless otherwise specified, to avoid results being unduly influenced by overall time to complete the task. Where sample summaries are presented (rather than per choice set), this ratio is further adjusted for number of choice sets for which eye-tracking data was captured.

Linear regressions found that the first four choice tasks had a longer mean duration than the last four tasks, even when adjusted for age and gender (*p*-value <0.01), suggesting learning or fatigue. Ratio of relevant to irrelevant visual attention duration increased in the last compared to the first 4 tasks (*p*-value <0.01)

Stated Attendance

One hundred percent of respondents stated they sometimes or always considered price, and 95% and 99% stated they sometimes or always considered volume and beverage type, respectively. All beverage types were sometimes or always considered by more than 30% of the sample.

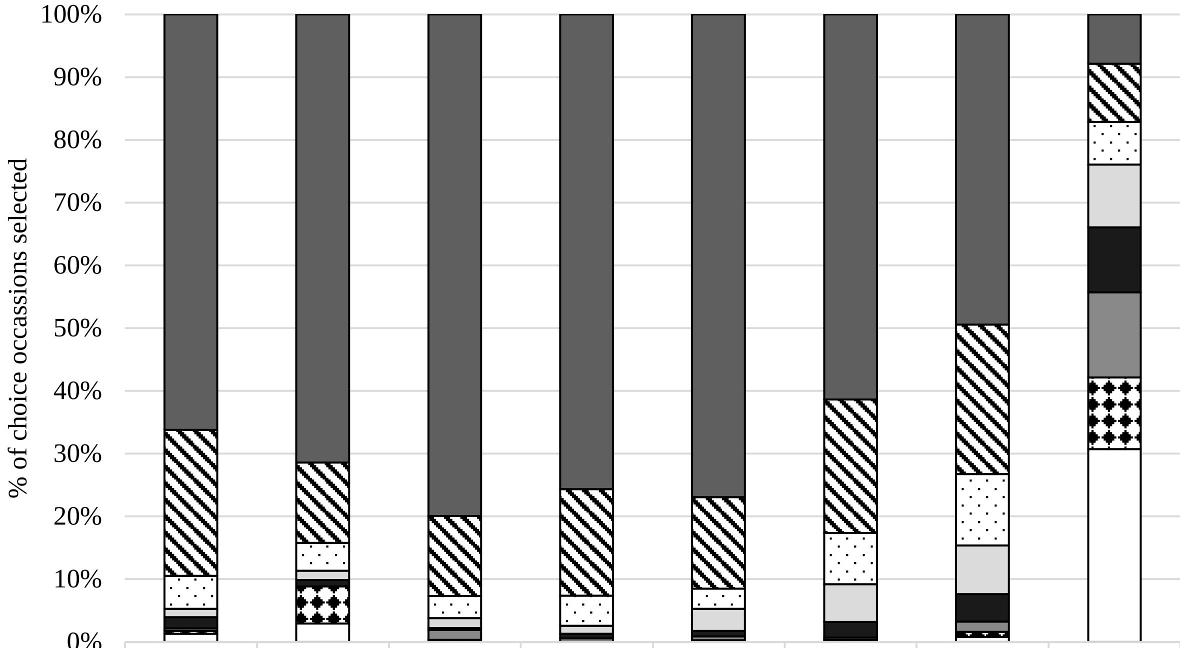
Relationship Between Stated and Visual Attendance

No significant difference was found in fixation duration on beverage, price or volume labels by stated importance on a 5-point Likert scale as per participants using an ANOVA (all *p*-values ≥ 0.34). Attribute and alternative fixation duration were not predicted by relevant stated attribute or alternative non-attendance using linear regression (all *p*-values >0.05). A higher score on strength of habit questionnaire (stronger SSB consumption habit) was positively related to fixation duration on energy drinks (*p*-value=0.06) and flavored milk (*p*-

value=0.03), and negatively related to fixation duration on “no drink” alternative (*p*-value=0.01) using linear regression when adjusted for age and gender. This suggests that SSB consumption habit may be related to visual attention, but this unadjusted analysis was unable to distinguish the direction of effect. No significant relationships were seen between stage of readiness to drink fewer SSBs and fixation duration by beverage type or overall time on choice task.

Relationship between visual attention and choice

Respondents spent less visual fixation time on the chosen alternative across choice tasks compared to other alternatives. **Fig. E.1** shows a detailed breakdown of visual attention time spent on chosen alternatives. On more than 50% of occasions, the chosen alternative received the least amount of visual attention.



	Energy drink	Plain low-fat milk	Flavoured milk	Bottled water	Soft drink (regular)	Soft drink (diet)	Fruit juice	No beverage
R8	66%	71%	80%	76%	77%	61%	49%	8%
R7	23%	13%	13%	17%	15%	21%	24%	9%
R6	5%	4%	4%	5%	3%	8%	11%	7%
R5	1%	1%	2%	1%	4%	6%	8%	10%
R4	2%	1%	0%	1%	1%	2%	4%	10%
R3	0%	0%	2%	1%	1%	0%	2%	14%
R2	0%	6%	0%	0%	0%	0%	1%	11%
R1	1%	3%	0%	0%	0%	0%	1%	31%

Chosen alternative

Fig. E.1: Distribution of chosen beverage alternative as a function of amount of time spent looking at that particular alternative. R1 to R8 indicate the ranking in ascending order of time spent looking at an alternative

WEB APPENDIX A: DISCRETE CHOICE EXPERIMENT DETAILS

Experimental Design

In the labelled DCE, participants selected a beverage within a hypothetical convenience store setting. Each participant completed 20 choice tasks involving three SSB alternatives (energy drink, flavored milk, regular soft drink (soda)), four non-sugar-sweetened alternatives (non-SSBs: plain low-fat milk, fruit juice, diet soft drink, bottled water), and a "no drink" alternative (meaning that they would "consume no drink on this occasion"). Each beverage was described by alternative-specific prices and generic volume attributes which varied over four levels each. An orthogonal design was generated using Ngene software (Rose, Collins, Bliemer & Hensher 2009). An example choice task (**Figure A.1**) and list of attribute levels for each alternative (**Table A.1**) are given below. Prior to completing the choice tasks, half of participants were randomly exposed to a real-world educational message designed to discourage selection of SSBs. The other half did not see any message. After the 10th choice task all participants were presented with a message reminding them to "consider their options carefully", to ameliorate potential fatigue effects. As described later, we tested for the impact of the education message in the analysis and found no significant effect on beverage choice, hence sub-samples were pooled and we used the full sample in the estimation results we present later.

Following the DCE, participants completed questions on stated attendance to attributes and alternatives as well as strength of SSB consumption habit. This included an 11-point scale of readiness to consider reducing SSB intake based on a validated tool to assess readiness to quit smoking (Biener & Abrams 1991) and the Self-Report Behavioral Automaticity Index, a 4-item measure of habit strength measured on a 5-point Likert scale with higher scores signifying a stronger habit (Gardner, Abraham, Lally & de Bruijn 2012).

Please read the information below carefully:

For this survey, imagine that you are now going into your local convenience store (e.g. newsagent, 7-Eleven, independent milk bar) with the intention to buy a pre-packaged drink (in a bottle, can or carton) for yourself to drink immediately. Please note that this does not include supermarkets or petrol stations, hospital, sports and recreation facilities etc. where you may have entered the store for another main purpose.

You will be presented with a number of individual shopping scenarios. In each scenario, you will be presented with 7 drink options, each drink will be described by its price and size (volume). The information describing price and volume will change between each task. Assume the displayed products are the only available options.

Please note that ‘energy drink’ refers to a drink specifically designed to give a short term ‘energy’ boost such as those with added taurine, guarana or caffeine. It does not include ‘sports drinks’.

In each scenario, please indicate which **one option** you would choose. Either select the drink that you would buy OR select ‘no beverage’ if you would exit the convenience store without having purchased a drink in this situation, after already having entered the convenience store with the intention to buy a pre-packaged drink. This would mean you would not consume a drink on this occasion.

Please also treat each scenario as separate (i.e. as if you had not just made the previous choice).

Please note: there are no right or wrong answers, the researchers are interested in your individual preference among the options presented.

On the next page will be a practice scenario.

Fig. A.1 (part 1): Discrete Choice Experiment scenario explanation and sample choice scenario

You have gone into your local convenience store now (e.g. newsagent, 7-Eleven, independent milk bar) with the intention to buy a pre-packaged drink (in a bottle, can or carton) to drink immediately yourself. Select the option below that you would choose.

	Energy drink	Plain low-fat milk	Flavoured milk	Bottled water	Soft drink (regular)	Soft drink (diet)	Fruit juice	No drink
Price	$5.90	$5.00	$6.50	$1.00	$6.50	$6.50	$5.90	N/A
Volume (size)	200mL	200mL	200mL	600mL	200mL	200mL	200mL	N/A
Which would you choose?								

Fig. A.1 (part 2): Discrete Choice Experiment scenario explanation and sample choice scenario

Table A.1: Alternative attribute levels

Alternative	Experimental volumes tested	Experimental prices tested (AUD)
Energy drink	200mL, 330mL, 460mL, 600mL	$2.00, $3.30, $4.60, $5.90
Soft drink (regular)		$2.00, $3.50, $5.00, $6.50
Soft drink (diet)		$2.00, $3.50, $5.00, $6.50
Plain low-fat milk		$1.00, $2.30, $3.70, $5.00
Flavoured milk		$2.00, $3.50, $5.00, $6.50
Bottled water		$1.00, $2.30, $3.70, $5.00
Fruit juice		$2.00, $3.30, $4.60, $5.90

Implementation of Eye-Tracking Measurements

All participants completed the DCE in an eye-tracking laboratory at the study university. The task involved sitting and completing the DCE on a computer-screen. A discrete, web-cam like device tracked eye movements (Tobii Pro, 2011, Tobii TX300; Stockholm, Sweden). The choice tasks were presented through a web-browser using Tobii Studio version 3.2 (Tobii Pro, 2012, Stockholm, Sweden). Eye movements were recorded at 300 Hz on a screen resolution of 1920 x 1080 pixels. Minimum fixation duration was 60ms.

Participants were positioned with their head 64cm from the screen as per recommended Tobii T-series validity requirements. Participants' eye-movements were calibrated before the experiment using nine static calibration locations on the screen. Participants were eye-tracked during the entire survey, however only visual attention data corresponding to the DCE are analyzed here.

References

Biener, L., & Abrams, D. B. (1991). The Contemplation Ladder: validation of a measure of readiness to consider smoking cessation. *Health Psychology*, *10*(5), 360.

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102.

Rose, J. M., Collins, A. T., Bliemer, M. C., & Hensher, D. A. (2009). *Ngene stated choice experiment design software*, (Version 1.1.2). Sydney, Australia: University of Sydney.

WEB APPENDIX B: DETAILED METHODOLOGY

Our model has three components: continuous (visual attention duration), ordered (habit measures), and nominal (choice outcome) variables. We first describe the construction of each component separately and then bring them together using a covariance approach.

Visual Attention Model

Let $\tilde{t}$ be the index for task instance ($\tilde{t}$ = 1,2,…, $\tilde{T}$) and $\tilde{h}$ be the index for the continous outcome ($\tilde{h}$ = 1,2,…, $\widetilde{H}$). Then, we can write in the usual linear regression form:

(1) $\tilde{y}_{\tilde{h},\tilde{t}} = \tilde{\rho}_{\tilde{h}}\tilde{y}_{\tilde{h},\tilde{t}-1}+\gamma'_{\tilde{h}}x_{\tilde{h},\tilde{t}} + \xi_{\tilde{h}}$

Where $\tilde{\rho}_{\tilde{h}}$ is the autoregressive (AR-1) coefficient which ranges between -1 to 1, $x_{\tilde{h},\tilde{t}}$ is a $(k_{\tilde{h}} \times 1)$ vector of exogenous variables (including a constant), $\gamma_{\tilde{h}}$ is the corresponding $(k_{\tilde{h}} \times 1)$ vector of coefficients, and $\xi_{\tilde{h}}$ is a normally distributed error term. The autoregressive coefficient helps us capture the time-multiplier effect (i.e., the effect of previous time period on the current time period for both observed and unobserved variables). Now, stack all the $\widetilde{H}$ continuous outcomes for all task instances $\tilde{T}$ in a vector $\tilde{y} = (\tilde{y}_{1,1}, \tilde{y}_{2,1}, \dots, \tilde{y}_{\widetilde{H},1}, \dots, \tilde{y}_{\widetilde{H},\tilde{T}})$ $(\widetilde{H}\tilde{T} \times 1)$, autoregressive coefficient $\tilde{\rho}_{\tilde{h}}$ for all the $\widetilde{H}$ continuous outcomes in a vector $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, \tilde{\rho}_2, \dots, \tilde{\rho}_{\widetilde{H}})$ of size $(\widetilde{H} \times 1)$, exogenous variable's coefficients in a matrix $\gamma = (\gamma'_1, \gamma'_2, \dots, \gamma'_{\widetilde{H}})$ of size $(\widetilde{H} \times k_{\tilde{h}})$, exogenous variables in a matrix $x_{\widetilde{H},\tilde{T}} = (x'_{1,1}, x'_{2,1}, \dots, x'_{\widetilde{H},1} \dots, x'_{\widetilde{H},\tilde{T}})$ of size $(\widetilde{H}\tilde{T} \times k_{\tilde{h}})$ and all the error terms in $(\xi = \xi_1, \xi_2, \dots, \xi_{\widetilde{H}})$ of size $(\widetilde{H} \times 1)$. Where (,…) inside the bracket refers to placement of next variable in the next row. Also, let Ξ be the covariance matrix of ξ.

Now, to write the equation (1) in the matrix form, define the following matrices:

construct a matrix $\mathbf{F}_{\widetilde{H}\tilde{T}}$ of size $(\widetilde{H}\tilde{T} \times \widetilde{H}\tilde{T})$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of matrix $\mathbf{F}_{\widetilde{H}\tilde{T}}$.

for *j* = 2 to $\tilde{T}$

for *i* = 1 to $\widetilde{H}$

$$\mathbf{F}_{\widetilde{H}\tilde{T}}\left[(j-1) * \widetilde{H} + i, (j-1) * \widetilde{H} + i\right] = \tilde{\rho}[i,1]$$

end

end

For example: a $F_{\widetilde{H}\tilde{T}}$ matrix with $\widetilde{H} = 2$ and $\tilde{T} = 3$ will take the following form:

$$\mathbf{F}_{\widetilde{H}\tilde{T}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \tilde{\rho}_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \tilde{\rho}_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{\rho}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{\rho}_2 & 0 & 0 \end{bmatrix}$$

Also, construct a matrix $\mathbf{I}_{\widetilde{H}\tilde{T}}$ of size $(\widetilde{H}\tilde{T} \times \widetilde{H}\tilde{T})$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix $\mathbf{I}_{\widetilde{H}\tilde{T}}$

for j = 2 to $\widetilde{T}$

for i = 1 to $\widetilde{H}$

$$I_{\widetilde{H}\widetilde{T}}\left[(j-1) * \widetilde{H} + i, (j-2) * \widetilde{H} + i\right] = 1$$

end

end

With this, equation (1) may be written in the matrix form as follows:

(2) $\tilde{y}$=S*[sumc[$(\tilde{\gamma}.* x_{\widetilde{H}\widetilde{T}})'$] + $\tilde{\xi}$]

where $\tilde{\gamma}$=ones$(\widetilde{T}, 1).*.\gamma$, $\tilde{\xi} = \text{ones}(\widetilde{T}, 1).*.\xi$, ".*." refers to Kronecker product, ".*" refers to element by element multiplication, the operator sumc(.) returns the sum of columns of matrix in a column vector, ones$(\widetilde{T}, 1)$ indicates a vector of size $\widetilde{T}$ whose all the elements are filled with a value of "1", $\mathbf{1}_{\widetilde{H}\widetilde{T}}$ refers to an identity matrix of size $\widetilde{H}\widetilde{T}$ and $\mathbf{S} = [\mathbf{1}_{\widetilde{H}\widetilde{T}} - (\boldsymbol{F}_{\widetilde{H}\widetilde{T}}.*.\boldsymbol{I}_{\widetilde{H}\widetilde{T}})]^{-1}$ of size $(\widetilde{H}\widetilde{T} \times \widetilde{H}\widetilde{T})$.

From equation (2), it can be observed that ỹ is distributed normally with mean S*[sumc[(γ̃.*x_(H̃T̃))'] and covariance $\mathbf{S}*[\mathbf{1}_{\widetilde{T}}.*.\Xi] * \boldsymbol{S}'$ [1]. Also, to maintain the bound on the autoregressive parameter vector $\widetilde{\boldsymbol{\rho}}$, we parametrize the parameter as $\widetilde{\boldsymbol{\rho}} = \widetilde{\boldsymbol{\rho}}_{\boldsymbol{p}}/[\mathbf{1} + (\widetilde{\boldsymbol{\rho}}_{\boldsymbol{p}})^{\mathbf{2}}]^{\mathbf{0.5}}$. Where $\tilde{\rho}_p$ is the value passed to the optimization module.

Habit and Goal Variable Model

Strength of habit and goals were considered on an ordinal scale. Let $\ddot{t}$ be the index for task instance ($\ddot{t}$ = 1,2,…,$\ddot{T}$) and $\ddot{n}$ be the index for the ordinal outcome ($\ddot{n}$ = 1,2,…,$\ddot{N}$). Also, let $J_{\ddot{n}}$ (>1) be the number of categories for the $\ddot{n}^{th}$ordinal outcome and the correponding index be $j_{\ddot{n}} =$

[1] In a time-series based regression such as the one described here, the dependence between a particular continuous variable's task instances or time periods is generated through the autoregressive parameter and the dependence across continuous variables is captured through the covariance matrix Ξ. This allows the analyst to exclude random taste heterogeneity in the model. Our experience with the model suggests that recovery of random parameters in such a highly non-linear model is relatively difficult. Therefore, we suggest the inclusion of either autoregressive parameters or random taste parameters in the model depending upon the analyst's requirement. Random taste parameters can be included in a straighforward manner as follows: let $\mathbf{\Omega}$ be a $(k_{\tilde{h}} \times k_{\tilde{h}})$ covariance matrix of exogenous variables. Then, stack the exogenous variables in a matrix of size $(\widetilde{H}\widetilde{T} \times k_{\tilde{h}}\,\widetilde{T}$ as follows:

$$X_{\widetilde{H}\widetilde{T}} = \begin{bmatrix} x'_{1,1} & 0 & 0 & 0 \\ x'_{2,1} & 0 & 0 & 0 \\ x'_{\widetilde{H},1} & 0 & 0 & 0 \\ 0 & \vdots & 0 & 0 \\ 0 & \vdots & 0 & \vdots \\ 0 & 0 & \cdots & x'_{\widetilde{H}\widetilde{T}} \end{bmatrix}$$

With this $\widetilde{\boldsymbol{y}} \sim \boldsymbol{N}[\mathbf{S} * [\mathbf{sumc}[(\widetilde{\boldsymbol{\gamma}}.* \boldsymbol{x}_{\widetilde{H}\widetilde{T}})'], \mathbf{S} * [\mathbf{1}_{\widetilde{T}}.*.\Xi + \boldsymbol{X}_{\widetilde{H}\widetilde{T}} * (\mathbf{1}_{\widetilde{T}}.*.\mathbf{\Omega}) * \boldsymbol{X}'_{\widetilde{H}\widetilde{T}}] * \boldsymbol{S}']$

$(1,2,\ldots,J_{\ddot{n}})$. [2]Let $\ddot{y}^*_{\ddot{n},\ddot{t}}$ be the underlying latent variable. Then in the usual ordered-response formulation, we may write:

(3) $\ddot{y}^*_{\ddot{n},\ddot{t}} = \delta'_{\ddot{n}} x_{\ddot{n},\ddot{t}} + \zeta_{\ddot{n}}$, and $\Psi_{\ddot{n},a_{\ddot{n},\ddot{t}}-1} < \ddot{y}^*_{\ddot{n},\ddot{t}} < \Psi_{\ddot{n},a_{\ddot{n},\ddot{t}}}$, if $\ddot{y}_{\ddot{n},\ddot{t}}=a_{\ddot{n},\ddot{t}}$

where $x_{\ddot{n},\ddot{t}}$ is a $(k_{\ddot{n}} \times 1)$ vector of exogenous variables (including constant)[3] , $\delta_{\ddot{n}}$ is the corresponding $(k_{\ddot{n}} \times 1)$ vector of parameters, $a_{\ddot{n},\ddot{t}}$ is the observed outcome category at time period $\ddot{t}$ for the $\ddot{n}^{th}$ ordinal variable, and $\zeta_{\ddot{n}}$, is a standard normal error term[4]. Further, the thresholds for the ordinal outcome should be in ascending order (i.e.,

$\Psi_{\ddot{n},0} < \Psi_{\ddot{n},1} < \cdots < \Psi_{\ddot{n},J_{\ddot{n}}-1} < \Psi_{\ddot{n},J_{\ddot{n}}}$; $\Psi_{\ddot{n},0} = -\infty, \Psi_{\ddot{n},1}$=0, and $\Psi_{\ddot{n},J_{\ddot{n}}} = \infty$).

Now, stack the threshold elements as follows:

$\Psi_{\ddot{n}} = (\Psi_{\ddot{n},0}, \Psi_{\ddot{n},1}, \ldots, \Psi_{\ddot{n},J_{\ddot{n}}})$ $[(J_{\ddot{n}} + 1) \times 1]$ vector,

$\Psi_{\ddot{t}} = (\Psi'_1, \Psi'_2, \ldots, \Psi'_{\ddot{N}})'$ $[\ddot{N}(J_{\ddot{n}} + 1) \times 1]$ vector,

$$\Psi_{low} = (\Psi_{1,a_{1,1}-1}, \Psi_{1,a_{2,1}-1}, \ldots, \Psi_{1,a_{\ddot{N},1}-1}, \ldots, \Psi_{1,a_{\ddot{N},\ddot{T}}-1})\ [\ddot{N}\ddot{T} \times 1\ ] \text{ vector, and}$$

$\Psi_{up} = (\Psi_{1,a_{1,1}}, \Psi_{1,a_{2,1}}, \ldots, \Psi_{1,a_{\ddot{N},1}}, \ldots, \Psi_{1,a_{\ddot{N},\ddot{T}}})[\ddot{N}\ddot{T} \times 1]$ vector [5].

Further, stack the $\ddot{N}\ddot{T}$ underlying latent variables in a $(\ddot{N}\ddot{T} \times 1)$ vector $\ddot{\boldsymbol{y}}^* = (\ddot{y}^*_{1,1}, \ddot{y}^*_{2,1}, \ldots, \ddot{y}^*_{\ddot{N},1}, \ldots, \ddot{y}^*_{\ddot{N},\ddot{T}})$, exogenous variables in a matrix $\boldsymbol{x}_{\ddot{N},\ddot{T}} = (\boldsymbol{x}'_{1,1}, \boldsymbol{x}'_{2,1}, \ldots, \boldsymbol{x}'_{\ddot{N},1}, \ldots, \boldsymbol{x}'_{\ddot{N},\ddot{T}})$ of size $(\ddot{N}\ddot{T} \times k_{\ddot{n}})$, exogenous variables' coefficients in a matrix $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2, \ldots, \boldsymbol{\delta}'_{\ddot{N}})$ of size $(\ddot{N} \times k_{\ddot{n}})$, and all the error terms in $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_{\ddot{N}})$ of size $(\ddot{N} \times 1)$. Also, let Γ be the correlation matrix of ζ. Then, we may write, equation (3) in the matrix form as follows:

(4) $\ddot{y}^* = \text{sumc}\left[\left(\ddot{\boldsymbol{\delta}}.* \boldsymbol{x}_{\ddot{N},\ddot{T}}\right)'\right] + \boldsymbol{\zeta}, \boldsymbol{\Psi_{low}} < \ddot{\boldsymbol{y}}^* < \boldsymbol{\Psi_{up}}$ [6] where $\boldsymbol{\delta} = ones(\ddot{T},1).*.\boldsymbol{\delta}$ and $\ddot{\boldsymbol{\zeta}} = ones\ (\ddot{T},1).*.\boldsymbol{\zeta}$.

---

[2] The requirement of number of categories to be greater than 1 instead of 2 enables us to model binary outcomes as ordinal outcomes with no additional thresholds being estimated.

[3] We fix the second threshold to a value of zero and thus estimate the constant for every ordinal outcome.

[4] The normalization on the error term is needed for identification, as in the usual ordered-response model; see McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*(1), 103-120..

[5] Here for ease in notation, we assume that all the ordinal outcomes have same number of categories. However, this may not be the case. In situations with different number of categories, one can fill the remaining/extra cells with zeros.

[6] If the ordinal outcomes are observed for more than one time period, then one would be tempted to include random-taste parameters in order to capture the dependence across time-periods. Similar to the continuous variable model, the incorporation of random-taste parameter is straightforward. Let $\boldsymbol{\Psi}$ be a $(k_{\ddot{n}} \times k_{\ddot{n}})$, covariance matrix of exogenous variables. Then, stack the exogenous variables in a matrix of size $(\ddot{N}\ddot{T} \times k_{\ddot{n}}\ddot{T})$ as follows: $X_{\ddot{N}\ddot{T}} = \begin{bmatrix} x'_{1,1} & 0 & 0 & 0 \\ x'_{2,1} & 0 & 0 & 0 \\ x'_{\ddot{N},1} & 0 & 0 & 0 \\ 0 & \vdots & 0 & 0 \\ 0 & \vdots & 0 & \vdots \\ 0 & 0 & \ldots & x'_{\ddot{N},\ddot{T}} \end{bmatrix}$

With this the covariance matrix for $\ddot{\boldsymbol{y}}^*$ becomes $[1_{\ddot{T}} .*.\Gamma + X_{\ddot{N}\ddot{T}} * (1_{\ddot{T}}.*.\Psi) * X'_{\ddot{N}\ddot{T}}]$

Choice Model

Let *t* be the index for choice occasion ($t = 1,2, \dots, T$), *i* be the index for nominal outcome ($i = 1,2, \dots, I$), and *k* be the index for number of alternatives per nominal outcome ($k = 1,2, \dots, K$)[7]. Then, we can write the utility of alternative *k* from the i^{th} nominal variable in the time period *t* as:

$$(5) \qquad U_{i_k t} = \boldsymbol{\beta}_i' \boldsymbol{x}_{i_k t} + \boldsymbol{\varepsilon}_{i_k}$$

where $x_{i_k t}$ is a (g_i ×1) vector of exogenous variables at choice occasion *t*, β_i is the corresponding (g_i ×1) vector of coefficients, and ε_{i_k} is a normally distributed error term (all the notations correspond to the nominal outcome *i*). Now, define the following notations:

I_k (total number of alternatives) = $\sum_{t=1}^{I} \boldsymbol{i_K}$,

U_{it} = $(U_{1t}, U_{2t}, \dots, U_{i_K t})[(i_K \times 1)]$ vector, U_t = $(U_{1t}, U_{2t}, \dots, U_{It})[(I_K \times 1)]$ vector,

$\boldsymbol{U} = (\boldsymbol{U_1}, \boldsymbol{U_2}, \dots, \boldsymbol{U_T})[(I_K \times 1)]$ vector, $\boldsymbol{\beta}=(\boldsymbol{\beta}_{11}', \boldsymbol{\beta}_{12}', \dots, \boldsymbol{\beta}_{1K}', \dots, \boldsymbol{\beta}_{IK}')[(I_K \times g_i)]$ vector,

$x=(x_{11t}', x_{12t}', \dots, x_{1Kt}', \dots, x_{IKt}'[(I_K \times g_i)]$ matrix, $x=(x_1, x_2, \dots, x_T)[(TI_K \times g_i)]$ matrix,

$\varepsilon_i = (\varepsilon_{i_1}, \varepsilon_{i_2}, \dots, \varepsilon_{i_K})[(i_k \times 1)]$ vector, $\boldsymbol{\varepsilon_t} = (\varepsilon_{1_1}, \varepsilon_{1_2}, \dots, \varepsilon_{1_k}, \dots, \varepsilon_{I_k})$ $[(I_K \times 1)]$ vector,

$\tilde{\beta} = [\text{ones}(T,1).*.\beta][(TI_K \times g_i)]$ matrix, and $\varepsilon = [\text{ones}(T,1).*.\ \varepsilon_t][(TI_K \times 1)]$ vector.

Also, let $\boldsymbol{\Lambda}_i$ be the covariance matrix of ε_i. Then, we may write, equation (5) in the matrix form as follows:

$$(6) \qquad \boldsymbol{U} = \mathbf{sumc}\left[(\tilde{\boldsymbol{\beta}}.* \boldsymbol{x})'\right] + \boldsymbol{\varepsilon}$$

With this, we may write the distribution of $\boldsymbol{U}$ as

$\boldsymbol{U} \sim N_{(TI_K \times TI_K)}\left[\text{sumc}\left[\left(\tilde{\beta}.* x\right)'\right],\ \mathbf{1}_T.*.\boldsymbol{\Lambda}\right]$. Where,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda_1} & \boldsymbol{\Lambda_{12}} & \boldsymbol{\Lambda_{1,I-1}} & \boldsymbol{\Lambda_{1,I}} \\ \boldsymbol{\Lambda_{12}'} & \boldsymbol{\Lambda_2} & \boldsymbol{\Lambda_{2,I-1}} & \boldsymbol{\Lambda_{2,I}} \\ \boldsymbol{\Lambda_{1,I-1}'} & \boldsymbol{\Lambda_{2,I-1}'} & \ddots & \boldsymbol{\Lambda_{I-1,I}} \\ \boldsymbol{\Lambda_{1,I}'} & \boldsymbol{\Lambda_{2,I}'} & \boldsymbol{\Lambda_{I-1,I}'} & \boldsymbol{\Lambda_I} \end{bmatrix}, \text{ and}$$

In the $\boldsymbol{\Lambda}$ matrix, the off-diagonal elements capture dependencies across nominal variables through correlation in unobserved variables[8].

---

[7] We supress the index for the individual participant (i) for ease in presentation as it is a non-spatial model.

[8] This is not to say that this is the only way to capture dependencies across nominal variables. Another way to capture dependency may be achieved by random-taste parameter. However, this would require the analyst to have a common exogenous variable in all the nominal variables and in all the alternatives. This could be rather difficult given the differential impact of the same exogenous variable on different choice dimensions. On the other hand, one is free to incorporate random-taste parameters at the nominal variable level (with full or no correlation) with no cross-correlation across nominal variables. It could be incorporated as follows: Let Σ_i be the ($i_G \times i_G$) covariance matrix of exogenous variables for the *i*[th] nominal variable. Where $G = \sum_{r=1}^{K} i_r$ is the total number of exogenous variables in the *i*[th] nominal variable. Then, stack the exogenous variables for all the nominal variables in a matrix of size ($I_K \mathrm{T} \times TG$) and all the random-taste parameter matrices into a **Σ** matrix as follows:

Since only the differences in utility matter, only the difference of error-terms are identifiable and not the actual error terms after performing the normalization to fix the scale of utility. Therefore, we normalize the top diagonal element to 1 for estimation purposes (Keane 1992). However, all the differenced error matrices must originate from the same un-differenced error matrix. To do so, append the matrices Λ_i by adding a row and column of zeros on the top (Sidharthan & Bhat 2012) i.e., $\Lambda_i = \begin{bmatrix} 0 & 0_{1,i_K-1} \\ 0_{i_K-1,1} & \Lambda_i \end{bmatrix}$ or multiply the matrix $\mathbf{\Lambda}$ with a matrix **D** (i.e., expanded differenced matrix $\mathbf{D\Lambda}$ for all the nominal variables) constructed as follows:

Define a matrix **D** of size$[(I_K) \times (I_K - I)]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **D**.

for m=1 to *I*

if(m==1)

st_row =2

end_row=m_K

st_col =1

end_col= $m_K - 1$

else

st_row =$[\sum_{n=1}^{m-1} n_K]$ + 2

end_row=$[\sum_{n=1}^{m} n_K]$

st_col =$[\sum_{n=1}^{m-1} (n_K - 1)]$ +1

end_col= $[\sum_{n=1}^{m} (n_K - 1)]$

end

D[st_row: end_row, st_col:end_col]= 1_{m_K-1}

end

$$X_{I_K T} = \begin{bmatrix} x'_{1,1} & 0 & 0 & 0 \\ x'_{2,1} & 0 & 0 & 0 \\ x'_{1_k,1} & 0 & 0 & 0 \\ 0 & x'_{I,1} & 0 & 0 \\ 0 & x'_{I,1} & 0 & 0 \\ 0 & x'_{I_k,1} & 0 & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & \vdots & 0 \\ 0 & 0 & x'_{I-1_k,T-1} & 0 \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & x'_{I_k,T} \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_I \end{bmatrix}$$

With this, we may write the distribution of $\boldsymbol{U}$ as
$\boldsymbol{U} \sim N_{I_K \times I_K}\left[sumc[(\tilde{\beta}.* x)'], [1_T.*.\Lambda + X_{1_K T} * \left(1_T.*.\Sigma * X'_{I_K T}\right)]\right]$.

Now, similar to the continuous variable model, we introduce the AR-1 structure in the unobserved part of the utility as follows:

$$\varepsilon_{i_k t} = \lambda_i \varepsilon_{i_k t-1} + \eta_{i_k}$$

where λ_i is the autoregressive coefficient for the i^{th} nominal variable and η_{i_k} is the time-independent component of the error-term. That is, η_{i_k}' can be correlated for a nominal variable in a given time period, but are independent across time-periods. With this, we may re-write the equation (5) as follows with all the notations as above:

(7) $$U_{i_k t} = \boldsymbol{\beta}_i' x_{i_k t} + \varepsilon_{i_k t}$$

Now, stack the time-independent error terms and the nominal variable specific AR coefficients as follows:

$$\boldsymbol{\eta_i} = (\eta_{i_1}, \eta_{i_2}, \dots, \eta_{i_K})[(\mathrm{i}_k \times 1)] \text{ vector}, \boldsymbol{\eta_t} = (\eta_{1_1}, \eta_{1_2}, \dots, \eta_{1_k}, \dots, \eta_{1_K})[(\mathrm{I}_K \times 1)] \text{ vector},$$

$$\boldsymbol{\eta} = [\mathrm{ones}(T,1).*.\boldsymbol{\eta_t}][(TI_K \times 1)] \text{ vector, and } \lambda = (\lambda_1, \lambda_2, \dots, \lambda_I)[(I \times 1)] \text{ vector.}$$

With this, we assume that Λ_i is the covariance matrix of $\boldsymbol{\eta_i}$[9]. Now, define the additional matrices in order to write equation (7) in the matrix form:

Define a matrix **R** of size $[(TI_K) \times (TI_K)]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **R**.

for *m*=2 to *T*

    for *n*=1 to *I*

        if(*n*==1)

            for *j* =1 to n_K

                row = (*m*-1) * $I_K + j$

                col = (*m*-2) * $I_K + j$

                R[row,col]=1

            end

        else

            for *j* =1 to n_K

                row = (*m*-1) * $I_K + [\sum_{r=1}^{n-1} r_K] + j$

                col = (*m*-2) * $I_K + [\sum_{r=1}^{n-1} r_K] + j$

                R[row,col]=1

            end

[9] Here we use the same notation for the covariance matrix of $\boldsymbol{\eta_i}$ as $\boldsymbol{\varepsilon_i}$ to avoid redundancy. To be precise, one can motivate the model directly by incorporating AR-1 structure, avoiding the need for redundancy.

end

end

end

Next, construct a matrix $\boldsymbol{F_{I_K T}}$ of size $(TI_K \times TI_K)$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of matrix $\boldsymbol{F_{I_K T}}$.

```
for m=2 to T
    for n=1 to I
        if(n==1)
            for j =1 to n_K
                row = (m-1) * I_K + j
                col = (m-2) * I_K + j
                F_{I_K T}[row,col]=λ[n.1]
            end
        else
            for j =1 to n_K
                row = (m-1) * I_K + [∑_{r=1}^{n-1} r_K] + j
                col = (m-2) * I_K + [∑_{r=1}^{n-1} r_K] + j
                F_{I_K T}[row,col]= λ[n.1]
            end
        end
    end
end
```

With this, equation (7) can be written in the matrix form as follows:

(8) $\boldsymbol{U} = \text{sumc}\left[(\tilde{\beta}.* x)'\right] + \boldsymbol{C_\eta}$

where $\mathbf{C} = [\mathbf{1}_{I_K T}$ - $(\mathbf{F_{I_K T}}.*|.\mathbf{R}_{I_K T})]^{-1}$ of size $(TI_K \times TI_K)$.

From equation (8), it is easy to observe that U is distributed normally with mean $sumc\left[(\tilde{\beta}.* x')\right]$ and covariance $\mathbf{C} * [\mathbf{1_T}.*.\boldsymbol{D\Lambda}] * \mathbf{C}'$. Also, to maintain the bound on autoregressive parameter vector $\boldsymbol{\lambda}$, we parametrize the parameter as $\boldsymbol{\lambda} = \boldsymbol{\lambda_\rho}/[\mathbf{1} + (\boldsymbol{\lambda_\rho})^{2}]^{\mathbf{0.5}}$, where $\boldsymbol{\lambda_\rho}$ is the value passed to the optimization module.

Joint Model Estimation

Now, we bring the individual components of the model together to form a joint model followed by model estimation approach. To write the joint model in a matrix form, define the following vector and matrices:

$Y_tU_t = \left(\tilde{y}_{1,t}, \tilde{y}_{2,t}, \dots, \tilde{y}_{\tilde{H},t}, \ddot{y}^*_{1,t}, \ddot{y}^*_{2,t}, \dots, \ddot{y}^*_{\ddot{N},t}, U_t\right)[\left(\tilde{H} + \ddot{N} + I_K\right) \times 1]$ vector, 1

YU **=** $[(\boldsymbol{Y_1U_1}), (\boldsymbol{Y_2U_2}), \dots, (\boldsymbol{Y_TU_T})]'\left[T * \left(\tilde{H} + \ddot{N} + I_K\right) \times 1\right]$ vector, [10]

$X_t = \left(x'_{1,t}, x'_{2,t}, \dots, x'_{\tilde{H},t}, x'_{1,t}, x'_{2,t}, \dots, x'_{\ddot{N},t}, x'_t\right)[\left(\tilde{H} + \ddot{N} + I_K\right) \times \max(k_{\tilde{h}}, k_{\ddot{n},} \mathrm{g}_i)]$ matrix,

X **=** $(\boldsymbol{X_1X_2}, \dots, \boldsymbol{X_T})\left[T * \left(\tilde{H} + \ddot{N} + I_K\right) \times \max\ (k_{\tilde{h}}, k_{\ddot{n},} \mathrm{g}_i)\right]$ matrix,

$\overleftrightarrow{\boldsymbol{\beta}} = (\boldsymbol{\gamma}', \boldsymbol{\delta}', \boldsymbol{\beta}')[\left(\tilde{H} + \ddot{N} + I_K\right) \times \max\ (k_{\tilde{h}}, k_{\ddot{n},} \mathrm{g}_i)]$ matrix,

$\overleftrightarrow{\beta} = ones(T, 1).*.\overleftrightarrow{\beta}\left[T * \left(\tilde{H} + \ddot{N} + I_K\right) \times \max\ (k_{\tilde{h}}, k_{\ddot{n},} \mathrm{g}_i)\right]$ matrix.

Define a matrix **D_Mat** of size $[\left(\tilde{H} + \ddot{N} + I_K\right) \times \left(\tilde{H} + \ddot{N} + I_K - I\right)]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **D_Mat**.

D_Mat $[1{:}\,\tilde{H} + \ddot{N}, 1{:}\,\tilde{H} + \ddot{N}] = \mathbf{1}_{\tilde{H}+\ddot{N}}$

for *m* =1 to *I*

if(m==1)

st_row = $\tilde{H} + \ddot{N} + 2$

end_row = $\tilde{H} + \ddot{N} + m_K$

st_col = $\tilde{H} + \ddot{N} + 1$

end_col = $\tilde{H} + \ddot{N} + m_K - 1$

else

st_row = $\tilde{H} + \ddot{N} + [\sum_{n=1}^{m-1} n_K] + 2$

end_row = $\tilde{H} + \ddot{N} + [\sum_{n=1}^{m} n_K]$

st_col = $\tilde{H} + \ddot{N} + [\sum_{n=1}^{m-1}(n_K - 1)] + 1$

end_col = $\tilde{H} + \ddot{N} + [\sum_{n=1}^{m}(n_K - 1)]$

end

D_Mat[st_row: end_row, st_col: end_col]=1$_{\boldsymbol{m_K-1}}$

end

[10] The assumption here is that $\tilde{T} = \ddot{T} = T$. However, this need not be the case. If $\tilde{T} \neq \ddot{T} \neq T$, we assume that $T \geq \tilde{T} \,\&\, T \geq \ddot{T}$ given the focus of discrete choice models to model the choice outcome. Later we provide a design matrix which can be multiplied with the vector ***YU*** to extract the relevant components. In the meantime, all the missing values can be replaced by zero. Thus, from now on we assume $T \geq \tilde{T} \,\&\, T \geq \ddot{T}$ and thus all the matrices/vector will be created to accommodate the highest dimension *T*.

Construct a matrix **Cap_RI** of size $[T(\widetilde{H} + \ddot{N} + I_K) \times T(\widetilde{H} + \ddot{N} + I_K)]$ with all the elements being equal to zero. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **Cap_RI**.

for *m* =2 to *T*

for *n* =1 to $\widetilde{H}$

row = $(m-1) * (\widetilde{H} + \ddot{N} + I_K) + n$

col = $(m-2) * (\widetilde{H} + \ddot{N} + I_K) + n$

Cap_RI[row,col]=1

end

end

for *m* =2 to *T*

for *n* =1 to *I*

if(n==1)

for $j = 1$ to n_K

row = $(m-1) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + j$

col = $(m-2) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + j$

Cap_RI[row,col]=1

end

else

for $j = 1$ to n_K

row = $(m-1) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + [\sum_{r=1}^{n-1} r_K] + j$

col = $(m-2) * (\widetilde{H} + \ddot{N} + I_K) + (\widetilde{H} + \ddot{N}) + [\sum_{r=1}^{n-1} r_K] + j$

Cap_RI[row,col]=1

end

end

end

end

Finally, construct two matrices **I_Mean** and **I_Error** of size $[T(\widetilde{H} + \ddot{N} + I_K) \times T(\widetilde{H} + \ddot{N} + I_K)]$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **I_Mean** and **I_Error**.

for *m* =2 to *T*

for *n* =1 to $\widetilde{H}$

if(m==1)

row = $(m-1)*\left(\widetilde{H}+\ddot{N}+I_K\right)+n$

col = $(m-2)*\left(\widetilde{H}+\ddot{N}+I_K\right)+n$

I_Mean[row,col]= $\widetilde{\boldsymbol{\rho}}[\boldsymbol{i},\mathbf{1}]$

I_Error[row,col]= $\widetilde{\boldsymbol{\rho}}[\boldsymbol{i},\mathbf{1}]$

end

end

for *m* =2 to *T*

for *n* =1 to *I*

if(n==1)

for $j=1$ to n_K

row = $(m-1)*\left(\widetilde{H}+\ddot{N}+I_K\right)+(\widetilde{H}+\ddot{N})+\ j$

col = $(m-2)*\left(\widetilde{H}+\ddot{N}+I_K\right)+(\widetilde{H}+\ddot{N})+\ j$

I_Error[row,col]= $\boldsymbol{\lambda}[\boldsymbol{n},\mathbf{1}]$

end

else

for $j=1$ to n_K

row = $(m-1)*\left(\widetilde{H}+\ddot{N}+I_K\right)+(\widetilde{H}+\ddot{N})+[\sum_{r=1}^{n-1}r_K]+\ j$

col = $(m-2)*\left(\widetilde{H}+\ddot{N}+I_K\right)+(\widetilde{H}+\ddot{N})+[\sum_{r=1}^{n-1}r_K]+\ j$

I_Error[row,col]= $\boldsymbol{\lambda}[\boldsymbol{n},\mathbf{1}]$

end

end

end

end

Also, collect all the error-covariance matrices as follows:

$$\overleftrightarrow{\boldsymbol{\Sigma}}=\begin{bmatrix}\Xi & \mathrm{Cov}(\boldsymbol{\Xi};\boldsymbol{\Gamma})' & \mathrm{Cov}(\boldsymbol{\Xi};\boldsymbol{\Lambda})' \\ \mathrm{Cov}(\boldsymbol{\Xi};\boldsymbol{\Gamma}) & \boldsymbol{\Gamma} & \mathrm{Cov}(\boldsymbol{\Gamma};\boldsymbol{\Lambda})' \\ \mathrm{Cov}(\boldsymbol{\Xi};\boldsymbol{\Lambda}) & \mathrm{Cov}(\boldsymbol{\Gamma};\boldsymbol{\Lambda}) & \boldsymbol{\Lambda}\end{bmatrix}[(\widetilde{H}+\ddot{N}+I_K)\times(\widetilde{H}+\ddot{N}+I_K)]$$

where off-diagonal elements capture the dependence across different type of variables (continuous, ordered, and nominal variables).

With this, we can write the distribution of joint model as follows:

$$\boldsymbol{YU} \sim MVN(B_{T*(\widetilde{H}+\ddot{N}+I_K)}, \Theta_{T*(\widetilde{H}+\ddot{N}+I_K)\times T*(\widetilde{H}+\ddot{N}+I_K)}),$$

where $\boldsymbol{B} = \mathbf{F\_Mean} * \mathbf{sumc}\left[\left(\overleftrightarrow{\boldsymbol{\beta}}.*\boldsymbol{X}\right)'\right]$,

$$\boldsymbol{\Theta} = \mathbf{F\_Error} * \left[\mathbf{1}_{\boldsymbol{T}}.*.\left(\boldsymbol{D}_{\boldsymbol{MAT}} * \overleftrightarrow{\boldsymbol{\Sigma}}\right)\right] * \mathbf{F\_Error}',$$

$$\mathbf{F\_Mean} = [\mathbf{1}_{\boldsymbol{T}(\widetilde{\boldsymbol{H}}+\ddot{\boldsymbol{N}}+\boldsymbol{I_K})} - \mathbf{I\_Mean}.*.\mathbf{Cap\_RI})]^{-\mathbf{1}}, \text{and}$$

$$\mathbf{F\_Error} = [\mathbf{1}_{\boldsymbol{T}(\widetilde{\boldsymbol{H}}+\ddot{\boldsymbol{N}}+\boldsymbol{I_K})} - \mathbf{I\_Error}.*.\mathbf{Cap\_RI})]^{-\mathbf{1}}$$

Next, to estimate the model, we take the utility difference between the chosen alternative (i_{m_k}) and non-chosen alternatives for all the nominal variables. To perform utility difference, construct a matrix **M_mat** of size $[T(\widetilde{H} + \ddot{N} + I_K - I) \times T(\widetilde{H} + \ddot{N} + I_K)]$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **M_mat**.

For m=1 to T

M=zeros$((\widetilde{H} + \ddot{N} + I_K - I), (\widetilde{H} + \ddot{N} + I_K))$

M $[\mathbf{1}: \widetilde{H} + \ddot{N}), 1: (\widetilde{H} + \ddot{N})] = \mathbf{1}_{(\widetilde{H}+\ddot{N})}$

for n=1 to I

Iden_mat=$\mathbf{1}_{n_K-1}$

O_neg =-1*ones(n_K-1,1)

if($n_{m_K} == 1$)

temp_mat =O_neg ~ Iden_mat

else if($n_{m_K} == n_K$)

temp_mat=Iden_mat[.,1: n_{m_K}1]~O_neg~Iden_mat[., n_{m_K}: $n_K - 1$]

end

if(n==1)

row1 = $(\widetilde{H} + \ddot{N}) + 1$

row2 = $(\widetilde{H} + \ddot{N}) + n_K - 1$

col1 = $(\widetilde{H} + \ddot{N}) + 1$

col2 = $(\widetilde{H} + \ddot{N}) + n_K$

else

row1 = $\left(\widetilde{H} + \ddot{N}\right) + (\sum_{j=1}^{n-1}(j_K - 1)) + 1$

row2 = $\left(\widetilde{H} + \ddot{N}\right) + (\sum_{j=1}^{n}(j_K - 1)) + 1$

col1 = $\left(\widetilde{H} + \ddot{N}\right) + (\sum_{j=1}^{n-1}(j_K)) + 1$

col2 = $\left(\widetilde{H} + \ddot{N}\right) + (\sum_{j=1}^{n-1}(j_K))+n_K$

end

M[row1:row2,col1:col2]=temp_mat

end

s_row1 = $(m-1) * \left(\widetilde{H} + \ddot{N} + I_K - I\right) + 1$

s_row2 = $(m) * \left(\widetilde{H} + \ddot{N} + I_K - I\right)$

s_col1 = $(m-1) * \left(\widetilde{H} + \ddot{N} + I_K\right) + 1$

s_col2 = $(m) * \left(\widetilde{H} + \ddot{N} + I_K\right)$

M_mat[s_row1:s_row2,s_col1:s_col2]=**M**

end

where "~" refers to horizontal concatenation.

With this we may write the distribution of $\overline{\boldsymbol{Y}}\overline{\boldsymbol{U}}$ (same as ***YU*** but with utility difference w.r.t the chosen alternative for all the nominal variables) as $\bar{Y}\bar{U} \sim MVN_{T*(\widetilde{H}+\ddot{N}+I_K-I)}\left(\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{\Theta}}\right)$ where $\widetilde{\boldsymbol{B}}$ =**M_mat*B, and** $\widetilde{\boldsymbol{\Theta}} = \mathbf{M\_mat} * \widetilde{\boldsymbol{\Theta}} * \mathbf{M\_mat}'$.

Next, we define a matrix to re-arrange the elements of mean and covariance matrix of $\overline{\boldsymbol{Y}}\overline{\boldsymbol{U}}$ in the following order: continuous, ordered, and nominal. This makes it easy to find the conditional distribution of non-continuous variables in a matrix format. To do so, define a matrix **R_mat** of size $[T\left(\widetilde{H} + \ddot{N} + I_K - I\right) \times T\left(\widetilde{H} + \ddot{N} + I_K - I\right)]$ with all the cells filled with zeros. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **R_mat.**

---------------For continuous variables---

For *m* =1 to *T*

row1=(*m*-1) * $\widetilde{H}$+1

row2=(*m*) *$\widetilde{H}$

col1=(*m*-1) * $\left(\widetilde{H} + \ddot{N} + I_K - I\right) + 1$

col2=(*m*-1) * $\left(\widetilde{H} + \ddot{N} + I_K - I\right) + \widetilde{H}$

R_mat[row1: row2, col1:col2]= $1_{\widetilde{H}}$

end

-----------------For ordered variables---

For m =1 to T

row1=$\widetilde{H}T$+ (m-1) * $\ddot{N}$+1

row2=$\widetilde{H}T$+ (m) * $\ddot{N}$+1

col1=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\widetilde{H}+1$

col2=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\widetilde{H}+\ddot{N}$

R_mat[row1: row2, col1:col2]= $\mathbf{1}_{\ddot{N}}$

end

----------------------For nominal variables--

For m =1 to T

row1=$(\widetilde{H}+\ddot{N})T$ + (m-1) * (I_K-I) +1

row2=$(\widetilde{H}+\ddot{N})T$ + (m) * (I_K-I) +1

col1=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\left(\widetilde{H}+\ddot{N}\right)+1$

col2=(m-1) * $\left(\widetilde{H}+\ddot{N}+I_K-I\right)+\left(\widetilde{H}+\ddot{N}\right)+(I_K-I)$

R_mat[row1: row2, col1:col2]= $1_{(I_K-I)}$

end

With this, we may write:

$$\overline{\boldsymbol{Y}}\overline{\boldsymbol{U}}\sim MVN_{T*(\widetilde{H}+\ddot{N}+I_K-I)}\left(\ddot{\boldsymbol{B}},\ddot{\boldsymbol{\Theta}}\right)$$

where $\widetilde{\boldsymbol{B}}$ =**R_mat***$\widetilde{\boldsymbol{B}}$, and $\overleftrightarrow{\boldsymbol{\Theta}}=\mathbf{R\_mat}*\widetilde{\boldsymbol{\Theta}}*\mathbf{R\_mat}'$.

Next, to account for un-balanced panel data structure, we define a matrix **RM_mat** of size $[\tilde{T}\widetilde{H}+\ddot{T}\ddot{N}+T(I_K-I)\times T(\widetilde{H}+\ddot{N}+I_K-I)]$ with all the cells filled with zeros. It will allow us to collect the relevant elements from the vector $\overleftrightarrow{B}$ and matrix $\overleftrightarrow{\Theta}$. Now, follow the pseudo-code provided below to fill-up the cells of the matrix **RM_mat**.

--------------For continuous variables---

For m =1 to $\tilde{T}$

row1=(m-1) * $\widetilde{H}$+1

row2=(m) *$\widetilde{H}$

col1=(m-1) * $\widetilde{H}+1$

col2=(m) * $\widetilde{H}$

R_mat[row1: row2, col1:col2]= $\mathbf{1}_{\widetilde{H}}$

end

-----------------For ordered variables---

For *m* =1 to $\ddot{T}$

row1=$\widetilde{H}\widetilde{T}$+ (*m*-1) * $\ddot{N}$+1

row2=$\widetilde{H}\widetilde{T}$+ (*m*) * $\ddot{N}$

col1=$\widetilde{H}T$ + (*m*-1) * $\ddot{N} + 1$

col2=$\widetilde{H}T$ + (*m*) * $\ddot{N}$

R_mat[row1: row2, col1:col2]= $1_{\ddot{N}}$

end

----------------------For nominal variables---

For *m* =1 to *T*

row1=$\widetilde{H}\widetilde{T} + \ddot{N}\ddot{T}$ + (*m*-1) * $(I_K - I)$ +1

row2=$\widetilde{H}\widetilde{T} + \ddot{N}\ddot{T}$ + (*m*) * $(I_K - I)$ +1

col1= $\left(\widetilde{H} + \ddot{N}\right)T + (m-1) * \left(I_K - I\right) + 1$

col2=$\left(\widetilde{H} + \ddot{N}\right)T + (m) * \left(I_K - I\right)$

R_mat[row1: row2, col1:col2]= $\mathbf{1}_{(I_K - I)}$

end

Now we may write:

$$\overline{\boldsymbol{Y}}\overline{\boldsymbol{U}} \sim MVN_{\widetilde{H}\widetilde{T}+\ddot{N}\ddot{T}+T(I_K-I)}\left(\overleftrightarrow{\boldsymbol{B}}, \overleftrightarrow{\boldsymbol{\Theta}}\right)$$

where $\overleftrightarrow{\boldsymbol{B}} =$ **RM_mat***$\overleftrightarrow{\boldsymbol{B}}$, and $\overleftrightarrow{\boldsymbol{\Theta}} = \mathbf{RM\_mat} * \overleftrightarrow{\boldsymbol{\Theta}} * \mathbf{RM\_mat}'$.

Next, partition the $\overleftrightarrow{B}$ and $\overleftrightarrow{\boldsymbol{\Theta}}$ into the continuous and non-continuous variables as follows:

$$\overleftrightarrow{B} = \begin{bmatrix} \overleftrightarrow{B}_{\widetilde{H}} \\ \overleftrightarrow{B}_{\ddot{N}\overline{U}} \end{bmatrix} \begin{bmatrix} \widetilde{H}\widetilde{T} \times 1 \\ \ddot{N}\ddot{T} + T(I_K - I) \times 1 \end{bmatrix}, \text{and } \overleftrightarrow{\boldsymbol{\Theta}} = \begin{bmatrix} \overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H}} & \overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H},\ddot{N}\overline{U}} \\ \overleftrightarrow{\boldsymbol{\Theta}}'_{\widetilde{H},\ddot{N}\overline{U}} & \overleftrightarrow{\boldsymbol{\Theta}}_{\ddot{N}\overline{U}} \end{bmatrix}.$$

With this, the conditional distribution of non-continuous variables can be written as:

$\overline{\boldsymbol{B}}_{\ddot{N}\overline{U}}$ = $\overleftrightarrow{\boldsymbol{B}}_{\ddot{N}\overline{U}} + \overleftrightarrow{\boldsymbol{\Theta}}'_{\widetilde{H},\ddot{N}\overline{U}}(\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H}})^{-1}(\widetilde{\boldsymbol{y}}\ [\mathbf{1}{:}\ \widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{T}}] - \overleftrightarrow{\boldsymbol{B}}_{\widetilde{H}})$,

$\overline{\boldsymbol{\Theta}}_{\ddot{N}\overline{U}} = \overleftrightarrow{\boldsymbol{\Theta}}_{\ddot{N}\overline{U}} - \overleftrightarrow{\boldsymbol{\Theta}}'_{\widetilde{H},\ddot{N}\overline{U}}(\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H}})^{-1}\overleftrightarrow{\boldsymbol{\Theta}}_{\widetilde{H},\ddot{N}\overline{U}}$.

Also, append the threshold vectors as follows:

$\overline{\Psi}_{low}$=$[(\Psi_{low}[1{:}\ \ddot{N}\ddot{T}])', (-\infty_{T(I_K-I)})'] \left[\left(\ddot{N}\ddot{T} + T(I_K - I)\right) \times 1\right]$ vector, and,

$\overline{\Psi}_{up}$=$[(\Psi_{up}[1:\ddot{N}\ddot{T}])',(0_{T(I_K-I)})']'$ $\left[\left(\ddot{N}\ddot{T}+T(I_K-I)\right)\times 1\right]$ vector. Where $-\infty_{T(I_K-I)})$ and $0_{T(I_K-I)}$ are column vectors of size $T(I_K-I)$with all the cells filled with a value of "-∞" and "0" respectively.

Then the likelihood function may be written as:

$$L(\theta)=f_{\widetilde{H}\widetilde{T}}(\tilde{y}\,[1:\widetilde{H}\widetilde{T}]\overleftrightarrow{B}_{\widetilde{H}},\overleftrightarrow{\mathbf{\Theta}}_{\widetilde{\boldsymbol{H}}})\times\int_{\overline{\Psi}_{low}}^{\overline{\Psi}_{up}}f_{\ddot{N}\ddot{T}+T(I_K-I)}(\boldsymbol{r}|\,\overline{\boldsymbol{B}}_{\ddot{\boldsymbol{N}}\overline{\boldsymbol{U}}},\overleftrightarrow{\mathbf{\Theta}}_{\ddot{\boldsymbol{N}}\overline{\boldsymbol{U}}})dr \quad (9)$$

where $\boldsymbol{\theta}=[\boldsymbol{\gamma}',\boldsymbol{\delta}',\boldsymbol{\beta}',\boldsymbol{\rho}',\boldsymbol{\lambda}',(\boldsymbol{Vech}(\overleftrightarrow{\mathbf{\Sigma}}))']$ and Vech (.) operator vectorizes the unique element of a matrix.

The likelihood function involves computation of a $\widetilde{H}\widetilde{T}$ dimensional multi-variate normal probability density (MVNPD) function and $\ddot{N}\ddot{T}+T(I_K-I)$ dimensional multi-variate normal cumulative density (MVNCD) function. While the MVNPD function has a closed form expression, increase in dimensionality can lead to calculation of numerical value very close to zero and thus causing issues during estimation[11]. On the other hand, the computation of a MVNCD function is a well-known challenge in the literature (Genz 1992; Heiss 2010; Connors et al. 2014). Even the powerful GHK simulator armed with sophisticated quasi-random sequences can calculate the value accurately only up to a limited number of dimensions (Sándor & András 2004)[12]. At the same time, it is well known and established that any simulation-based method loses its accuracy with increases in dimension due to simulation noise, not to mention the unreasonable computation time (Train 2000; Bhat 2003; Craig 2008). For example: the analysis section of the paper has 8 continuous variables with 20 time periods, 5 ordinal variables with 1 time period, and 1 nominal variable with 8 alternatives and 20 choice occasions. In the maximum likelihood (ML) approach, this translates to a computation of a 160 dimensional MVNPD function and a 145 dimensional MVNCD function. Therefore it may be quite challenging to solve equation (9) using ML approach.

While one can use Bayesian approach to solve such a complicated likelihood function involving a series of draws from conditional distribution, a review of literature involving Probit kernel shows that Bayesian approach has not performed as expected in terms of recovering parameters and their standard errors (Franzese et al. 2010; Patil et al. 2017). On the other hand, there have been few studies (Daziano 2015; Zhou et al. 2016) which have found the performance of Bayesian approach to be quite good. However, these studies did not compare the performance of Bayesian approach against ML or other approaches. This is not to say that the Bayesian approach may not work. A comprehensive evaluation of the present model using Bayesian approach is outside the scope of the paper and we leave this for future explorations. Therefore, we use a composite marginal likelihood (CML) approach which has been established in the last decade as one of the powerful approach for solving likelihood functions with high dimensional integrals. A comprehensive discussion on the CML approach is outside the scope of this paper and readers are refer to Varin & Vidoni 2005; Varin 2008; Varin, Reid & Firth 2011 for a detailed discussion on CML and see Bhat & Dubey (2014) for its application in the context of discrete choice models. Further Bhat and colleagues have performed

[11] Consider a situation where there are 20 continuous dependent variables. Now, estimate a uni-variate regression for each of the 20 continuous variables which may include parameters apart from a constant. Now, if one wish to estimate a joint model for all the 20 continuous variables, even with a good starting value (obtained from uni-variate regression), the MVNPD value may be very close to zero (numerically).

[12] The assumption is that the number of draws are finite (less than 1000) to maintain reasonable estimation time.

extensive simulation using CML approach for complex econometric models and have observed highly accurate results (Paleti & Bhat 2013; Bhat & Dubey 2014; Bhat 2015; Bhat et al. 2016).

Composite Marginal Likelihood Approach

The likelihood function can be written as follows using the CML approach:

$$L_{CML}(\theta) = \left(\prod_{h=1}^{\widetilde{H}\widetilde{T}-1} \prod_{h'=h+1}^{\widetilde{H}\widetilde{T}} f_2(\tilde{y}_{hh'} | \overleftrightarrow{B}_{hh'}, \overleftrightarrow{\Theta}_{hh'}) \right) \times$$

$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}-1} \prod_{r'=r+1}^{\ddot{N}\ddot{T}} \Pr(\ddot{y}_n = a_n, \ddot{y}_{n'} = a_{n'}) \right) \times$$

$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}} \prod_{t=1}^{T} \prod_{r'=1}^{I} \Pr\left(\ddot{y}_n = a_n, i_{r'\,t} = i_{r'_{m k'_t}}\right) \right) \times$$

$$\left(\prod_{r=1}^{TI-1} \prod_{r'=r+1}^{TI} \Pr(i_r = i_{r_{m_k}}, i_{r'} = i_{r'_{m_k}}) \right) \qquad (10)$$

In the above CML expression, the first expression corresponds to the pairing of two continuous variables at a time reducing the dimension of MVNPD function from $\widetilde{H}\widetilde{T}$ to a maximum of 2 avoiding any numerical issues in calculation of MVNPD function due to high dimensionality[13]. The second expression corresponds to the pairing of two ordinal variables reducing the dimensionality of integration to 2 from $\ddot{N}\ddot{T}$. The third expression corresponds to the pairing between an ordinal and nominal variable with a maximum dimensionality of integration equal to max($i_K \forall I$). Finally the fourth expression corresponds to the pairing between nominal variables with a highest dimensionality of integration being equal to 2*max($i_K \forall I$).

To explicitly write out the equation (10) in terms of MVNPD and MVNCD functions, we define a set of selection matrices: (1) construct a selection matrix **D_HH** of size [2× $\widetilde{H}\widetilde{T}$] with all the cells filled with zeros. Now, place a value of '1' in 1st row and *h*th column and in 2nd row and *h'* th column. This matrix is designed to collect relevant elements for pairing between continuous variables within and across time-periods, (2) define a selection matrix **D_NI** of size $[i_K \times (\ddot{N}\ddot{T} + (I_K - I)T]$ with all the cells filled with zeros. This matrix is designed to collect relevant elements for pairing between ordered and nominal variables. Now, place a value of '1' in the 1st row and r^{th} column. Next if $r' = 1$, then place an identity matrix of size $r'_K - 1$ in the rows 2 to r'_K and columns $\ddot{N}\ddot{T} + (t-1)(I_K - I) + 1$ to $\ddot{N}\ddot{T} + (t-1)(I_K - I) + r'_K - 1$, otherwise place an identity matrix of size $r'_K - 1$ in the rows 2 to r'_K and columns $\ddot{N}\ddot{T} + (t-1)(I_K - I) + \left(\sum_{j=1}^{r'-1}(j_K - 1)\right)$ +1 to $\ddot{N}\ddot{T} + (t-1)(I_K - I) + \left(\sum_{j=1}^{r'-1}(j_K - 1)\right)$, and (3) define two variables as follows: alt_1 = $r - (\text{ceil}(r/1) - 1) * I$ and alt_2 = $r' - (\text{ceil}(r/I) - 1) * I$. Where **ceil(.)** operator rounds the value in parenthesis to next largest integer. Now, construct a selection matrix **D_II** of size $[(r_{alt_1} + r'_{alt_2} - 2) \times (\ddot{N}\ddot{T} + (I_K - I))\text{T}]$ with all the cells filled with zeros. This matrix is designed to collect relevant

[13] For all the pairings, different continuous variables in the same time-period and all continuous variables across time-periods are used. This also holds for all pairing between ordinal and ordinal, and nominal and nominal variables.

elements for pairing between nominal variables within and across time-periods. Now, use the pseudocode provided below to fill-up the cells of **D_II** matrix.

if (alt_1 == 1)

row1=1

row2=$r_{alt_1} - 1$

col1 =$\ddot{N}\ddot{T} + (\text{ceil}(r/1) - 1) * (I_K - I) + I$

col2 =$\ddot{N}\ddot{T} + (\text{ceil}(r/1) - 1) * (I_K - I) + r_{alt_1} - 1$

else

row1=1

row2=$r_{alt_1} - 1$

col1 =$\ddot{N}\ddot{T} + (\text{ceil}(r/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt1-1}(j_K - 1)\right) + 1$

col2 =$\ddot{N}\ddot{T} + (\text{ceil}(r/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt1}(j_K - 1)\right)$

end

D_II[row1:row2,col1:col2]=$1_{i_{alt\_1}-1}$

if(alt_2==1)

row1=r_{alt_1}

row2=$r_{alt\_1} \quad + r'_{alt\_2} - 2$

col1 =$\ddot{N}\ddot{T} + (\text{ceil}(r'/1) - 1) * (I_K - I) + 1$

col2 =$\ddot{N}\ddot{T} + (\text{ceil}(r'/1) - 1) * (I_K - I) + r'_{alt\_2} \quad - 1$

else

row1=r_{alt_1}

row2=$r_{alt_1} + r'_{alt_2} - 2$

col1 =$\ddot{N}\ddot{T} + (\text{ceil}(r'/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt\_2-1}(j_K - 1)\right) + 1$

col2 =$\ddot{N}\ddot{T} + (\text{ceil}(r'/1) - 1) * (I_K - I) + \left(\sum_{j=1}^{alt\_2}(j_K - 1)\right)$

end

D_II[row1:row2,col1:col2] $1_{i_{alt\_2}-1}$

With the selection matrices defined, now we define the appropriate mean vector and covariance matrix for pairing of dependent variables. Define the following vectors and matrices:

$$\widehat{\boldsymbol{B}}_{\boldsymbol{hh'}} = \mathbf{D\_HH} * \overleftrightarrow{\boldsymbol{B}}_{\boldsymbol{hh'}}, \widehat{\boldsymbol{\Theta}}_{\boldsymbol{hh'}} = \mathbf{D\_HH} * \overleftrightarrow{\boldsymbol{\Theta}}_{\boldsymbol{hh'}}, * \mathbf{D\_HH'},\ \widehat{\boldsymbol{y}}_{\boldsymbol{hh'}} = \mathbf{D\_HH} * \widetilde{\boldsymbol{y}}_{\boldsymbol{hh'}},$$

$$v_{r,low}=\frac{[\Psi_{low}]_r-[\overline{\mathrm{B}}_{\ddot{N}\bar{U}}]_r}{\sqrt{[\overline{\Theta}_{\ddot{N}\bar{U}}]_{rr}}},v_{r,up}=\frac{[\Psi_{up}]_r-[\overline{\mathrm{B}}_{\ddot{N}\bar{U}}]_r}{\sqrt{[\overline{\Theta}_{\ddot{N}\bar{U}}]_{rr}}},v_{r',low}=\frac{[\Psi_{low}]_{r'}-[\overline{\mathrm{B}}_{\ddot{N}\bar{U}}]_{r'}}{\sqrt{[\overline{\Theta}_{\ddot{N}\bar{U}}]_{r'r'}}},$$

$$v_{r',up}=\frac{[\Psi_{up}]_{r'}-[\overline{\mathrm{B}}_{\ddot{N}\bar{U}}]_{r'}}{\sqrt{[\overline{\Theta}_{\ddot{N}\bar{U}}]_{r'r'}}},\vartheta_{rr}=\frac{[\overline{\Theta}_{\ddot{N}\bar{U}}]_{rr'}}{\sqrt{[\overline{\Theta}_{\ddot{N}\bar{U}}]_{rr}}\sqrt{[\overline{\Theta}_{\ddot{N}\bar{U}}]_{r'r'}}},\hat{B}_{rr'}=\mathrm{D\_NI}*\bar{B}_{\ddot{N}\bar{U}},$$

$$\widehat{\Theta}_{rr'}=\mathrm{D\_NI}*\overline{\Theta}_{\ddot{N}\bar{U}},*\,\mathrm{D\_NI}',\widehat{\Psi}_{rr',low}=\mathrm{D\_NI}*\Psi_{low},$$

$$\widehat{\Psi}_{rr',low}\left[2{:}\,rows\left(\widehat{\Psi}_{rr',low}\right)\right]=\mathrm{zeros}(\mathrm{rows}\left(\widehat{\Psi}_{rr',low}\right),1),$$

$$\widehat{\Psi}_{rr',up}=\mathbf{D\_NI}*\Psi_{up},\breve{B}_{rr'}=\mathbf{D\_II}*\bar{B}_{\ddot{N}\bar{U}},\text{and }\breve{\Theta}_{rr'}=\mathbf{D\_II}*\overline{\Theta}_{\ddot{N}\bar{U}},*\,\mathbf{D\_II}'$$

With the help of above defined notations, we may write the equation (10) in the explicit form as follows:

$$L_{CML}(\theta)=\left(\prod_{h=1}^{\widetilde{H}\widetilde{T}-1}\prod_{h'=h+1}^{\widetilde{H}\widetilde{T}}\phi_2\left(\hat{y}_{hh'}|\hat{B}_{hh'},\widehat{\Theta}_{hh'}\right)\right)\times$$
$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}-1}\prod_{r'=r+1}^{\ddot{N}\ddot{T}}\begin{bmatrix}\Phi_2\left(v_{r,up},v_{r',up},\vartheta_{rr'}\right)-\Phi_2\left(v_{r,up},v_{r',low},\vartheta_{rr'}\right)\\-\Phi_2\left(v_{r,low},v_{r',up},\vartheta_{rr'}\right)+\Phi_2\left(v_{r,low},v_{r',low},\vartheta_{rr'}\right)\end{bmatrix}\right)\times$$
$$\left(\prod_{r=1}^{\ddot{N}\ddot{T}}\prod_{t=1}^{T}\prod_{r'=1}^{I}[\Phi_{r_K'}\left[\left(\hat{\psi}_{rr',up}-\hat{B}_{rr'}\right);\widehat{\Theta}_{rr'}\right]-\Phi_{r_K'}\left[\left(\hat{\psi}_{rr',low}-\hat{B}_{rr'}\right);\widehat{\Theta}_{rr'}\right]]\right)\times$$
$$\left(\prod_{r=1}^{TI-1}\prod_{r'=r+1}^{TI}[\Phi_{r_K+r_K'-2}\left[\breve{B}_{rr'};\breve{\Theta}_{rr'}\right]]\right)\qquad(11)$$

where $\phi_r(.)$ and $\Phi_r(.)$ represents a MVNPD and MVNCD function of dimension *r*, respectively. The parameters Θ are obtained by maximizing the $\log[L_{CML}(\theta)]$. Further, unlike the ML approach, in the CML approach, the equivalence between the inverse of Hessian matrix $H(\theta)\left[-\frac{\partial^2 L_{CML}(\theta)}{\partial\theta*\partial\theta'}\right]^{-1}$ and the information matrix $I(\theta)\left[\left(\frac{\partial L_{CML}(\theta)}{\partial\theta}\right)\times\left(\frac{\partial L_{CML}(\theta)}{\partial\theta}\right)'\right]$does not exist and therefore the standard errors are calculated using the inverse of sandwich matrix $G(\theta)^{-1}=H(\theta)^{-1}I(\theta)\,\mathrm{H}(\theta)^{-1}$. Now that the dimension of MVNCD function has been reduced to a computationally acceptable range, one may use the Geweke- Hajivassiliou-Keane (GHK) simulator (Hajivassiliou et al. 1996) with quasi-random sequences or Bhat's Maximum Approximate Composite Marginal Likelihood (MACML) approach (Bhat 2011). While the GHK simulator is a simulation based estimator, the MACML is an analytic approximation and thus is computationally faster than the GHK. However, based on extensive testing of both methods, we have found that the MACML approach is a good method up to a dimension of 8-10. Its performance starts to degrade rather quickly beyond a dimension of 12 in comparison with the GHK simulator[14]. In our empirical analysis, the highest dimension of integral is 14 and thus based on equation (11) we use GHK simulator with 200 Halton-draws for the estimation of MVNCD function[15]. Finally, since the standard errors are calculated using sandwich estimator, one will need to calculate the Hessian matrix numerically or analytically. However, unlike logit kernel where the Hessian matrix can be computed numerically using central difference method, the same is not true for the Probit kernel due to relatively slow rate of change of MVNCD function in comparison

[14] The simulation design corresponds to a multinomial Probit model estimation for both cross-section and panel data structure with 5 alternatives, 5 choice occasions, and 5 random parameters with full and no cross-correlation.

[15] In our simulation experiments, we found that the 200 Halton draws are sufficient up to 20 dimensions.

to the exponential function[16]. To avoid any such issues, we analytically calculated the first and second order derivatives of the CML function involving MVNCD function.

Positive Definiteness of Covariance Matrices

To maintain the positive definiteness of the error covariance and random taste parameter covariance matrices, we work with the Cholesky decomposition of the matrices during estimation. i.e., if we are working with the full joint model, we pass the lower triangular Cholesky decomposition of the matrix $\overleftrightarrow{\boldsymbol{\Sigma}}$. Also, since the error covariance matrix for ordered variables are restricted to be a correlation matrix along with the first row of each of the nominal variables, we need to ensure that the during estimation, proper restrictions are maintained. Therefore, for all the rows of the matrix $\overleftrightarrow{\boldsymbol{\Sigma}}$ where the diagonal element is constrained to be 1, parametrize such rows of the lower triangular Cholesky decomposition of matrix $\overleftrightarrow{\boldsymbol{\Sigma}}$ as follows:

Let $\boldsymbol{LL}' = \ddot{\boldsymbol{\Sigma}}$**,** where ***L*** is the lower triangular Cholesky matrix. Then, for each of the required rows do the following: Let $a_i = [1 + L[i, 1{:}\, i-1].\,\hat{}\,2]^2$ where *i* refers to the row number and the operator ".^" refers to element by element exponentiation. Then parametrize all non-diagonal elements of the *i*th row as $L[i,r] = \frac{L[i,r]}{a_i}\,\forall r = 1 \text{ } to \text{ } i-1 \text{ and the diagonal element as } L[i,i] = \frac{1}{a_i}$.

The same strategy can be used if one wishes to estimate the models independently. In this case just work with Cholesky decomposition of matrices **Ξ, Ω, ψ, Γ, Λ** and **Σ**.

The above described model treats the visual attention data as a means to drive the preferences. The continuous model component of the system models the visual attention in terms of time spent on various alternatives, including its labels, which is then used as an explanatory variable in the choice model component). On the other hand, to test the hypothesis that habits, goals, and constraints work as a screening mechanism, we use the visual attention as an explanatory variable in the choice model but passed as a penalty. That is, we add a penalty term to the utility equation on each alternative which may be a function of individuals' habits and time-spent on alternatives.

$$U_{alt} = V_{alt} + \ln\left[{}^{1}\!/_{1} + \exp(\mu_{alt})\right] + \xi_{alt}$$

Where U_{alt} is the utility of the alternative, V_{alt} is the deterministic component of the utility, ξ_{alt} is the normally distributed error term, and μ_{alt} is the penalty function. Further μ_{alt} = $\mu_{alt} = f(\text{individuals'havits, time spent on the alternative})$. The first parametrization $\left[{}^{1}\!/_{1} + \exp(\mu_{alt})\right]$ ensures that the value in the square bracket is bounded between 0 and 1 so that the natural logarithm of the function is bounded between -**∞** and 0. This way, an alternative becomes unavailable or gets pushed out from the consideration set as soon as the expression $ln\left[{}^{1}\!/_{1} + \exp(\mu_{alt})\right]$ takes a value of -**∞**. Please note that there is no stochastic component in the penalty function. Adding the stochastic component creates additional computational challenges in the realm of Probit kernel.

[16] Most software (except "R" software) fails to calculate the Hessian matrix for the models built on Probit kernel. The "R" software uses Richardson extrapolation method for calculating the Hessian matrix which ensures the computation of a positive definite Hessian matrix, but its accuracy is low in most of the cases.

References

Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological*, *37*(9), 837-855.

Bhat, C. R. (2011). The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, *45*(7), 923-939.

Bhat, C. R. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B: Methodological*, *79*, 50-77.

Bhat, C. R., & Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B: Methodological*, *67*, 68-85.

Bhat, C. R., Pinjari, A. R., Dubey, S. K., & Hamdi, A. S. (2016). On accommodating spatial interactions in a generalized heterogeneous data model (GHDM) of mixed types of dependent variables. *Transportation Research Part B: Methodological*, *94*, 240-263.

Connors, R. D., Hess, S., & Daly, A. (2014). Analytic approximations for computing probit choice probabilities. *Transportmetrica A: Transport Science*, *10*(2), 119-139.

Craig, P. (2008). A new reconstruction of multivariate normal orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 227-243.

Daziano, R. A. (2015). Inference on mode preferences, vehicle purchases, and the energy paradox using a Bayesian structural choice model. *Transportation Research Part B: Methodological*, *76*, 1-26.

Franzese, R. J., Hays, J. C., & Schaffer, L. M. (2010). Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. *APSA 2010 Annual Meeting* https://ssrn.com/abstract=1643867

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, *1*(2), 141-149.

Hajivassiliou, V., McFadden, D., & Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics*, *72*(1), 85-134.

Heiss, F. (2010). The panel probit model: adaptive integration on sparse grids. In Greene, W., & Hill, R. C. (Eds.), *Maximum simulated likelihood methods and applications* (pp. 41-64). Bingley, UK: Emerald Group Publishing Limited.

Keane, M. P. (1992). A note on identification in the multinomial probit model. *Journal of Business & Economic Statistics*, *10*(2), 193-200.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, *4*(1), 103-120.

Paleti, R., & Bhat, C. R. (2013). The composite marginal likelihood (CML) estimation of panel ordered-response models. *Journal of Choice Modelling*, *7*, 24-43.

Patil, P. N., Dubey, S. K., Pinjari, A. R., Cherchi, E., Daziano, R., & Bhat, C. R. (2017). Simulation evaluation of emerging estimation techniques for multinomial probit models. *Journal of Choice Modelling*, *23*, 9-20.

Sándor, Z., & András, P. (2004). Alternative sampling methods for estimating multivariate normal probabilities. *Journal of Econometrics*, *120*(2), 207-234.

Sidharthan, R., & Bhat, C. R. (2012). Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis*, *44*(4), 321-349.

Train, K. (2000). *Halton sequences for mixed logit*: UC Berkeley: Department of Economics.

Varin, C. (2008). On composite marginal likelihoods. *AStA- Advances in Statistical Analysis*, *92*(1), 1-28.

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*, 5-42.

Varin, C., & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, *92*(3), 519-528.

Zhou, Y., Wang, X., & Holguín-Veras, J. (2016). Discrete choice with spatial correlation: A spatial autoregressive binary probit model with endogenous weight matrix (SARBP-EWM). *Transportation Research Part B: Methodological*, *94*, 440-455.

WEB APPENDIX C: PARTICIPANT FLOW DIAGRAM

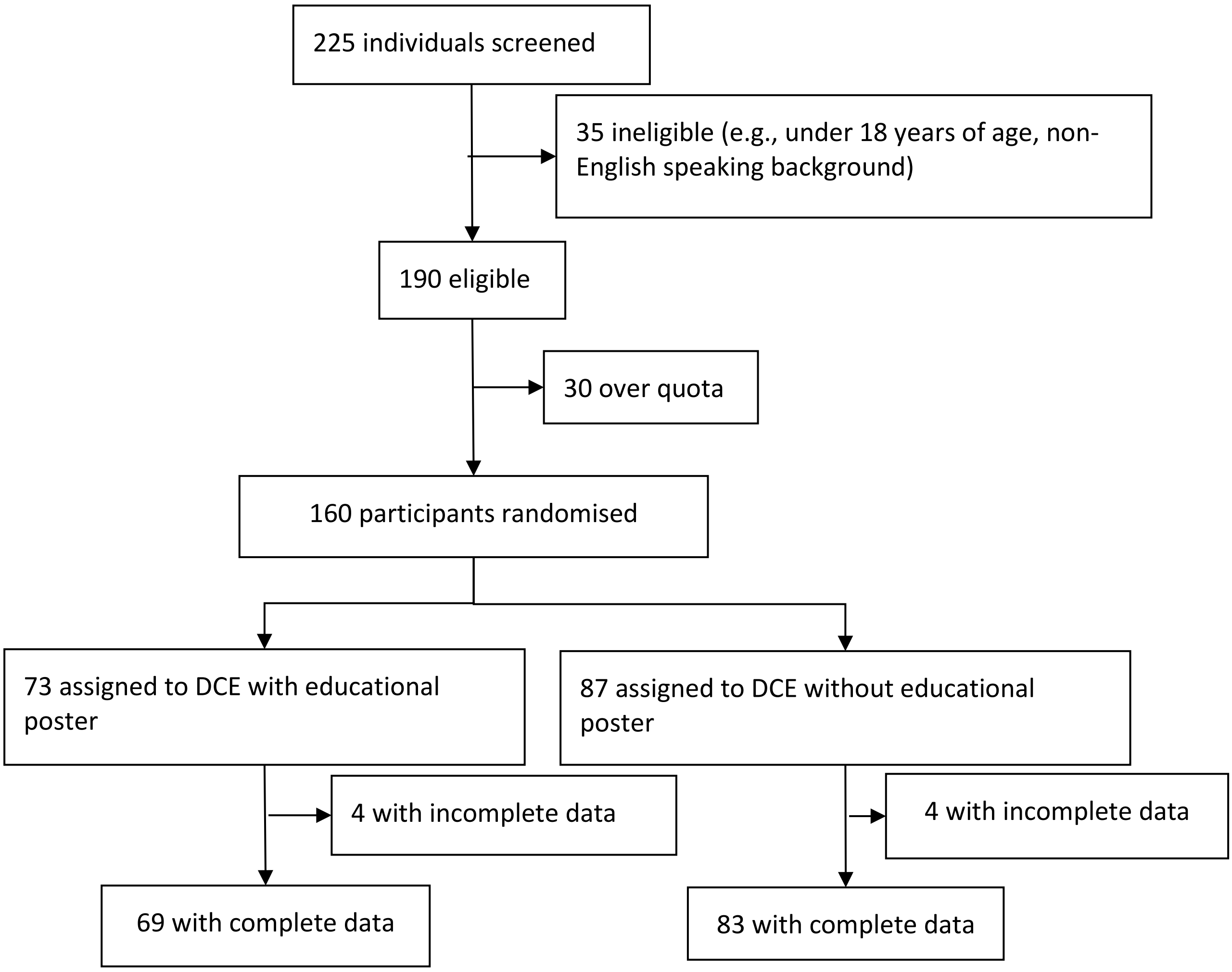


Fig. C.1. Flow chart of participants included/ excluded from eye-tracked discrete choice experiment (DCE)

Eye-tracking data were not captured for eight participants, due to technical errors, and therefore were excluded from this analysis. Of the remaining 152 participants, eye-tracking data were detected for some but not all of the 20 choice tasks for 13 individuals. These individuals were therefore excluded from the main analysis but used to test out-of-sample predictive power.

WEB APPENDIX D: DEMOGRAPHIC CHARACTERISTICS OF EYE-TRACKING STUDY PARTICIPANTS

Table D.1: Demographic characteristics of eye-tracking study participants in main estimation (n=139)

Characteristic	n (%)	Australian population
Females [a]		
18-35 years	41 (51.3%)	32.2%
36-59 years	32 (40.0%)	40.4%
60 years and over	7 (8.8%)	27.4%
Males [a]		
18-35 years	28 (47.5%)	34.0%
36-59 years	21 (35.6%)	40.7%
60 years and over	10 (17.0%)	25.3%
Equivalised household income quintile [b]		
Q1 (lowest income)	39 (28.1%)	20%
Q2	31 (22.3%)	20%
Q3	23 (16.6%)	20%
Q4	29 (20.9%)	20%
Q5 (highest income)	17 (12.2%)	20%
Highest educational attainment [c]		
Year 11 or below	4 (2.9%)	26%
Year 12 or equivalent	21 (15.1%)	18%
TAFE or Certificate, diploma	18 (13.0%)	21%
Undergraduate university	54 (38.9%)	29%
Postgraduate university	42 (30.2%)	6%
Body Mass Index (BMI) [d]		
<25kg/m^2 (normal or underweight)	78 (56.9%)	37.2%
25 to 30 kg/m^2 (overweight)	42 (30.7%)	35.3%
>30 kg/m^2 (obese)	17 (12.4%)	27.5%
SSB purchase frequency from convenience store in the past month [e]		
On about half of days or more	76 (34.7%)	-
A few times	56 (40.3%)	-
Never	7 (5.0%)	-

n=139 eye-tracking participants in main analysis (from total sample of 160). National statistics derived from: [a] Australian Bureau of Statistics (ABS) (2011), "Australian Demographic Statistics, Jun 2016, 'Table 1. Population Change, Summary - Australia ('000)', data cube: Excel spreadsheet, cat no. 3101.0," Available at: http://www.abs.gov.au/ ; [b] ABS (2013), "Household income and income distribution Australia." Available at: http://www.abs.gov.au/.[c] ABS (2016), "Education and Work, Australia, May 2016" Available at: http://www.abs.gov.au/; [d] BMI missing for 2 participants. ABS (2015)"National Health Survey: First Results, 2014-15, cat no. 4364.0.55.001," Available at http://www.abs.gov.au/. [e] 'Regular SSB (sugar-sweetened beverage) consumers' were defined as those who reported consumption of a SSB purchased from a convenience store at least a few times in the past month

WEB APPENDIX E: DESCRIPTIVE STATISTICS OF VISUAL ATTENDANCE

There was a significant correlation between fixation duration examining relevant choice set information with fixation duration out of choice set (R^2= 0.92, *p*-value <0.001). Based on this, the analyses below used ratio of time spent in and out of consideration set rather than absolute duration, unless otherwise specified, to avoid results being unduly influenced by overall time to complete the task. Where sample summaries are presented (rather than per choice set), this ratio is further adjusted for number of choice sets for which eye-tracking data was captured.

Linear regressions found that the first four choice tasks had a longer mean duration than the last four tasks, even when adjusted for age and gender (*p*-value <0.01), suggesting learning or fatigue. Ratio of relevant to irrelevant visual attention duration increased in the last compared to the first 4 tasks (*p*-value <0.01)

Stated Attendance

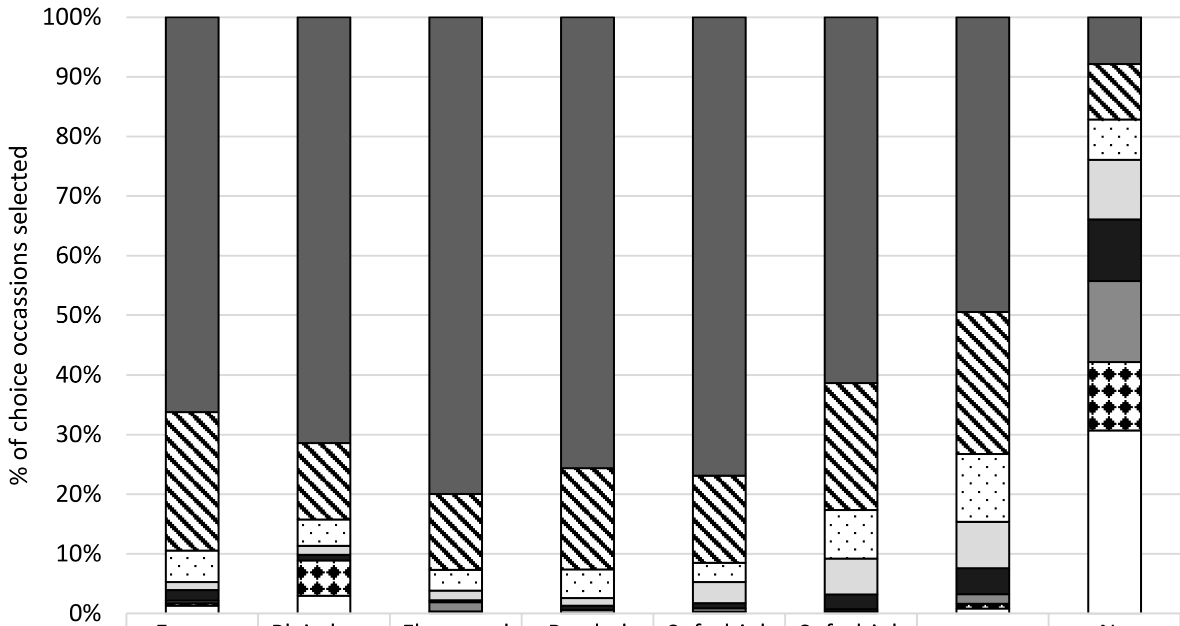
One hundred percent of respondents stated they sometimes or always considered price, and 95% and 99% stated they sometimes or always considered volume and beverage type, respectively. All beverage types were sometimes or always considered by more than 30% of the sample.

Relationship Between Stated and Visual Attendance

No significant difference was found in fixation duration on beverage, price or volume labels by stated importance on a 5-point Likert scale as per participants using an ANOVA (all *p*-values ≥ 0.34). Attribute and alternative fixation duration were not predicted by relevant stated attribute or alternative non-attendance using linear regression (all *p*-values >0.05). A higher score on strength of habit questionnaire (stronger SSB consumption habit) was positively related to fixation duration on energy drinks (*p*-value=0.06) and flavored milk (*p*-value=0.03), and negatively related to fixation duration on “no drink” alternative (*p*-value=0.01) using linear regression when adjusted for age and gender. This suggests that SSB consumption habit may be related to visual attention, but this unadjusted analysis was unable to distinguish the direction of effect. No significant relationships were seen between stage of readiness to drink fewer SSBs and fixation duration by beverage type or overall time on choice task.

Relationship between visual attention and choice

Respondents spent less visual fixation time on the chosen alternative across choice tasks compared to other alternatives. **Fig. E.1** shows a detailed breakdown of visual attention time spent on chosen alternatives. On more than 50% of occasions, the chosen alternative received the least amount of visual attention.



	Energy drink	Plain low-fat milk	Flavoured milk	Bottled water	Soft drink (regular)	Soft drink (diet)	Fruit juice	No beverage
R8	66%	71%	80%	76%	77%	61%	49%	8%
R7	23%	13%	13%	17%	15%	21%	24%	9%
R6	5%	4%	4%	5%	3%	8%	11%	7%
R5	1%	1%	2%	1%	4%	6%	8%	10%
R4	2%	1%	0%	1%	1%	2%	4%	10%
R3	0%	0%	2%	1%	1%	0%	2%	14%
R2	0%	6%	0%	0%	0%	0%	1%	11%
R1	1%	3%	0%	0%	0%	0%	1%	31%

Chosen alternative

Fig. E.1: Distribution of chosen beverage alternative as a function of amount of time spent looking at that particular alternative. R1 to R8 indicate the ranking in ascending order of time spent looking at an alternative