# An Empirical Model of Screening Rule Choice

**J. Jason Bell**
Saïd Business School
University of Oxford
jason.bell@sbs.ox.ac.uk

**Sanghak Lee**
W. P. Carey School of Business
Arizona State University
sanghak.lee@asu.edu

**Gary J. Russell**
Henry B. Tippie College of Business
University of Iowa
gary-j-russell@uiowa.edu

# Abstract

*We model consideration set formation involving non-compensatory screening rules. Rather than choosing from among alternatives or sets, consumers in our model choose rules. The best rules strike a balance between maximizing utility and minimizing costs of thinking. By adding psychological insights into an econometric framework, we obtain a parsimonious, interpretable model that delivers a wealth of insights from minimal data requirements. The model requires only stated (or revealed) consideration sets, and alternative attributes, and produces interpretable estimates of partworths, thresholds, common rules used, and thinking costs. In addition, the model can be used for managerial counterfactuals. We estimate the model using a dataset of consideration sets from the automobile industry, and explore substantive findings about common screening criteria in automobile consideration, brand competition and the dollar value of costs of thinking, especially in relation to demographics.*

# 1 Introduction

Consideration sets play an important role in explaining consumer behavior. When faced with many alternatives, consumers are unlikely to make a complete evaluation of every option. Numerous studies show that consumers are likely to first screen the entire set of alternatives using some type of elimination or inclusion heuristic (e.g., Montgomery and Svenson 1976, Payne 1976, Hauser and Wernerfelt 1990, Roberts and Lattin 1991, Shocker et al. 1991, Bronnenberg and Vanhonacker 1996, DeSarbo et al. 1996, Jedidi et al. 1996, Wu and Rangaswamy 2003). This process where consumers screen using an efficient heuristic and then make a final choice using compensatory, computationally burdensome comparison is known informally as "two stage choice."

Despite the evidence in favor of heuristics, most models still treat agents as performing an exhaustive search over sets or alternatives in order to form a consideration set. In the model of this paper, the consumer chooses among a relatively small set of screening rules, and uses the chosen rule to arrive at a consideration set. This approach aligns with the well known literature on "adaptive decisions", where agents are thought to construct decision criteria according to context and goals (Payne et al. 1988, Payne et al. 1993, Bettman et al. 1998). This novel treatment of rules is the major differentiating factor of our model.

The fact that we treat rules much like other models treat alternatives has several important effects. First, and most importantly, it allows us to understand which attributes drive screening behavior, and why. Second, it allows for the use of well understood, well proven methods: in essence we estimate a choice model, only with a different concept of what is being chosen. Finally, our approach lends a high degree of flexibility to rule specifications. Using our framework, it is almost trivially easy to change between models where consumers choose among all conjunctive rules, all disjunctive rules, or even disjunctions-of-conjunctions (Hauser 2014).

Previous models incorporating heuristics are difficult to interpret (Hauser et al. 2010), or they do not directly model consideration sets, but rather they model final choice and treat consideration as implicit to the process (Gilbride and Allenby 2004, Gilbride and Allenby 2006). Because our model is parametric, it is more interpretable than previous heuristic-based screening models. Additionally, the model is parsimonious in comparison to models that assume consumers select a consideration

set from the group of all possible subsets of alternatives.

In the next section, we review the literature on consideration sets. We then introduce the model and the likelihood function. Next we discuss identification. We then explain the Bayesian estimation procedure and data sources. Finally, we present and interpret the results of our estimation in the context of the automobile industry.

## 2 Literature Review

Consideration sets have received attention in marketing since Howard and Sheth (1969). Since the literature on this topic is vast, we review only those papers most relevant to our model.

Tversky (1972) introduced a model called "Elimination by Aspects," commonly referred to as EBA. The EBA model did not refer to the concept of a consideration set, but it did contribute to the idea of screening. The EBA model assumes that consumers choose screening criteria, and use them sequentially to remove items from their consideration set. Later, Fader and McAlister (1990) applied the concept of screening to grocery store purchasing data. However, screening was introduced mainly as a way of improving model fit. Fader and McAlister (1990) also highlight a contrast in the consideration set literature between compensatory and non-compensatory models. A compensatory model, such as a logit model, allows for trade-offs in characteristics, while non-compensatory models do not. In compensatory models, favorable levels in one attribute can compensate for unfavorable levels of other attributes. In linear, additive utility frameworks, given certain boundary conditions on the part-worths, very favorable levels for one attribute can compensate for unfavorable levels in all other attributes.[1]

Many papers have found that failing to account for consideration sets can lead to poor results. For example, Chiang et al. (1998) found that ignoring consideration set heterogeneity across consumers leads to underestimation of the effects of marketing variables and overestimation of the effects of preferences and history.

This paper is related to three papers on consideration sets in marketing. Gilbride and Allenby (2004) estimated a now well-known model to compare non-compensatory screening rule types: 'con-

[1]It is important to note that not all additive, linear utility models are compensatory. See Hauser (2014).

junctive rules' and 'disjunctive rules.' In conjoint data about cameras, the authors found the model with conjunctive rules to fit better.

Gilbride and Allenby (2006) compared an EBA model against an economic model of screening in a conjoint setting, and found the economic model to perform better on that dataset. The parametric specification of the economic screening model in that paper resembles this one. However, in that model, the target data is final choice data rather than consideration set data, and the screening is not as flexible to specify. In addition, that model was designed for conjoint data, where the model of this paper is designed for survey data. This is an important point, since analyses designed for conjoint data rely on special structures, such as orthogonality of attributes, pre-defined aspect levels, and a carefully controlled choice context. The model of this paper attempts to pull insight from observational data, to which none of these things apply.

Finally, Hauser et al. (2010) estimated consideration set rules using linear programming. The ultimate goal in that paper is to achieve prediction accuracy and to incorporate a wide range of rule types. Because of the estimation approach, the model predicts well but is not interpretable.

Table 1 compares this model to the three models described in the preceding paragraphs.

Table 1 about here.

# 3 Model and Estimation

Here we provide a conceptual overview of the model, before detailing the equations. Figure 7 shows the flow of the model from the consumer perspective.

Figure 7 about here.

## 3.1 Model Outline

The consumer's objective is to choose a screening rule from among those available. The rules that are available to each consumer are defined by the researcher. We define the set of all rules (as assumed by the researcher) as $\mathbb{S}$. Since rules in this model operate in tandem with thresholds, we

first explain our approach to modeling thresholds, and then define the operation of rules.

In Figure 7, the first component of the model are the attribute thresholds. Each consumer has a vector of exogenously given thresholds ($\boldsymbol{\gamma}_i$). The vector is of length $J \times K$, where $J$ is the number of alternatives, and $K$ is the number of attributes. The thresholds are "cut-offs", or points below which an attribute is not satisfactory (in our model, consumers only screen alternatives out if they are below a threshold, as opposed to above it.) We model these thresholds as heterogeneous ($\boldsymbol{\gamma}_i$ depends on $i$), and as dependent upon past purchase variables. In addition, we assume different thresholds for different alternatives for the same consumer. This is an important aspect of our model. We will discuss this further below.

Rules in this model are vectors of 1s and 0s, whose lengths correspond to the number of attributes ($K$). A 1 indicates that an attribute is included in the rule, or in other words, that attribute is used for screening. Each attribute in a rule has an associated threshold. An alternative passes a screening rule if each of its attributes exceeds the associated thresholds included in the rule. For example, suppose we have two alternatives, A and B, and two attributes ($K = 2$). Suppose A has attribute values (0.2, 0.6) and B has values (0.5, 0.3). Now suppose a consumer's thresholds for A are $(0.5, 0.5)$ and her thresholds for B are $(0.4, 0.4)$, so $\boldsymbol{\gamma}_i = (0.5, 0.5, 0.4, 0.4)$. Then the rule $\{0, 1\}$, which indicates that the second attribute is included in the rule, will not remove A (0.6 exceeds 0.5) but will remove B (0.3 does not exceed 0.4). Notice here again that the thresholds applied to A and B for the same attribute are *different*, which is a unique and powerful feature of our model.

As noted in previous literature (Gilbride and Allenby 2004, Hauser et al. 2010, Hauser 2014), there are different types of rules that involve thresholds. Some of the more prominent types in the literature are conjunctive rules, disjunctive rules, and disjunction-of-conjunction rules. Though our model is capable of accommodating any of these types of rules, we use conjunctive rules only. The main reason for using a single type of rule is to allow the reader to focus on the unique aspects of our model. We chose conjunctive rules over disjunctive rules based on early experiments with model fit.

### 3.2 How Consumers Apply Rules

For what follows, note that we index rules in the model by $s$, alternatives by $j$, and attributes by $k$. The set of all rules is $\mathbb{S}$, the set of all alternatives is $U$, and the size of $U$ is $J$. The number of attributes is $K$. Each rule generates a consideration set. Because this consideration set is a deterministic result of the rule, $s$, and because rules lead to different sets depending on $\boldsymbol{\gamma_i}$, we use the following notation to signify a consideration set which is the result of screening using rule $s$ while having the thresholds $\boldsymbol{\gamma_i}$: $c_s(\boldsymbol{\gamma_i})$. We can now write the exact mathematical specification for screening using rule $s$:

$$c_s(\boldsymbol{\gamma}_i) = \{j : x_{jk} > \gamma_{ijk}, \forall k \in s\}. \tag{1}$$

The notation $\gamma_{ijk}$ is the entry in $\boldsymbol{\gamma}_i$ corresponding to attribute $k$ and alternative $j$. It is useful to think of the above equation as defining a function that takes thresholds and alternatives as inputs and which returns considerations sets. Such a function exists for every single rule $s$.

Given thresholds ($\boldsymbol{\gamma}_i$), each consumer in the model will select a screening rule. A rule, $s$, is evaluated by the set it produces ($c_s(\boldsymbol{\gamma_i})$). At its core, choosing a rule is a balancing act between simplicity and utility. Because consumers have thinking costs (Shugan 1980), rules that remove more alternatives from consideration are better, all else equal. However, rules must not remove desirable attributes. The best rules, then, are those which remove many low-utility alternatives. We will of course make this intuition more specific shortly.

We think of consumers as anticipating the final choice stage that will come after the screening stage. Consumers have some expectation about the value of making a final choice from a particular set. For this reason, consumers use an expected utility of the set produced by each rule as part of the rule selection process.

To remind the reader: consumers enter the choice environment with exogenously given thresholds. Given those thresholds, they are faced with the task of finding a rule which best balances expected utility and thinking costs.

### 3.3 Rule Utility

We now define 'rule utility' ($v_s(\cdot)$). Each alternative removed from consideration 'saves' the consumer a unit of thinking cost. The utility of the cost of thinking is weighted by the parameter $\lambda_i$. While consumers wish to make the final choice easier, they also wish to preserve expected utility. This expected utility will depend on the set produced by the rule, $c_s(\boldsymbol{\gamma}_i)$, and the part-worths associated with the attributes. We use $\boldsymbol{\beta}$ to represent the vector of part-worths. The value of screening using rule $s$ is

$$v_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i}) = \lambda_i\big(J - |c_s(\boldsymbol{\gamma_i})|\big) - \big(E[u(U, \boldsymbol{\beta})] - E[u(c_s(\boldsymbol{\gamma_i}), \boldsymbol{\beta})]\big). \tag{2}$$

The first term captures the main benefit of screening. As discussed, the parameter $\lambda_i$ represents the added cost of thinking for an extra item in the consideration set. We use 'pipe' notation to indicate set size, so $|c_s(\boldsymbol{\gamma_i})|$ is the size of the set $c_s(\boldsymbol{\gamma_i})$. The second term represents the loss of expected utility that the rule induces by eliminating alternatives from consideration. It can be thought of as a penalty for screening. We use an expected utility instead of a deterministic value since at this point in the consideration set formation process, the consumer has not computed an exact value for the utility of each alternative. This captures a notion of utility that is not as well-formed as compensatory utility.

### 3.4 Thinking Costs ($\lambda$) and Partworths ($\boldsymbol{\beta}$)

The cost of thinking parameter ($\lambda_i$) and the partworths ($\boldsymbol{\beta}$) play different roles in the evaluation of rules and their associated sets. Increases in the cost of thinking parameter for a particular consumer $i$ lead to small consideration sets, holding all else constant. The mechanism by which costs of thinking lead to smaller sets is rule choice. Rules that produce large sets are penalized in proportion to $\lambda_i$. For larger values of $\lambda_i$ (i.e., larger costs of thinking) this penalty is more severe. The model also allows for for a consumer to have a $\lambda_i$ which is *negative*, indicating that such a consumer receives utility from the process of considering more vehicles. Such a consumer would be very likely to consider every alternative available.

The partworths ($\boldsymbol{\beta}$) drive expected utility, and are attribute weights. A high positive value of $\beta_k$ indicates that attribute $k$ has a large impact on the expected utility of consideration sets. Since the utility of rules is determined only solely by the sets they create, this attribute importance passes through to the rule choice process. For high positive values of $\beta_k$, alternatives with high levels of attribute $k$ increase the expected utility of any sets which include those alternatives. In turn, rules which lead to sets which include alternatives with high levels of attribute $k$ are more likely to be chosen. Thus, for high, positive values of $\beta_k$, we expect to see alternatives with high levels of attribute $k$ included in consideration sets often.

We model $\lambda_i$ as an intercept term plus coefficients on demographics. We denote by $D$ the matrix of demographic variables, with each row being called $d_i$, and containing the demographics of consumer $i$. Then we have

$$\lambda_i = \lambda_0 + d_i \boldsymbol{\lambda} \tag{3}$$

### 3.5 Rule Selection

Given her parameters, the consumer selects the rule with the highest utility. From the perspective of the researcher, however, this choice is probabilistic because of unobservables, and each consumer selects a rule with probability proportional to $v_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i})$. Given rule utilities, this is equivalent to a multinomial choice model, but instead of final alternatives, the consumer chooses among rules. We assume a logit form for this choice process, so

$$w_{is} = v_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i}) + \epsilon_{is}, \quad \epsilon_{is} \sim EV(0,1), \tag{4}$$

where $EV(0,1)$ is the Standard Extreme Value Distribution with PDF

$$f(x) = e^{-(x+e^{-x})},$$

and $\epsilon_{is}$ represents utility associated with rule $s$ that is not captured by $v_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i})$. As mentioned, we assume that $\epsilon_{is}$ is observable to consumer $i$ but unobservable to the researcher. Then from the

researcher's point of view, the probability that consumer $i$ selects rule $s$ for screening is

$$\tau_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i}) = \frac{e^{v_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i})}}{\sum_{s'} e^{v_{s'}(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i})}}. \tag{5}$$

Conditional on $\boldsymbol{\gamma_i}$, applying rule $s$ leads deterministically to a consideration set.

### 3.6 Threshold Model

We now turn to the model of individual threshold vectors. The model posits that all consumers have a common baseline threshold vector, to which we add terms which account for individual-level heterogeneity and individual-alternative-level inertia. In our model, consumers diverge from the baseline threshold as a function of inertia. This inertia term is designed for a context where we can observe the brand of the consumers' most recently purchased products. Then, if consumer $i$ last purchased brand $b$, then she may have different thresholds for alternatives of brand $b$. The size of this shift is determined by an inertia parameter ($\theta$). The exact specification follows.

$$\begin{aligned} \gamma_{ijk} &= \overline{\gamma}_k + \theta m_{ij} + \zeta_{ijk}, \\ \zeta_{ijk} &\sim N(0, \sigma^2) \end{aligned} \tag{6}$$

We use a fairly standard hierarchical setup for the latent threshold vectors, with a common intercept ($\overline{\gamma}_k$), an observed, individual-specific component ($\theta m_{ij}$) and an unobserved component ($\zeta_{ijk}$). The indicator variable captures inertia and heterogeneity and is defined as:

$$m_{ij} \equiv \begin{cases} 1 & \text{if consumer i's last brand was} \\ & \text{the same brand as alternative j} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Thus, $m_{ij}$ acts on the thresholds of each attribute (i.e., it doesn't depend on $k$) and if inertia is present, we expect $\theta$ to be negative, or in other words, we expect that cars that share the same brand as $i$'s previous car have an easier time entering the consideration set. While this term is meant to capture inertia, it may also capture some heterogeneity in the form of time-invariant differences

across consumers which are correlated with preferences.

The term $\zeta_{ijk}$ represents residual heterogeneity remaining after that captured by $m_{ij}$. More specifically, it is consumer $i$'s deviation from the population mean (conditional on $m_{ij}$) for a screening threshold associated with attribute $k$ and alternative $j$. Since $\zeta_{ijk}$ is specific to each alternative, any rule can lead to any consideration set in this model, if the $\zeta_{ijk}$s are sufficiently extreme. Because the researcher doesn't observe this term but the consumer does, $\zeta_{ijk}$ has the effect of making the model likelihood smoother from the researcher's perspective while preserving the non-compensatory nature of screening from the consumer's perspective.

As mentioned, we assume that consumer $i$ knows $\zeta_{ijk}$ but the researcher only knows the distribution of $\zeta_{ijk}$ across people and alternatives.[2]

The functional form for the expected utility of $c_s(\boldsymbol{\gamma}_i)$ is

$$E_{\eta_{ij}}[u(c_s(\boldsymbol{\gamma_i}), \boldsymbol{\beta})] = \log \left[ \sum_{j \in c_s(\boldsymbol{\gamma_i})} \exp\{X_j \boldsymbol{\beta}\} \right]. \tag{8}$$

This form is often referred to as the 'inclusive value' (Train 2003) of the set $c_s(\boldsymbol{\gamma_i})$.

Substituting (8) into (2), we have

$$v_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i}) = \lambda_i\big(J - |c_s(\boldsymbol{\gamma_i})|\big) + \log \left[ \frac{\sum_{j \in c_s(\boldsymbol{\gamma_i})} \exp\{X_j \boldsymbol{\beta}\}}{\sum_{j \in U} \exp\{X_j \boldsymbol{\beta}\}} \right] \tag{9}$$

This leads to screening rule probabilities. For each rule $s$, the probability that a consumer selects rule $s$ to use for screening is given by

$$\tau_s(\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma_i}) = \frac{e^{-\lambda_i |c_s(\boldsymbol{\gamma_i})|} \sum_{j \in c_s(\boldsymbol{\gamma_i})} e^{X_j \boldsymbol{\beta}}}{\sum_{s'} \left[ e^{-\lambda_i |c_{s'}(\boldsymbol{\gamma_i})|} \sum_{j \in c'_s(\boldsymbol{\gamma_i})} e^{X_j \boldsymbol{\beta}} \right]}. \tag{10}$$

[2] Our specification implies that consumers may use different cutoffs for relatively similar alternatives. Intuitively, two likely reasons for differing cutoffs might be: attribute level trade-offs (e.g. perhaps the consumer accepts a lower mpg in return for more power) and context effects (e.g. a convincing salesperson downplaying the F-250's lower mpg).

The intuition for this form is that rules which lead to large sets are penalized (i.e., less likely to be chosen), while rules that lead to sets with more expected utility are favored (i.e., more likely to be chosen). The severity of the 'larger-set' penalty depends on the parameter $\lambda_i$, and the importance of an attribute is measured by $\boldsymbol{\beta}$. The numerator displays the balancing act at the heart of the model: larger sets will tend to have higher utility, but they will also lead to increased thinking costs. The best rules lead to small, high-utility sets.

### 3.7 Likelihood Function

The likelihood for an individual with observed set $c_i$ is the probability of $c_i$, $s_i$ and $\boldsymbol{\gamma_i}$ conditional on the parameters. For concision, in the following exposition, as well as in Section 3.8, we will suppress the dependency of the likelihood on observed data other than $\{c_i\}$. These suppressed values are: the matrix of attributes values, $X$, the indicator variables for previous brand owned, $\{m_{ij}\}_{i=1,...,N,j=1,...,J}$, and $D$, the matrix of demographics. We start with the joint probability of a consideration set, rule, and a vector of thresholds: $p(c_i, s, \boldsymbol{\gamma_i}|\boldsymbol{\beta}, \lambda_i, \theta, \overline{\boldsymbol{\gamma}}, \sigma^2)$, and integrate over the distributions of $\boldsymbol{\gamma}_i$ and $s$, which are continuous and discrete, respectively. This gives the marginal probability of $c_i$, $p(c_i|\boldsymbol{\beta}, \lambda_i, \theta, \overline{\boldsymbol{\gamma}}, \sigma^2) \equiv \ell(c_i|\boldsymbol{\beta}, \lambda_i, \theta, \overline{\boldsymbol{\gamma}}, \sigma^2)$.[3] After simplification, the exact form, for individual $i$, is:

$$\ell(c_i|\boldsymbol{\beta}, \lambda_i, \theta, \overline{\boldsymbol{\gamma}}, \sigma^2) = \int_{-\infty}^{\infty} |\mathbb{S}_{c_i,g}|\tau_s(\boldsymbol{\beta}, \lambda_i, g) f(g|\theta, \overline{\boldsymbol{\gamma}}, \sigma^2) dg, \tag{11}$$

Where $\mathbb{S}_{c_i,\boldsymbol{\gamma_i}}$ is the set of all rules that produce the observed set $c_i$ given a set of thresholds $\boldsymbol{\gamma_i}$.[4] The size of $\mathbb{S}_{c_i,\boldsymbol{\gamma_i}}$ is denoted by $|\mathbb{S}_{c_i,\boldsymbol{\gamma_i}}|$.

### 3.8 Bayesian Estimation

Rather than performing the integration in (11), we use data augmentation. We remind the reader that $\boldsymbol{\gamma}_i$ is used to mean the entire vector $(\gamma_{i11}, ..., \gamma_{iJK})$. Just as in Section 3.7 we suppress $X$, $D$,

[3] Details are in the appendix.

[4] There is no guarantee that $\mathbb{S}_{c_i,\boldsymbol{\gamma_i}}$ is a singleton.

and $\{m_{ij}\}_{i=1,\ldots,N,j=1,\ldots,J}$ for notational simplicity. The full posterior of the unknowns is

$$p\left(\boldsymbol{\beta}, \{\lambda_i\}, \boldsymbol{\gamma_i}, \theta, \overline{\gamma}, \sigma^2 \middle| \{c_i\}\right) \propto \ell\left(\{c_i\}|\boldsymbol{\beta}, \{\lambda_i\}, \boldsymbol{\gamma_i}\right) p\left(\boldsymbol{\gamma_i}|\theta, \overline{\gamma}, \sigma^2\right) p\left(\theta\right) p\left(\overline{\gamma}\right) p\left(\sigma^2\right) p\left(\boldsymbol{\beta}\right) p\left(\{\lambda_i\}\right) \quad (12)$$

In the above, for ease of notation we denote $\{c_i\}_{i=1,\ldots,N}$ as $\{c_i\}$. In words, $\{c_i\}$ is the set of all observed consideration sets. We assume that consumers form consideration sets independently of one another (this amounts to an assumption that $\epsilon_{is}$ is independent across consumers.) This means that the first term on the right-hand side of 12 can be expressed as a product over individuals:

$$\ell(\{c_i\}|\boldsymbol{\beta}, \{\lambda_i\}, \boldsymbol{\gamma_i}) = \prod_{i=1}^{N} \ell(c_i|\boldsymbol{\beta}, \lambda_i, \boldsymbol{\gamma}_i). \quad (13)$$

### 3.9 Drawing from the Posterior

The exact method used to draw from the posterior distribution is found in the web appendix. The outline is as follows. We use the Metropolis-Hastings algorithm to sample from the conditional distributions of $\boldsymbol{\beta}$, $\overline{\gamma}$, $\lambda_i$, and $\theta$. Additionally, we employ data augmentation (Tanner and Wong 1987) at each step in the chain by drawing an $N \times (J \times K)$ matrix $\boldsymbol{\gamma}$ whose rows contain $\overline{\gamma_i}$ for each consumer. This matrix is used in a Bayesian regression to estimate the vector of baseline thresholds ($\overline{\gamma}$) and the inertia parameter ($\theta$).

## 4 Identification and Simulation

In this section we use the explore identification, and simulate the model.

### 4.1 Mathematical Identification

First we explore the mathematical identification of the parameters, each in turn.

### 4.1.1 Identification of $\boldsymbol{\beta}$

In this model, a global intercept part-worth term (common to all alternatives) is not identified. It is easy to see this by including an intercept in equation (10), and noting that it cancels.[5] This implies that we cannot include all levels of any dummy variable in estimating this model, since doing so is equivalent to including an intercept.

In addition to this strict non-identification of a global intercept, there are some conditions that, in theory, could weaken the identification of $\boldsymbol{\beta}$.[6] However, in both the simulation and empirical dataset, we encountered no problems identifying $\boldsymbol{\beta}$.

### 4.1.2 Identification of $\overline{\gamma}$

The thresholds $\overline{\gamma}$ are identified by attribute levels within chosen consideration sets, in particular those near the low end of the consideration sets. baseline thresholds cannot be identified well if they are not 'binding' in the sense that they don't eliminate any alternatives.

## 4.2 Simulation

We simulate the model for three reasons: to confirm that the estimation procedure can recover parameters correctly, to explore the behavior of the model under different settings, and to test the predictive performance of the model on the simulated sample. The steps used to simulate the model follow.

1. Choose parameters $J$, $K$, $N$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda_i}$, $\overline{\gamma}$, $\theta$, and $\sigma^2$, along with data $X$, $\{m_{ij}\}$, $D$.

2. Draw $\zeta_{ijk}$ for each consumer and compute $\gamma_{ijk} = \overline{\gamma}_k + \theta m_{ij} + \zeta_{ik}$, for each attribute $k$.

3. Conditional on $\boldsymbol{\gamma}_i = (\gamma_{i1}, ..., \gamma_{iK})$ for each consumer, calculate $v_{is}$ for each $s$, and draw $\epsilon_{is}$. Consumer $i$ chooses the $s$ with the highest $w_{is} = v_{is} + \epsilon_{is}$.

4. Each consumer applies the rule they chose using $\boldsymbol{\gamma}_i$ to get a consideration set.

[5] See the appendix for details.
[6] See the appendix for a detailed discussion.

After generating a simulated dataset, we can measure the predictive performance of the model using different metrics. We present prediction tables of the mean squared error computed on aggregate consideration shares. To obtain this, we predict consideration sets for each consumer, and then for each alternative we compute the percentage of consumers who consider it in the predicted data. Then we measure this figure against the observed figure in the original simulated dataset to get the MSE.

Table 2 about here.

#### 4.2.1 Empirical Identification of $\boldsymbol{\beta}$ vs. $\gamma$

Table 3 compares the influence of $\boldsymbol{\beta}$ vs $\overline{\boldsymbol{\gamma}}$ on various statistics about an attribute within consideration sets. In each cell of the table, there are 4 fields: Mean, Min, Max, and SD. In order to obtain each field, we first simulate consideration sets using our model, with only a single attribute ($K = 1$), and with the parameter values for that attribute equal to those specified in the row and column of the cell. Though the attribute in the simulation is randomly generated, it helps for the purposes of understanding the tables below if we think of it as 'quality.' Then, in order to fill in each cell, we compute each field for each consumer's consideration set, one by one, and then take the average of that field over all consumers. For example, to get the Min field, we first find the 'quality' of the alternative with the lowest 'quality' for each consumer. Then, we take the average of these minimum values, and that becomes the Min field in the cell. The other fields are computed in an analogous way.

Table 3 about here.

What we see is that the marginal impact of $\boldsymbol{\beta}$ on attribute levels in the considerations set falls very quickly. This is because $\boldsymbol{\beta}$ allocates probabilities across rules, and the impact of this allocation is limited by various factors: the tightness of the thresholds, the number of alternatives, and the variation in the universal set, as discussed in Section 4.1. As $\boldsymbol{\beta}$ increases, probability reallocates across rules, but eventually the probability is "used up", so the impact of $\boldsymbol{\beta}$ reaches a limit.

Table 4 contains a more different breakdown of the impacts of $\boldsymbol{\beta}$ and $\overline{\boldsymbol{\gamma}}$. The rows correspond to different combinations of $\boldsymbol{\beta}$ and $\overline{\boldsymbol{\gamma}}$, and the columns correspond to the shares of each alternative

in the simulated dataset. In the simulation, we used $J = 6$ alternatives. The final column shows the average consideration set size in the dataset.

Table 4 about here.

Just as in Table 3, we see in Table 4 that the impact of $\boldsymbol{\beta}$ depends somewhat on $\overline{\boldsymbol{\gamma}}$, except in the case of alternatives with very high levels of the simulated attribute, which we have been referring to as "quality." If $\overline{\boldsymbol{\gamma}}$ is very loose, almost all of the power to identify $\boldsymbol{\beta}$ comes from the presence or absence of the alternative with the highest level of "quality." If $\overline{\boldsymbol{\gamma}}$ is tight, $\boldsymbol{\beta}$ has a larger influence on the shares of all of the alternatives.

These two tables help us to understand the different ways that $\overline{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}$ impact consideration sets in our model. In addition, they aid in the interpretation of these two parameter types.

## 5 Application to the Automobile Market

We use the model developed in this paper to analyze automobile purchases in the compact car category. We have several goals: to gauge whether the model parameters are sensible, to see if the model fits the data well, and to interpret parameters. We also use the model to explore the competitive impact of a counterfactual where one of the alternatives is improved. First, we explain our data sources and final estimation sample, and then we present and interpret parameter estimates, and finally we present and discuss the counterfactual.

To estimate the base model, we require the following information:

1. The consideration sets, demographics and the previously owned car for each consumer.

2. The attributes of each product in the consideration set.

These data come from two major sources. The first is a survey, conducted by the leading market research firm in the industry, of consumers purchasing new cars. Consideration sets, demographics, and the previously owned car come from this survey, while the vehicle attributes come from Edmunds.com.

The survey targets every car buyer in the United States as a subject. The survey is administered by mail roughly 3 months after the time of purchase. Our dataset comes from the year 2012. We refine the data by selecting consumers who considered only compact cars. However, since this tends to remove those consumers with a larger consideration set size, we use stratified sampling to match the original distribution of consideration set sizes. In other words, we impose the assumption that the distribution of consideration set sizes among consumers who considered *exclusively* compact cars corresponds to the distribution of consideration set sizes of consumers who consider *at least one* compact car. Finally, we randomly sampled 500 individuals.

Each observation in our dataset represents one car purchase for one individual. However, along with the purchased car, respondents were asked what models they most seriously considered besides the one they purchased. We use the answers to this question as the consideration set of that consumer.

Consumers are also asked a series of questions regarding demographics, such as their age, education, income, and whether they live in a more urban or rural area. We use the answers to these questions to estimate the model with thinking cost heterogeneity.

The data from Edmunds.com contains attributes of the compact cars, such as MPG and power. For price, we use the average over all final transaction prices paid by all buyers of each vehicle. The final transaction price is taken from the survey data, described above.

We have access to other variables from Edmunds.com, such as vehicle dimensions (length and width), ground clearance, GPS presence, upholstery type, and even driver legroom. However, after a long process of pre-testing both using descriptive methods and our model, we decided to use MPG, HP, price, and brand as the explanatory variables in the model. We made this decision for two reasons: the variables we omitted had limited impact on the model fit and parameter estimates, and secondly for parsimony. We face a somewhat steep estimation penalty for including many attributes in the model, since the rule space expands combinatorially with the number of attributes. We determined that the small gains in model fit would not be worth complicating the model.

### 5.1 Data Description

Here, we explore aspects of the dataset related to parameters of interest. First, Figure 2 shows the distribution of consideration set sizes. It is evident that most consideration sets have only a single car. This highlights a benefit of the approach we take in this paper of modeling consideration sets instead of final choice. If we were modeling final choice directly, consideration sets of size one pose a problem. This is because models of final choice require a choice set, and the most obvious candidate for the choice set is the consideration set, if it is observable. If the choice set is a singleton, however, the final choice is trivial, and contains no information about tradeoffs between attributes. This means the model would only use information from the minority of consumers. On the other hand, if a researcher were to assume that all consumers choose from the universal set (the set containing every alternative, in this case all of the compact cars), then consideration set information would not be used. Our model allows for consideration sets of any size, and so the parameters are better informed than a final choice model would be.

Figure 2 about here.

The continuous attributes of the automobile data are summarized in Table 5.

Table 5 about here.

The demographic data are summarized in Table 6.

Table 6 about here.

### 5.2 Parameter Estimates

Our model specifications includes power, mpg, and price as continuous variables, in addition to 14 brand intercepts. The matrix of attributes ($X$) is standardized.

Some care is needed when interpreting the price threshold. Since the rules are such that consumers screen out alternatives if they fall *below* their thresholds, this would intuitively mean the

model could only explain the screening of *inexpensive* alternatives. We used the negative of the standardized price values, so that we instead interpret screening as removing expensive alternatives. In the table with parameter estimates, we have adjusted the price part-worth so that it can be interpreted the same as the other part-worths. The price *threshold*, however, must be interpreted slightly differently. While the thresholds are typically interpreted as lower bounds, the price is interpreted as an upper bound.

Table 7 displays parameter estimates for the model.

Table 7 about here.

The preference parameter for price is negative and significant, as expected: higher prices are less desirable. More specifically, consideration sets are more attractive to the degree that they contain lower-priced alternatives, *ceteris paribus*. The part-worths for power and MPG are both positive and significant, unsurprisingly. To help interpret the brand intercepts, Figure 4 plots them against consideration set penetrations by brand.

Figure 4 about here.

There is a positive relationship between brand intercepts and the percent who consider each brand. The correlation is roughly 0.37 between these two variables, which suggests that screening behavior related to brands does explain a modest component of brand popularity, but not all of it. This provides evidence that attributes other than brand play a significant role in consideration set formation.

The inertia parameter ($\theta$) is negative and significant. Since the inertia parameter modifies thresholds, this suggests that the average consumer lowers her threshold for alternatives of the same brand as her previously owned alternative. This leads to the pattern that consumers are more likely to consider alternatives whose brands match their most recently owned ones.

## 5.3 Costs of Thinking

As a reminder: the thinking cost parameter measures the burden of considering a single alternative. We can change the units to make them easier to process, but in any units, higher values of the parameter correspond to smaller consideration sets. The mechanism in the model that leads to this is the rule choice process. Rules that lead to larger consideration sets are penalized proportionally to the thinking cost parameter ($\lambda_i$). The higher the thinking cost parameter, the higher the penalty for large sets, the more likely the consumer will choose a rule that leads to a smaller set.

The thinking cost intercept is positive and significant. This parameter is directly comparable to the part-worth parameters only for consideration set sizes of one. In that condition, it is interesting to compare thinking costs to price. Given the demographics of the consumers in the sample and the model estimates, the average thinking cost is imputed to be 5.09. This comes from multiplying the thinking cost parameters by the values of the demographic variables for each consumer in the sample and taking the average. To put this in dollar terms, we first divide by the price coefficient to get thinking cost in units of price standard deviations. This gives 0.90. Then, we multiply by the original standard deviation in the raw price data to get the value in dollar terms. This gives \$2478 for the average thinking cost. In words, in order to consider an additional vehicle, the average consumer requires an additional expected utility that converts roughly to \$2500.

To understand the impact of demographics on thinking costs, we can investigate three consumer demographic profiles. The first profile will illustrate the demographics associated with very low thinking costs (whom we would expect to have large consideration sets). The second will illustrate the profile for maximum thinking costs, and the third will illustrate the median.

1. **Low Costs of Thinking.** Age is positively correlated with thinking costs, education negatively, and living in a rural area positively. The youngest person in the sample is 19 years old, and the highest level of education is 13, which represents a PhD. However, for the sake of realism, we use the highest level of education that any 19-year old in the sample attained, which was 10, and which represents "some college." Finally, living in an urban area is associated with lower thinking costs. Our estimates predict that a 19-year old with "some college" living in an urban area has thinking costs of 3.92, which translates to \$1918 for the cost of

considering one alternative.

2. **High Costs of Thinking.** Using the opposites of the variables above, we see that the profile of a consumer with highest thinking costs is older, with less education, and who lives in a rural area. The oldest consumer in the sample is 95. Our estimates predict that a 95 year old with no education from a rural area has thinking costs of 7.63, which translates to $3733 for the cost of considering one alternative.

3. **Median Costs of Thinking.** A profile of the median consumer is a 45 year old with some college who lives in an urban area. Thinking costs for such a person are estimated at 4.95, which translates to $2421 for the cost of considering one alternative.

Tables 8 and 9 summarize the above profiles to highlight the impacts of individual and joint demographic differences.

Tables 8 and 9 about here.

### 5.4 Attributes Used in Screening

The threshold parameters only apply in the case when a consumer screens by the relevant attribute. Otherwise, they are meaningless. In order to help interpret them, we provide the model estimates of how often each attribute is included in a rule. To obtain this information, we simulate the model, find the rules with the highest probability of being chosen, and count up the attributes included. This gives an estimate of how often consumers screen using each attribute.

Table 10 about here.

Nearly every consumer screens by brand, according to the model. This is perhaps not surprising. Power and MPG are very close in percentage, but around three-quarters of the sample are predicted as using those attributes to screen. Though power and MPG are used almost the same amount for screening, since MPG has a much higher threshold, it 'cuts' more alternatives. Finally, price is used

only about half the time. This makes sense in the context we study, since we have selected a sample of compact car buyers, and so there is relatively less variation in price.

### 5.5 Model Fit and Counterfactuals

Figure 3 compares the predicted versus actual consideration shares for the compact car data. The model fit, as judged by inspection, is very good.

Figure 3 about here.

Table 11 presents some counterfactual results. To produce this table, we create a counterfactual attribute matrix ($X$) by replacing the price, MPG, or power of the Subaru Impreza with a value that is theoretically more favorable. For example, the counterfactual price of the Subaru Impreza is 10% lower than the real price, while the counterfactual MPG is 10% higher. With the counterfactual attribute matrix, we simulate a dataset of consideration sets and measure the consideration shares of each alternative.

Table 11 about here.

While the Impreza gains share on each counterfactual, it gains the most from an increase in MPG. The model predicts that, with a 10% increase in MPG, the Impreza would be considered more often than the Civic. The extra penetration means that roughly 45 incremental people out of every 500 would consider the Subaru it it had a 10% increase in MPG. Notably, some alternatives, such as the Ford Focus, *benefit* from the increase in MPG of the Impreza. This is a very surprising and important result. In the case of the Ford Focus, the intuition is clear: the Subaru Impreza becomes more attractive, and induces more people to screen on MPG, which in turn 'drags' the Focus into more consideration sets. This is because the Ford Focus has a high MPG, so it tends to pass rules that include MPG more easily. The idea that making one alternative better could benefit a competitor resembles effects such as the decoy effect in behavioral literature. In this model, however, the results comes as a side-effect of rule choice: when consumers are faced with a different

set of attributes, they change how they form their sets. Thus, our model captures counterfactual impacts at a deeper level than previous models, which assume static consideration set formation rules.

In order to learn the impact on brand competition in the counterfactual, we use coincidence matrices at the brand level to create multidimensional scaling maps. First, to show that the map from simulated data aligns well with the one from observed data, we superimpose them. The grey labels are from the model, the black are from observed data. We used a Procrustes Transformation (Borg and Groenen 2005) for comparability.

Figure 5 about here.

Each map uses a sequence of pairwise brand distances. The correlation between the sequences of pairwise brand distances is 0.8, providing evidence beyond inspection that the model captures the brand structure well. To show the impact of the counterfactual, we show the brand map from the model with the brand map of the MPG counterfactual superimposed. We chose the MPG counterfactual because it had the most impact of any of the three that we ran (price, MPG, power). The black text represents the model while the grey (with arrows) shows the counterfactual.

Figure 6 about here.

We see that, originally, Subaru competes most closely with Volkswagen and Mazda. In the counterfactual, the Subaru brand shifts slightly away from Volkswagen and Mazda toward Ford. In addition, the Ford brand moves a significant distance toward Volkswagen, Mazda, and Subaru. In this dataset, the only car with the Ford brand is the Focus. The Focus has very high MPG, which explains Subaru's large shift in the direction of Ford. Ford's shift is probably due to the changes in screening rules, as consumers more often consider Subaru and Ford together. Subaru's impact on screening rules has a gravitational effect on the perceptual map, pulling Ford closer.

Table 12 gives the attribute values used for estimation, as well as the counterfactual attribute

values. Note that all continuous variables are standardized.

Table 12 about here.

# 6 Conclusion

In this paper, we have developed and estimated a model of screening rule choice. Our model takes a unique approach to consideration set formation, because it blends utility theory, from economics, with psychological models of screening rules. In addition, the model incorporates thinking costs, part-worths, and thresholds, using these sets of parameters to parsimoniously describe consideration set formation. Because consumers in the model choose rules instead of directly choosing sets, the model produces interesting and realistic counterfactual results. The realism arises from the adaptive nature of the consumers to changes in product attributes: when aspects of the decision change, the consumers use different mechanisms to form their consideration sets. In this way, the model is able to produce effects consistent with experimental psychological results. We refer to the effect in our counterfactual as a 'dragging' effect: when one alternative improves, that may trigger consumers to favor different rules, which in turn may benefit competing alternatives. The ability of our model to produce such patterns is an important distinguishing feature.

Our model can be estimated using survey data which is widely available in the industry of applicaiton, along with attributes from Edmunds.com. Both of these sources of data are available to, and already in use by, automobile manufacturers. More generally, the model can be estimated using only consideration sets and attributes (and, optionally, demographics), making it widely applicable for empirical work on consideration sets.

Using this model, managers can leverage pre-existing data to understand consideration set formation, key attributes, and costs of thinking. The model can also be used to run realistic counterfactuals.

### 6.1 Model Realism

We propose a consideration set formation model that relies on screening rule choice. We break with a common assumption from previous work that consumers form utility values for each alternative or each subset of alternatives. This assumption results either in an explosion of options, or overly rigid modeling restrictions. In addition, in all previous modeling frameworks, consumer decisions are static with respect to changes in the alternative set or the attributes thereof. Our approach assumes a parsimonious, flexible structure for consideration set formation in which consumers may change decision rules in response to changes in the attributes of alternatives they face. This allows for counterfactual results that mirror effects found in the behavioral literature, such as the decoy effect (Huber et al. 1982).

### 6.2 Minimal Data Requirements

All of the most similar models to this one use conjoint data. While conjoint data are higher quality, they also may be more costly to collect, especially if managers want quarterly results. This relegates previous work of this nature to settings where experimental studies are feasible and cost effective. On the other hand, our model can be estimated using only consideration sets and alternative attributes. These data are commonly available. Not only does our method produce the same results from survey data that other models produce from conjoint (partworths), it actually produces several extra results (thresholds, rules, and thinking costs).

### 6.3 Limitations and Extensions

The current model assumes conjunctive rules. Because our model treats rules like other models treat alternatives, it is straightforward to implement other rule types (see Hauser et al. 2010), and in fact it is possible to mix rule types together. Although we did some minor testing with other rule types using this model (not reported here), it would be interesting to conduct a more thorough exploration. In particular, it would be interesting to compare parameter estimates and model fit under a different set of rule types to those found here.

We assume a relatively simple form for cost of thinking. In the model of this paper, costs accrue

with each additional alternative in a linear fashion. Perhaps other specifications for thinking costs could be tested. Thought the specification found here leads to a well fitting model with reasonable parameter estimates, it is possible that a more complex specification could lead to improvements.

Our model assumes that consumers form preferences across entire groups of attributes simultaneously. However, as highlighted by previous work in economic search (De los Santos et al. 2012), it is possible that the process is sequential in nature. While our dataset would not allow us to empirically distinguish between the two frameworks, other datasets might. This would be an interesting avenue for further research.

Our model does not directly include heterogeneity in the partworths ($\beta$s). This is possibly a limitation. However, it is unclear whether modeling heterogeneity of the partworths would lead to much greater predictive performance. The main reason for this comes from the dataset. Since we don't have repeated measures on individuals, there are boundaries on our ability to identify heterogeneity patterns. The second reason is that we already have heterogeneity in the threshold equation, as well as in the cost of thinking equation. Because of this, there is less important variation left in the data that could be captured by adding partworth heterogeneity. Given that the performance of the model is already rather good, the gains from more complex models may or may not be very significant. Still, the exploration of this topic is a possible avenue for further research.

Finally, while our model is a step forward in terms of realistically modeling consideration set formation, it is by no means a perfect representation of reality. In this model, although consumers are not assumed to form preferences over all possible subsets of alternatives, they are still assumed to have preferences for each rule. Thus, we assume that consumers have costs of thinking for alternatives, but we rely on the fact that there are relatively few rules to avoid introducing thinking costs for choosing the *rules themselves.* While we believe this is a reasonable approach, it would be interesting to focus on this issue, perhaps using conjoint analysis. Our model can be thought of as the net effect of a sequential process, but tracing out the intermediate decisions using granular process data is likely to be a fruitful endeavor.

# References


Bettman, James R, Mary Frances Luce, John W Payne. 1998. Constructive consumer choice processes. *Journal of consumer research* **25**(3) 187–217.

Borg, Ingwer, Patrick JF Groenen. 2005. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.

Bronnenberg, Bart J, Wilfried R Vanhonacker. 1996. Limited choice sets, local price response and implied measures of price competition. *Journal of Marketing Research* 163–173.

Chiang, Jeongwen, Siddhartha Chib, Chakravarthi Narasimhan. 1998. Markov chain monte carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics* **89**(1) 223–248.

De los Santos, Babur, Ali Hortaçsu, Matthijs R. Wildenbeest. 2012. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review* **102**(6) 2955–2980.

DeSarbo, Wayne S, Donald R Lehmann, Gregory Carpenter, Indrajit Sinha. 1996. A stochastic multidimensional unfolding approach for representing phased decision outcomes. *Psychometrika* **61**(3) 485–508.

Fader, Peter S, Leigh McAlister. 1990. An elimination by aspects model of consumer response to promotion calibrated on upc scanner data. *Journal of Marketing Research* 322–332.

Gilbride, Timothy J., Greg M. Allenby. 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science* **23**(3) 391–406.

Gilbride, Timothy J., Greg M. Allenby. 2006. Estimating heterogeneous eba and economic screening rule choice models. *Marketing Science* **25**(5) 494–509.

Hauser, John R. 2014. Consideration-set heuristics. *Journal of Business Research* **67**(8) 1688–1699.

Hauser, John R, Olivier Toubia, Theodoros Evgeniou, Rene Befurt, Daria Dzyabura. 2010. Disjunctions of conjunctions, cognitive simplicity, and consideration sets. *Journal of Marketing Research* **47**(3) 485–496.

Hauser, John R, Birger Wernerfelt. 1990. An evaluation cost model of consideration sets. *Journal of consumer research* **16**(4) 393–408.

Howard, John A., Jagdish Sheth. 1969. *A Theory of Buyer Behavior*. John Wiley & Sons, New York.

Huber, Joel, John W Payne, Christopher Puto. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research* **9**(1) 90–98.

Jedidi, Kamel, Rajeev Kohli, Wayne S DeSarbo. 1996. Consideration sets in conjoint analysis. *Journal of Marketing Research* **33**(3) 364.

Montgomery, Henry, Ola Svenson. 1976. On decision rules and information processing strategies for choices among multiattribute alternatives. *Scandinavian Journal of Psychology* **17**(1) 283–291.

Payne, John W. 1976. Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational behavior and human performance* **16**(2) 366–387.

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

Payne, John W, James R Bettman, Eric J Johnson. 1993. *The adaptive decision maker*. Cambridge University Press.

Roberts, John H, James M Lattin. 1991. Development and testing of a model of consideration set composition. *Journal of Marketing Research* 429–440.

Shocker, Allan D, Moshe Ben-Akiva, Bruno Boccara, Prakash Nedungadi. 1991. Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters* **2**(3) 181–197.

Shugan, Steven M. 1980. The cost of thinking. *Journal of consumer Research* **7**(2) 99–111.

Small, Kenneth A, Harvey S Rosen. 1981. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society* 105–130.

Tanner, Martin A, Wing Hung Wong. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association* **82**(398) 528–540.

Train, Kenneth. 2003. *Discrete choice methods with simulation*. Cambridge university press.

Tversky, Amos. 1972. Elimination by aspects: A theory of choice. *Psychological review* **79**(4) 281.

Wu, Jianan, Arvind Rangaswamy. 2003. A fuzzy set model of search and consideration with an application to an online market. *Marketing Science* **22**(3) 411–434.

# 7 Figures and Tables

Figure 1: Consumer Screening Process

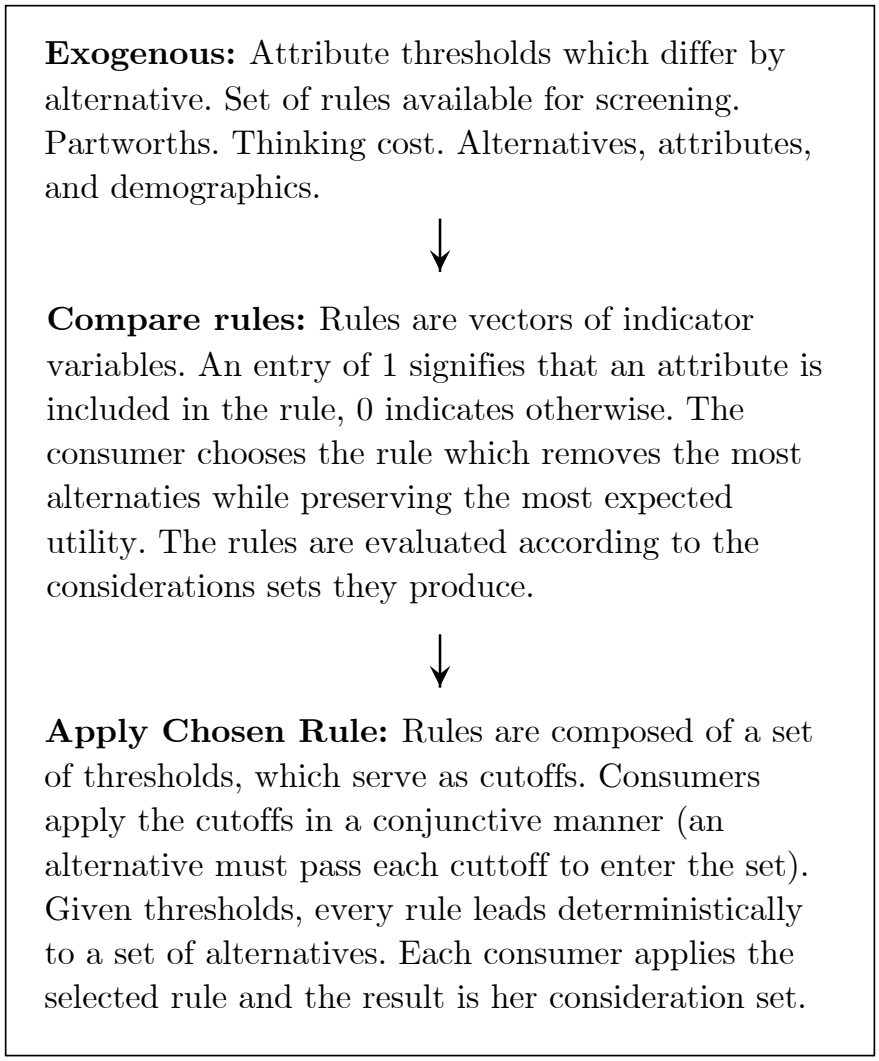

Figure 2: Consideration Set Size Histogram

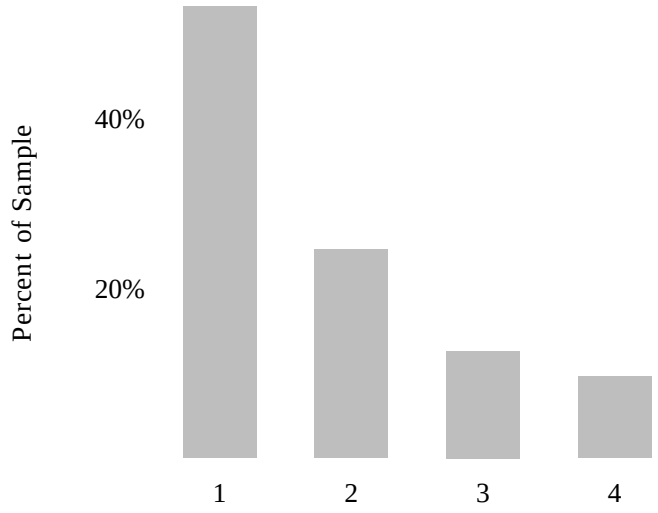


Figure 3: Consideration Shares: Real vs. Model

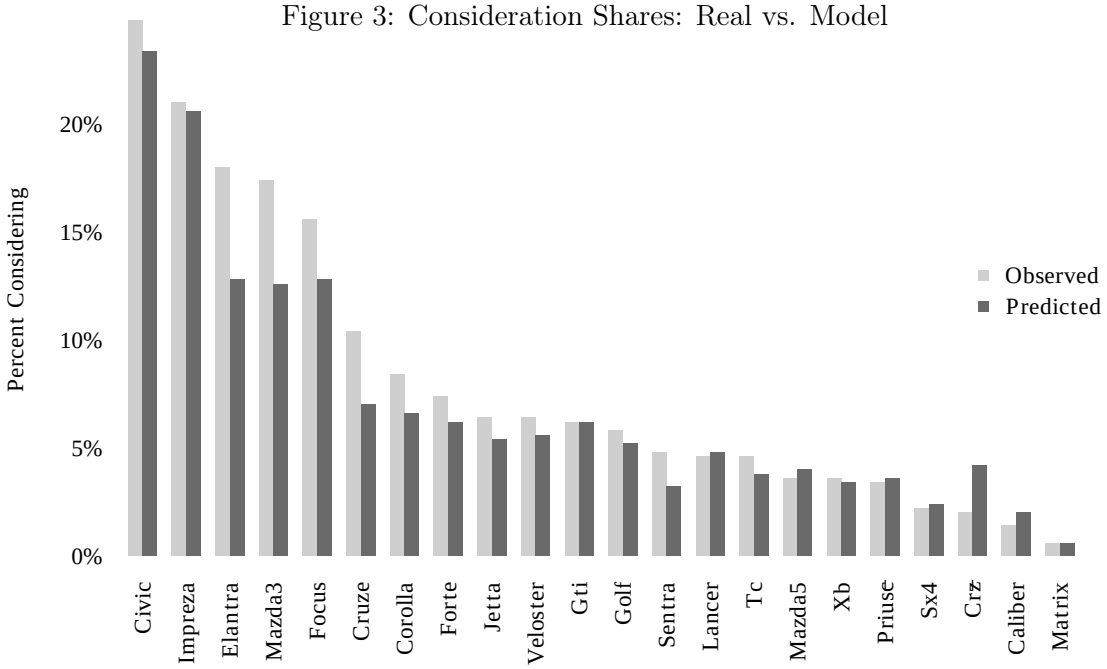

Figure 4: Consideration Shares by Brand vs. Brand Intercepts

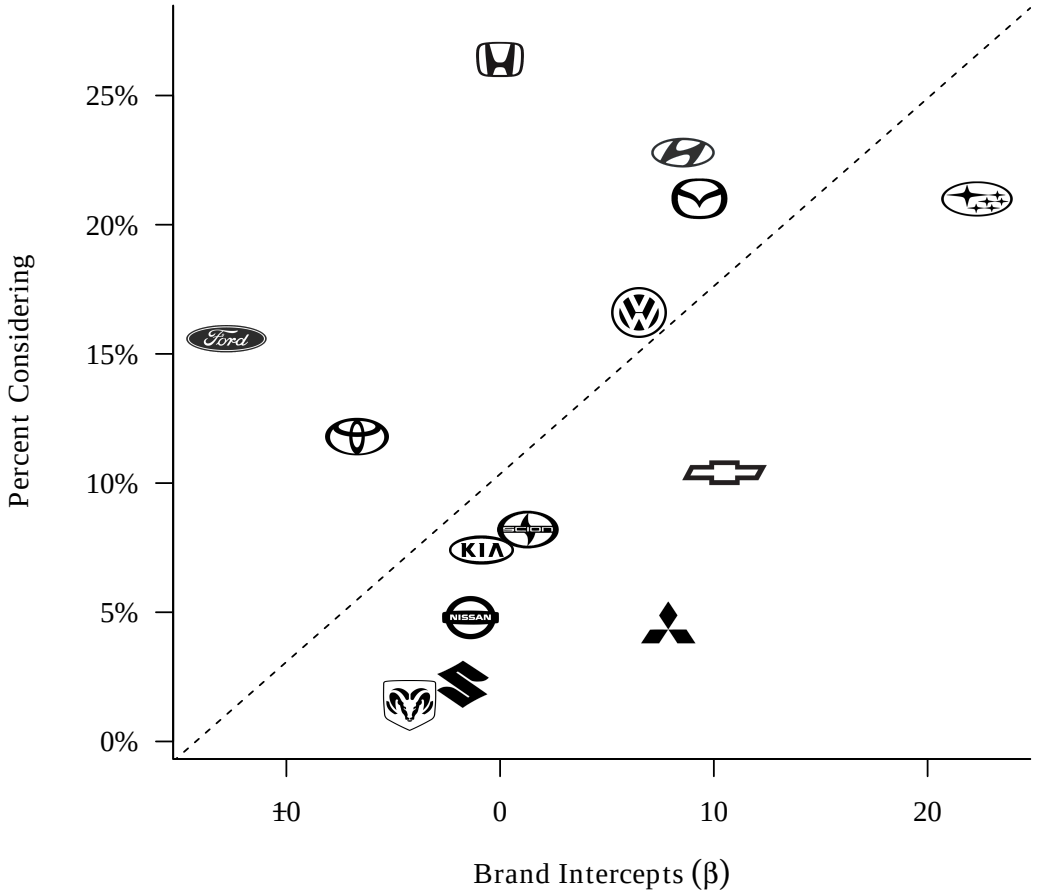

Figure 5: Brand Map: Correspondence

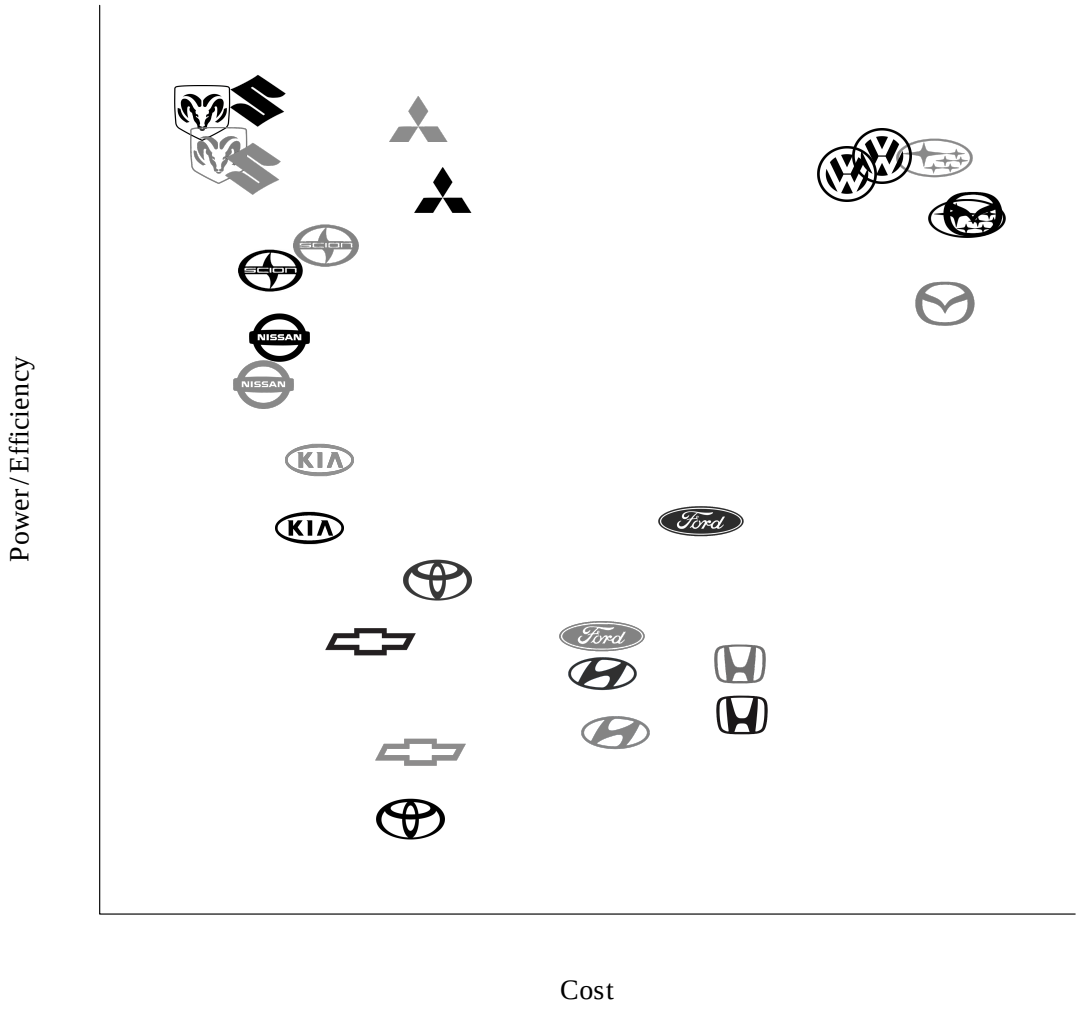

Figure 6: Brand Map: Counterfactual

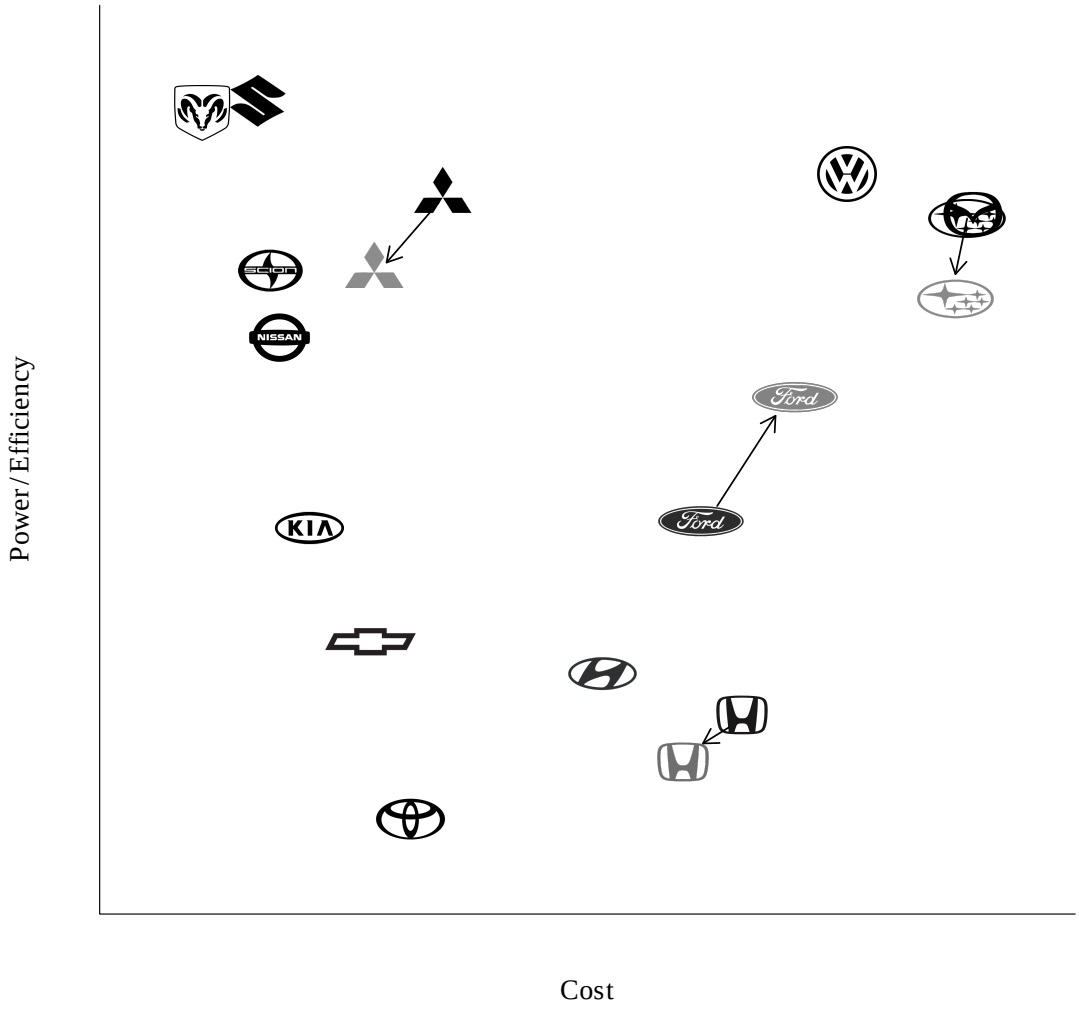

Table 1: Related Literature

| Model | Dependent Variable | Data Type | Outputs |
|---|---|---|---|
| Gilbride, Allenby '04 | Choice | Conjoint | Thresholds, Part-worths |
| Gibride, Allenby '06 | Choice | Conjoint | Thresholds, Part-worths, Costs of Thinking |
| Hauser et al. '10 | Consideration | Conjoint | Rules |
| This paper | Consideration | Survey | Rules, Part-worths, Costs of Thinking, Thresholds, Inertia |

Table 2: MSE on aggregated shares for Different $K$

| | True $K = 3$ | True $K = 4$ |
|---|---|---|
| Est. $K$=3 | .0051 | .0263 |
| Est. $K$ =4 | .0054 | .0068 |

Table 3: Attribute Level Impacts of $\boldsymbol{\beta}$ and $\gamma$

| | $\overline{\boldsymbol{\gamma}} = 0.5$ | $\overline{\boldsymbol{\gamma}} = 1.5$ |
|---|---|---|
| $\boldsymbol{\beta} = 10$ | Mean: 1.38<br>Min: 0.54<br>Max: 2.68 | Mean: 1.75<br>Min: 1.11<br>Max: 2.68 |
| $\boldsymbol{\beta} = 2$ | Mean: 1.35<br>Min: 0.51<br>Max: 2.66 | Mean: 1.72<br>Min: 1.08<br>Max: 2.67 |
| $\boldsymbol{\beta} = 0.5$ | Mean: 1.27<br>Min: 0.48<br>Max: 2.45 | Mean: 1.52<br>Min: 0.84<br>Max: 2.50 |

Table 4: Share Impacts of $\boldsymbol{\beta}$ and $\overline{\boldsymbol{\gamma}}$

| | Alt 1 | Alt 2 | Alt 3 | Alt 4 | Alt 5 | Alt 6 | Avg. Size |
|---|---|---|---|---|---|---|---|
| Attribute Value $\rightarrow$ | 0.06 | 0.72 | 0.88 | 0.77 | 1.16 | 2.68 | |
| $\boldsymbol{\beta} = 0.5, \overline{\boldsymbol{\gamma}} = 0.5$ | 0.58 | 0.56 | 0.57 | 0.71 | 0.62 | 0.87 | 3.9 |
| $\boldsymbol{\beta} = 2, \overline{\boldsymbol{\gamma}} = 0.5$ | 0.53 | 0.55 | 0.56 | 0.70 | 0.62 | 0.99 | 3.9 |
| $\boldsymbol{\beta} = 0.5, \overline{\boldsymbol{\gamma}} = 1$ | 0.54 | 0.50 | 0.51 | 0.66 | 0.56 | 0.89 | 3.6 |
| $\boldsymbol{\beta} = 2, \overline{\boldsymbol{\gamma}} = 1$ | 0.46 | 0.46 | 0.47 | 0.59 | 0.52 | 0.99 | 3.5 |
| $\boldsymbol{\beta} = 0.5, \overline{\boldsymbol{\gamma}} = 1.5$ | 0.52 | 0.45 | 0.47 | 0.58 | 0.50 | 0.89 | 3.4 |
| $\boldsymbol{\beta} = 2, \overline{\boldsymbol{\gamma}} = 1.5$ | 0.42 | 0.37 | 0.37 | 0.48 | 0.41 | 0.99 | 3.0 |

Table 5: Attribute Summary Statistics

| | Median | Mean | Min | Max | S.D. |
|---|---|---|---|---|---|
| Power | 155 | 153 | 99 | 200 | 21 |
| Mpg | 28 | 30 | 24 | 50 | 6 |
| Price | 22145 | 22520 | 18802 | 28415 | 2754 |

Table 6: Demographic Summary Statistics

| | Median | Mean | Min | Max | S.D. | Urban | Rural |
|---|---|---|---|---|---|---|---|
| Age | 46 | 46 | 19 | 95 | 17 | | |
| Education | 10 | 9 | 0 | 13 | 2 | | |
| Location | | | | | | 346 | 154 |

Table 7: Parameter Estimates

| | Variable | Consideration Sets Only | | | Expected Sign |
|---|---|---|---|---|---|
| | | Est. | Low | High | |
| Utility Weights | $\beta$:Power | **5.29** | (3.33, | 7.79) | + |
| | $\beta$:Mpg | **5.21** | (3.39, | 7.36) | + |
| | $\beta$:Price | **-5.63** | (-8.33, | -3.17) | − |
| | $\beta$:Chevrolet | **10.49** | (4.38, | 17.93) | |
| | $\beta$:Dodge | **-4.25** | (-8.55, | -0.54) | |
| | $\beta$:Ford | **-12.81** | (-18.5, | -8.19) | |
| | $\beta$:Honda | 0 | | | |
| | $\beta$:Hyundai | **8.56** | (5.33, | 12.65) | |
| | $\beta$:Kia | -0.86 | (-4.19, | 2.63) | |
| | $\beta$:Mazda | **9.32** | (5.58, | 13.6) | |
| | $\beta$:Mitsubishi | **7.87** | (2.74, | 13.06) | |
| | $\beta$:Nissan | -1.38 | (-5.79, | 2.71) | |
| | $\beta$:Scion | 1.3 | (-1.5, | 4.3) | |
| | $\beta$:Subaru | **22.32** | (15.54, | 30.02) | |
| | $\beta$:Suzuki | -1.73 | (-5.72, | 2.49) | |
| | $\beta$:Toyota | **-6.69** | (-11.83, | -3.5) | |
| | $\beta$:Volkswagen | **6.5** | (1.49, | 10.98) | |
| Cost of Thinking | $\lambda$:Intercept | **3.59** | (2.23, | 5.26) | + |
| | $\lambda$:Age | **0.04** | (0.01, | 0.08) | + |
| | $\lambda$:Education | -0.04 | (-0.18, | 0.09) | − |
| | $\lambda$:Location | 0.35 | (-0.39, | 1) | + |
| Intertia | $\theta$:Inertia | **-0.92** | (-1.01, | -0.8) | − |
| Thresholds | $\overline{\gamma}$:Power | **-0.26** | (-0.41, | -0.11) | |
| | $\overline{\gamma}$:Mpg | **0.46** | (0.23, | 0.65) | |
| | $\overline{\gamma}$:Price | **2.63** | (2.1, | 3.08) | |
| | $\overline{\gamma}$:Chevrolet | **-3.9** | (-4.21, | -3.58) | |
| | $\overline{\gamma}$:Dodge | **-5.12** | (-5.62, | -4.64) | |
| | $\overline{\gamma}$:Ford | **-5.45** | (-5.69, | -5.22) | |
| | $\overline{\gamma}$:Honda | **1.25** | (0.48, | 1.63) | |
| | $\overline{\gamma}$:Hyundai | **-4.94** | (-5.29, | -4.61) | |
| | $\overline{\gamma}$:Kia | **14.98** | (14.76, | 15.19) | |
| | $\overline{\gamma}$:Mazda | **3.28** | (3.04, | 3.51) | |
| | $\overline{\gamma}$:Mitsubishi | **9.41** | (8.98, | 10.11) | |
| | $\overline{\gamma}$:Nissan | **-7.49** | (-7.74, | -7.21) | |
| | $\overline{\gamma}$:Scion | **-5.64** | (-6.05, | -5.19) | |
| | $\overline{\gamma}$:Subaru | **-12.5** | (-13.05, | -12.07) | |
| | $\overline{\gamma}$:Suzuki | **-3.44** | (-3.81, | -3.07) | |
| | $\overline{\gamma}$:Toyota | **2.36** | (2.13, | 2.59) | |
| | $\overline{\gamma}$:Volkswagen | **-2.44** | (-2.82, | -2.06) | |

*Note: 'High' and 'Low' refer to the 95% posterior density region.*

Table 8: Demographic Profiles and Costs of Thinking

| Profile | Demographics | Thinking Costs |
|---|---|---|
| 1 | Age: 19<br>Education: 10<br>Location: Urban | 3.92 ($1918) |
| 2 | Age: 95<br>Education: 0<br>Location: Rural | 7.63 ($3733) |
| 3 | Age: 45<br>Education: 10<br>Location: Urban | 4.95 ($2412) |

Table 9: Individual Demographic Variable Impacts on Costs of Thinking

| Demographic | Thinking Cost Impact of Age |
|---|---|
| Age<br>(Educ. & Loc.<br>at Medians) | Age: 19 → $1918<br>Age: 95 → $3362 |
| Education<br>(Age & Loc.<br>at Medians) | Education: 0 → $2610<br>Education: 13 → $2353 |
| Location<br>(Age & Educ.<br>at Medians) | Location: Urban → $2412<br>Location: Rural → $2585 |

Table 10: Attribute Screening Frequency

| | Percent Time Used |
|---|---|
| Brand | 97 |
| Power | 70 |
| Price | 55 |
| Mpg | 55 |

Table 11: Counterfactual Consideration Shares

| | Real | CC | Price↓ | MPG ↑ | Power↑ |
|---|---|---|---|---|---|
| Caliber | 0.01 | 0.02 | 0.02 | 0.03 | 0.02 |
| Civic | 0.25 | 0.23 | 0.23 | 0.22 | 0.22 |
| Corolla | 0.08 | 0.07 | 0.07 | 0.07 | 0.07 |
| Cr-z | 0.02 | 0.04 | 0.04 | 0.04 | 0.04 |
| Cruze | 0.10 | 0.07 | 0.07 | 0.08 | 0.07 |
| Elantra | 0.18 | 0.13 | 0.14 | 0.13 | 0.14 |
| Focus | 0.16 | 0.13 | 0.15 | 0.15 | 0.14 |
| Forte | 0.07 | 0.06 | 0.07 | 0.06 | 0.07 |
| Golf | 0.06 | 0.05 | 0.05 | 0.06 | 0.06 |
| Gti | 0.06 | 0.06 | 0.07 | 0.06 | 0.06 |
| Impreza | 0.21 | 0.21 | 0.22 | 0.27 | 0.22 |
| Jetta | 0.06 | 0.05 | 0.06 | 0.06 | 0.06 |
| Lancer | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| Matrix | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| Mazda3 | 0.17 | 0.13 | 0.13 | 0.12 | 0.13 |
| Mazda5 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| Prius-c | 0.03 | 0.04 | 0.03 | 0.03 | 0.03 |
| Sentra | 0.05 | 0.03 | 0.04 | 0.04 | 0.03 |
| Sx4 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| Tc | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 |
| Veloster | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |
| Xb | 0.04 | 0.03 | 0.04 | 0.03 | 0.04 |

Table 12: Compact Car Attributes

| | power | mpg | price | Chev | Dodg | Ford | Hond | Hyun | Kia | Mazd | Mits | Niss | Scio | Suba | Suzu | Toyo | Volk |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Caliber | 0.23 | -0.60 | 1.16 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Civic | -0.25 | 0.38 | 0.20 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Corolla | -1.01 | -0.02 | 1.12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Cr-z | -1.49 | 1.01 | -0.20 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cruze | -0.73 | 0.13 | 0.43 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Elantra | -0.31 | 0.42 | 0.63 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Focus | 0.33 | 1.75 | -0.04 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Forte | 0.46 | -0.31 | 0.79 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Golf | 0.23 | -0.06 | -1.43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Gti | 2.24 | -0.59 | -2.14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Impreza | 1.36 | -0.44 | -1.85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Jetta | -0.01 | -0.31 | -1.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Lancer | 0.77 | -0.60 | -1.43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Matrix | -0.27 | -0.70 | 0.19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Mazda3 | 0.24 | -0.35 | -0.06 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mazda5 | 0.18 | -0.94 | 0.19 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Prius-c | -2.59 | 3.46 | -0.06 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Sentra | 0.03 | -0.24 | 1.03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Sx4 | -0.20 | -0.61 | 1.35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Tc | 1.28 | -0.60 | 0.08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Veloster | -0.73 | 0.16 | 0.02 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Xb | 0.23 | -0.94 | 1.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Impreza price | 1.36 | -0.44 | -0.85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Impreza mpg | 1.36 | 0.02 | -1.85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Impreza power | 2.23 | -0.44 | -1.85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |